Managed Network

Managed network refers to an AHV network in which IPAM is enabled.

Unmanaged Network

Unmanaged network refers to an AHV network in which IPAM is not enabled or is disabled.

IPAM is enabled, or not, in the Create Network dialog box when you create a virtual network for Guest VMs. See *Configuring a Virtual Network for Guest VM Interfaces* topic in the *Prism Web Console Guide*.

> Note: You can enable IPAM only when you are creating a virtual network. You cannot enable or disable IPAM for an existing virtual network.

IPAM enabled or disabled status has implications. For example, when you want to reconfigure the IP address of a Prism Central VM, the procedure to do so may involve additional steps for managed networks (that is, networks with IPAM enabled) where the new IP address belongs to an IP address range different from the previous IP address range. See *Reconfiguring the IP Address and Gateway of Prism Central VMs* in *Prism Central Guide*.

Traffic Marking for Quality of Service

To prioritize outgoing (or egress) traffic as required, you can configure quality of service on the traffic for a cluster.

There are two distinct types of outgoing or egress traffic:

- Management traffic (mgmt)
- Data services (data-svc)

Data services traffic consists of the following protocols:

Table 5: Data Services Protocols

Protocol	Port	Nutanix Services
NFS	Source ports (TCP): 445, 2049, 20048, 20049, 20050, and 7508. Source ports (UDP): 2049, 20048, 20049, 20050, and 7508. Source and Destination ports (TCP for Replicator-dr): 7515.	-Nutanix Files-
SMB	Source ports (TCP): 445, 2049, 20048, 20049, 20050, and 7508. Source ports (UDP): 2049, 20048, 20049, 20050, and 7508. Source and Destination ports (TCP for Replicator-dr): 7515.	-Nutanix Files-

Protocol	Port	Nutanix Services
Cluster-to-cluster replications (external or inter-site)	Destination Ports: 2009 and 2020 on CVM.	Stargate and Cerebro
Node-to-node replications (internal or intra-site)	Destination Ports: 2009 and 2020 on CVM.	Stargate and Cerebro
iSCSI	Source Ports: 3260,3261,3205 on CVM. Destination Ports: 3260,3261,3205 on AHV.	Nutanix Files and Volumes

Traffic other than data services traffic is management traffic. Traffic marking for QoS is disabled by default.

When you enable QoS, you can mark both the types of traffic with QoS values. AOS considers the values in hexadecimal even if you provide the values in decimal. When you view or get the QoS configuration enabled on the cluster, nCLI provides the QoS values in hexadecimal format (0xXX where XX is hexadecimal value in the range 00–3f).

> Note: Set any QoS value in the range 0x0–0x3f. The default QoS values for the traffic are as follows:
>
> - Management traffic (mgmt) = 0x10
> - Data services (data-svc) = 0xa

Configuring Traffic Marking for QoS

Configure Quality of Service (QoS) for management and data services traffic using nCLI.

About this task

To perform the following operations for QoS on the egress traffic of a cluster, use the nCLI commands in this section:

- Enable traffic marking for QoS on the cluster. QoS traffic marking is disabled by default.
- View or get the QoS configuration enabled on the cluster.
- Set QoS values for all traffic types or specific traffic types.
- Disable QoS on the cluster.

When you run any of the QoS configuration commands and the command succeeds, the console displays the following output indicating the successful command run:

```
QoSUpdateStatusDTO(status=true, message=null)
```

Where:

- `status=true` indicates that the command succeeded.
- `message=null` indicates that there is no error.

When you run any of the QoS configuration commands and the command fails, the console displays the following sample output indicating the failure:

```
QoSUpdateStatusDTO(status=false, message=QoS is already enabled.)
```

Where:

- `status=false` indicates that the command failed.
- `message=QoS is already enabled.` indicates why the command failed. This sample error message indicates that the `net enable-qos` command failed because QoS enable command was run again when QoS is already enabled.

Procedure

- To enable QoS on a cluster, run the following command:

  ```
  ncli> net enable-qos [data-svc="data-svc value"][mgmt="mgmt value"]
  ```

  If you run the command as `net enable-qos` without the options, AOS enables QoS with the default values (`mgmt=0x10` and `data-svc=0xa`).

  > Note: After you run the `net enable-qos` command, if you run it again, the command fails and AOS displays the following output:
  >
  > `QoSUpdateStatusDTO(status=false, message=QoS is already enabled.)`

  > Note: If you need to change the QoS values after you enable it , run the `net edit-qos` command with the option (`data-svc` or `mgmt` or both as necessary).

  > Note: Set any QoS value in the range 0x0–0x3f.

- To view or get the QoS configuration enabled on a cluster, run the following command:

  ```
  ncli> net get-qos
  ```

  > Note: When you get the QoS configuration enabled on the cluster, nCLI provides the QoS values in hexadecimal format (0xXX where XX is hexadecimal value in the range 00–3f).

  A sample output on the console is as follows:

  ```
  QoSDTO(status=true, isEnabled=true, mgmt=0x10, dataSvc=0xa, message=null)
  ```

  > Note:
  >
  > Where:
  >
  > - `status=true` indicates that the `net get-qos` command passed. `status=false` indicates that the `net get-qos` command failed. See the `message=` value for the failure error message.
  > - `isEnabled=true` indicates that QoS is enabled. `isEnabled=false` indicates that QoS is not enabled.
  > - `mgmt=0x10` indicates that QoS value for Management traffic (*`mgmt`* option) is set to 0x10 (represented in hexadecimal value as `0x10`. If you disabled QoS, then this parameter is displayed as `mgmt=null`.
  > - `dataSvc=0xa` indicates that QoS value for data services traffic (*`data-svc`* option) is set to 0xa (represented in hexadecimal value as `0xa`. If you disabled QoS, then this parameter is displayed as `dataSvc=null`.
  > - `message=null` indicates there is no error message. `message=` parameter provides the command failure error message if the command fails.

- To set the QoS values for the traffic types on a cluster after you enabled QoS on the cluster, run the following command:

  ```
  ncli> net edit-qos [data-svc="data-svc value"][mgmt="mgmt value"]
  ```

  You can provide QoS values between 0x0-0x3f for one or both the options. The value is hexadecimal representation of a value between decimal 0-63 both inclusive.

- To disable QoS on a cluster, run the following command:

  ```
  ncli> net disable-qos
  ```

  ```
  QoSDTO(status=true, isEnabled=false, mgmt=null, dataSvc=null, message=null)
  ```

Layer 2 Network Management

AHV uses virtual switch (VS) to connect the Controller VM, the hypervisor, and the guest VMs to each other and to the physical network. Virtual switch is configured by default on each AHV node and the VS services start automatically when you start a node.

To configure virtual networking in an AHV cluster, you need to be familiar with virtual switch. This documentation gives you a brief overview of virtual switch and the networking components that you need to configure to enable the hypervisor, Controller VM, and guest VMs to connect to each other and to the physical network.

About Virtual Switch

Virtual switches or VS are used to manage multiple bridges and uplinks.

The VS configuration is designed to provide flexibility in configuring virtual bridge connections. A virtual switch (VS) defines a collection of AHV nodes and the uplink ports on each node. It is an aggregation of the same OVS bridge on all the compute nodes in a cluster. For example, vs0 is the default virtual switch is an aggregation of the br0 bridge and br0-up uplinks of all the nodes.

After you configure a VS, you can use the VS as reference for physical network management instead of using the bridge names as reference.

For overview about Virtual Switch, see Virtual Switch Considerations on page 61.

For information about OVS, see About Open vSwitch on page 65.

Virtual Switch Workflow

A virtual switch (VS) defines a collection of AHV compute nodes and the uplink ports on each node. It is an aggregation of the same OVS bridge on all the compute nodes in a cluster. For example, vs0 is the default virtual switch is an aggregation of the br0 bridge of all the nodes.

The system creates the default virtual switch vs0 connecting the default bridge br0 on all the hosts in the cluster during installation of or upgrade to the compatible versions of AOS and AHV. Default virtual switch vs0 has the following characteristics:

- The default virtual switch cannot be deleted.
- The default bridges br0 on all the nodes in the cluster map to vs0. thus, vs0 is not empty. It has at least one uplink configured.
- The default management connectivity to a node is mapped to default bridge br0 that is mapped to vs0.
- The default parameter values of vs0 - Name, Description, MTU and Bond Type - can be modified subject to aforesaid characteristics.

- The default virtual switch is configured with the **Active-Backup** uplink bond type.

  For more information about bond types, see the Bond Type table.

The virtual switch aggregates the same bridges on all nodes in the cluster. The bridges (for example, br1) connect to the physical port such as eth3 (Ethernet port) via the corresponding uplink (for example, br1-up). The uplink ports of the bridges are connected to the same physical network. For example, the following illustration shows that vs0 is mapped to the br0 bridge, in turn connected via uplink br0-up to various (physical) Ethernet ports on different nodes.

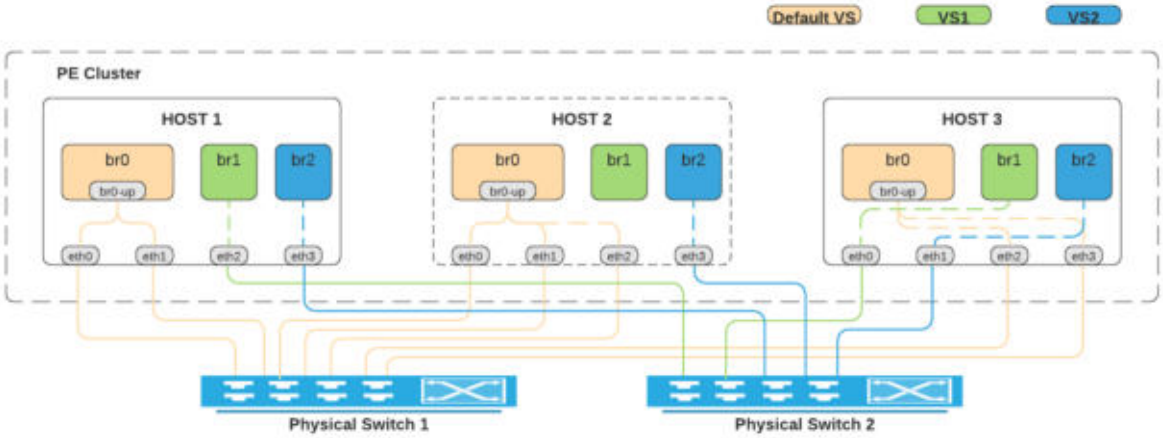


Figure 20: Virtual Switch

Uplink configuration uses bonds to improve traffic management. The bond types are defined for the aggregated OVS bridges. A new bond type - No uplink bond - provides a no-bonding option. A virtual switch configured with the No uplink bond uplink bond type has 0 or 1 uplinks.

When you configure a virtual switch with any other bond type, you must select at least two uplink ports on every node.

If you change the uplink configuration of vs0, AOS applies the updated settings to all the nodes in the cluster one after the other (the rolling update process). To update the settings in a cluster, AOS performs the following tasks when configuration method applied is **Standard**:

1. Puts the node in maintenance mode (migrates VMs out of the node)
2. Applies the updated settings
3. Checks connectivity with the default gateway
4. Exits maintenance mode
5. Proceeds to apply the updated settings to the next node

AOS does not put the nodes in maintenance mode when the **Quick** configuration method is applied.

Table 6: Bond Types

Bond Type	Use Case	Maximum VM NIC Throughput	Maximum Host Throughput
Active-Backup	Recommended. Default configuration, which transmits all traffic over a single active adapter.	10 Gb	10 Gb
Active-Active with MAC pinning Also known as balance-slb	Works with caveats for multicast traffic. Increases host bandwidth utilization beyond a single 10 Gb adapter. Places each VM NIC on a single adapter at a time. Do not use this bond type with link aggregation protocols such as LACP.	10 Gb	20 Gb
Active-Active Also known as LACP with balance-tcp	LACP and link aggregation required. Increases host and VM bandwidth utilization beyond a single 10 Gb adapter by balancing VM NIC TCP and UDP sessions among adapters. Also used when network switches require LACP negotiation. The default LACP settings are: • Speed—Fast (1s) • Mode—Active fallback-active-backup • Priority—Default. This is not configurable.	20 Gb	20 Gb
No Uplink Bond	No uplink or a single uplink on each host. Virtual switch configured with the No uplink bond uplink bond type has 0 or 1 uplinks. When you configure a virtual switch with any other bond type, you must select at least two uplink ports on every node.	-	-

Note the following points about the uplink configuration.

- Virtual switches are not enabled in a cluster that has one or more compute-only nodes. See Virtual Switch Limitations on page 68 and Virtual Switch Requirements on page 67.

- If you select the **Active-Active** policy, you must manually enable LAG and LACP on the corresponding ToR switch for each node in the cluster.
- If you reimage a cluster with the **Active-Active** policy enabled, the default virtual switch (vs0) on the reimaged cluster is once again the **Active-Backup** policy. The other virtual switches are removed during reimage.
- Nutanix recommends configuring LACP with fallback to active-backup or individual mode on the ToR switches. The configuration and behavior varies based on the switch vendor. Use a switch configuration that allows both switch interfaces to pass traffic after LACP negotiation fails.

Virtual Switch Considerations

Virtual Switch Deployment

A VS configuration is deployed using rolling update of the clusters. After the VS configuration (creation or update) is received and execution starts, every node is first put into maintenance mode before the VS configuration is made or modified on the node. This is called the **Standard** recommended default method of configuring a VS.

You can select the **Quick** method of configuration also where the rolling update does not put the clusters in maintenance mode. The VS configuration task is marked as successful when the configuration is successful on the first node. Any configuration failure on successive nodes triggers corresponding NCC alerts. There is no change to the task status.

> Note: If you are modifying an existing bond, AHV removes the bond and then re-creates the bond with the specified interfaces.
>
> Ensure that the interfaces you want to include in the bond are physically connected to the Nutanix appliance before you run the command described in this topic. If the interfaces are not physically connected to the Nutanix appliance, the interfaces are not added to the bond.
>
> Ensure that the pre-checks listed in *LCM Prechecks* section of the Life Cycle Manager Guide and the Always and Host Disruptive Upgrades types of pre-checks listed KB-4584 pass for Virtual Switch deployments.

The VS configuration is stored and re-enforced at system reboot.

The VM NIC configuration also displays the VS details. When you Update VM configuration or Create NIC for a VM, the NIC details show the virtual switches that can be associated. This allows you to change a virtual network and the associated virtual switch.

To change the virtual network, select the virtual network in the **Subnet Name** dropdown list in the Create NIC or Update NIC dialog box..

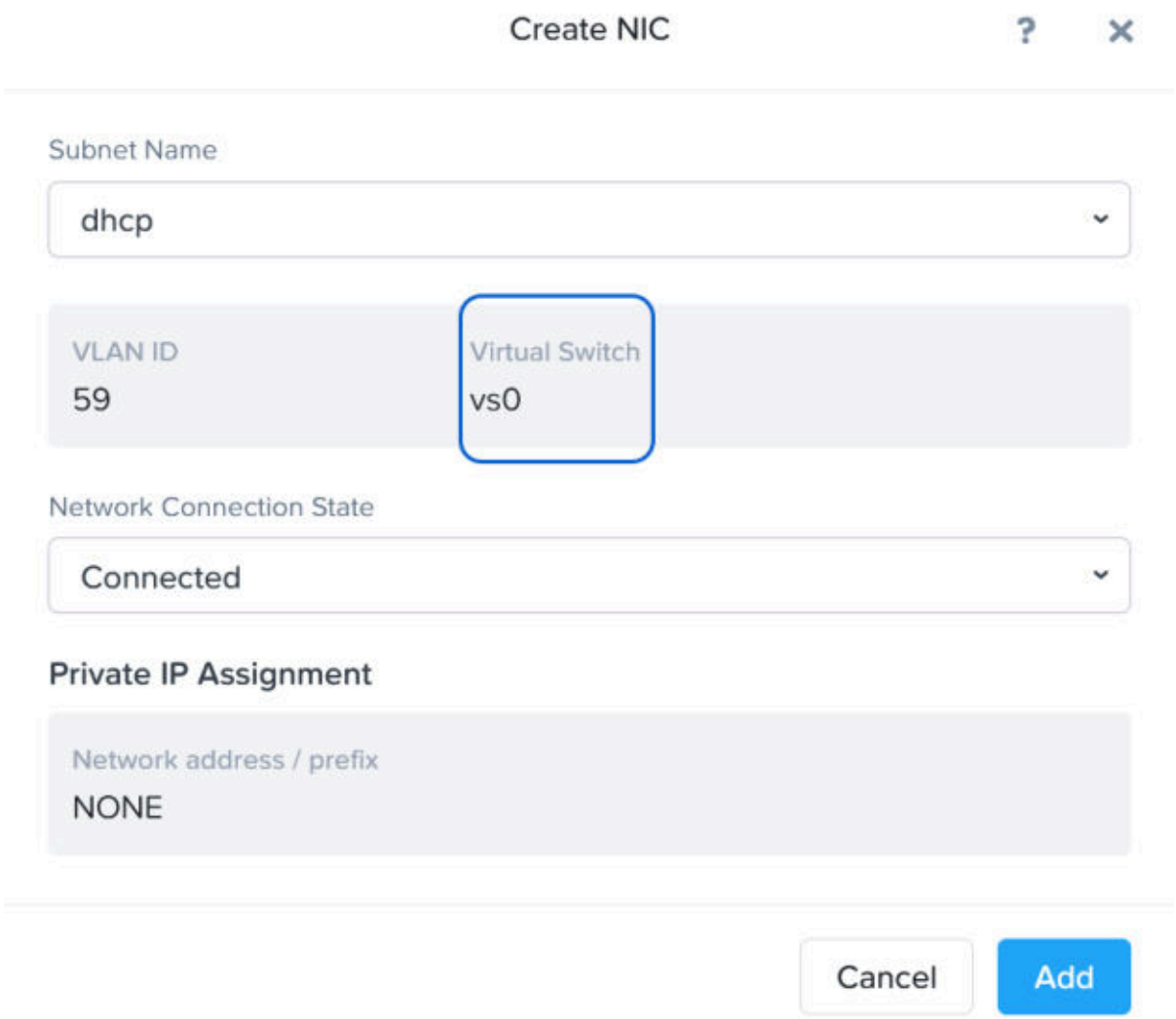


Figure 21: Create VM - VS Details

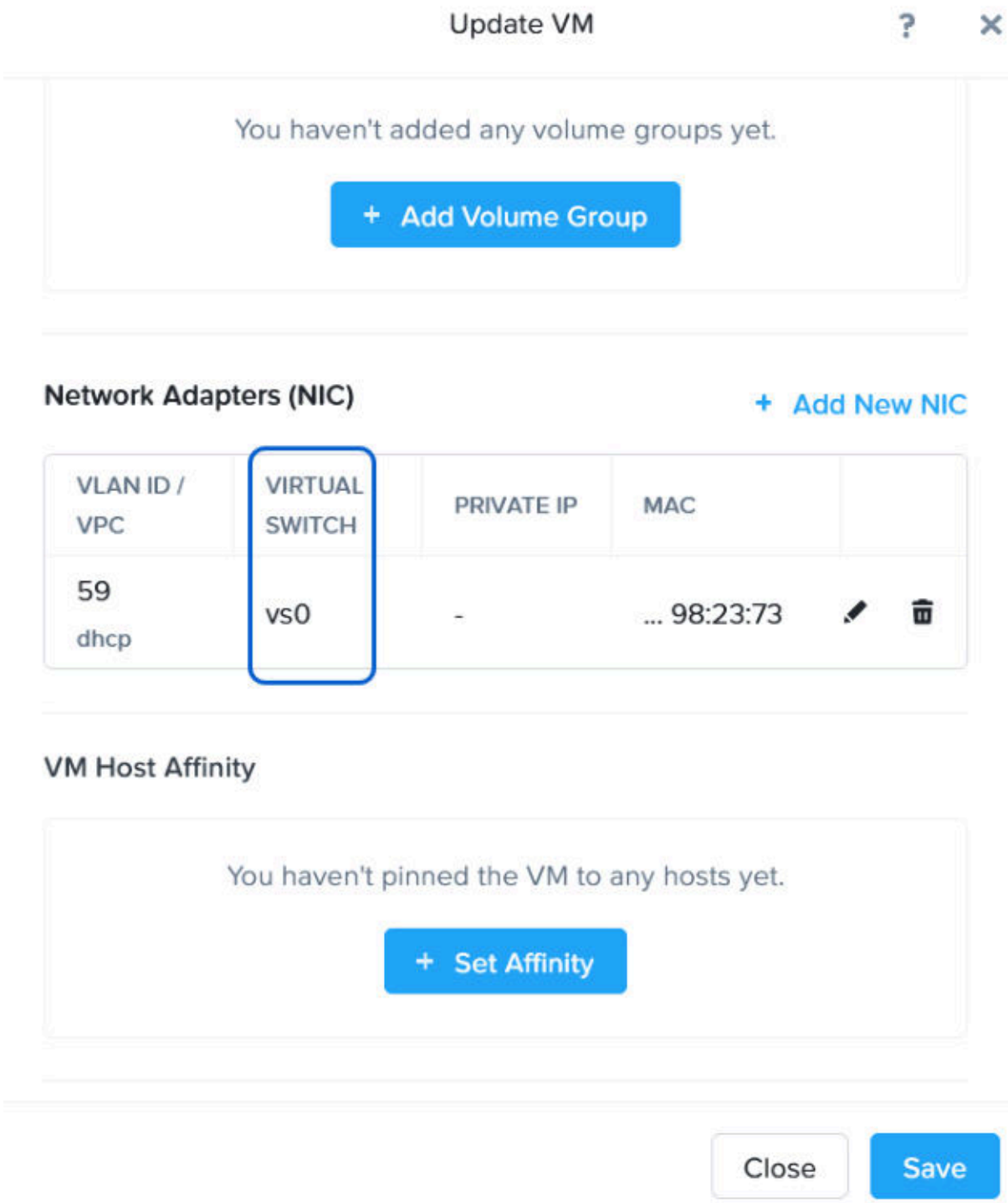


Figure 22: VM NIC - VS Details

Impact of Installation of or Upgrade to Compatible AOS and AHV Versions

See *Virtual Switch Requirements* for information about minimum and compatible AOS and AHV versions.

When you upgrade the AOS to a compatible version from an older version, the upgrade process:

- Triggers the creation of the default virtual switch vs0 which is mapped to bridge br0 on all the nodes.
- Validates bridge br0 and its uplinks for consistency in terms of MTU and bond-type on every node.

  If valid, it adds the bridge br0 of each node to the virtual switch vs0.

  If br0 configuration is not consistent, the system generates an NCC alert which provides the failure reason and necessary details about it.

  The system migrates only the bridge br0 on each node to the default virtual switch vs0 because the connectivity of bridge br0 is guaranteed.
- Does not migrate any other bridges to any other virtual switches during upgrade. You need to manually migrate the other bridges after instal or upgrade is complete.

Bridge Migration

After upgrading to a compatible version of AOS, you can migrate bridges other than br0 that existed on the nodes. When you migrate the bridges, the system converts the bridges to virtual switches.

See *Virtual Switch Migration Requirements* in Virtual Switch Requirements on page 67.

> Note: You can migrate only those bridges that are present on every compute node in the cluster. See *Migrating Bridges after Upgrade* topic in *Network Management* in the *Prism Web Console Guide*.

Cluster Scaling Impact

VS management for cluster scaling (addition or removal of nodes) is seamless.

Node Removal

When you remove a node, the system detects the removal and automatically removes the node from all the VS configurations that include the node and generates an internal system update. For example, a node has two virtual switches, vs1 and vs2, configured apart from the default vs0. When you remove the node from the cluster, the system removes the node for the vs1 and vs2 configurations automatically with internal system update.

Node Addition

When you add a new node or host to a cluster, the bridges or virtual switches on the new node are treated in the following manner:

> Note: If a host already included in a cluster is removed and then added back, it is treated as a new host.

- The system validates the default bridge br0 and uplink bond br0-up to check if it conforms to the default virtual switch vs0 already present on the cluster.

  If br0 and br0-up conform, the system includes the new host and its uplinks in vs0.

  If br0 and br0-up do not conform,then the system generates an NCC alert.
- The system does not automatically add any other bridge configured on the new host to any other virtual switch in the cluster.

  It generates NCC alerts for all the other non-default virtual switches.

- You can manually include the host in the required non-default virtual switches. Update a non-default virtual switch to include the host.

  For information about updating a virtual switch in Prism Element Web Console, see the *Configuring a Virtual Network for Guest VMs* section in the *Prism Web Console Guide*.

  For information about updating a virtual switch in Prism Central, see the *Network Connections* section in the *Prism Central Guide*.

VS Management

You can manage virtual switches from Prism Central or Prism Web Console. You can also use aCLI or REST APIs to manage them. See the *Acropolis API Reference* and *Command Reference* guides for more information.

You can also use the appropriate aCLI commands for virtual switches from the following list:

- net.create_virtual_switch
- net.list_virtual_switch
- net.get_virtual_switch
- net.update_virtual_switch
- net.delete_virtual_switch
- net.migrate_br_to_virtual_switch
- net.disable_virtual_switch

About Open vSwitch

Open vSwitch (OVS) is an open-source software switch implemented in the Linux kernel and designed to work in a multiserver virtualization environment. By default, OVS behaves like a Layer 2 learning switch that maintains a MAC address learning table. The hypervisor host and VMs connect to virtual ports on the switch.

Each hypervisor hosts an OVS instance, and all OVS instances combine to form a single switch. As an example, the following diagram shows OVS instances running on two hypervisor hosts.

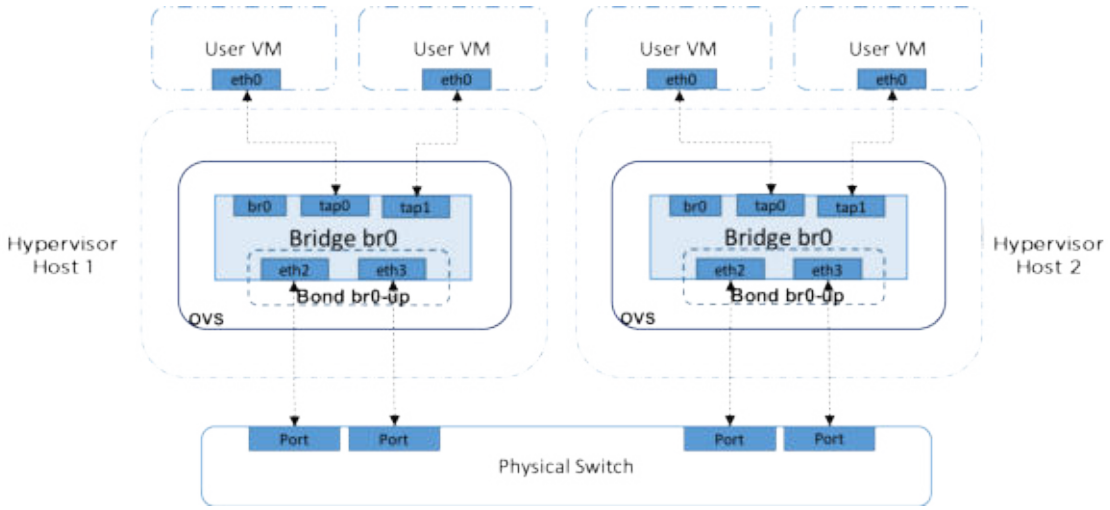


Figure 23: Open vSwitch

Default Factory Configuration

The factory configuration of an AHV host includes a default OVS bridge named br0 (configured with the default virtual switch vs0) and a native linux bridge called virbr0.

Bridge br0 includes the following ports by default:

- An internal port with the same name as the default bridge; that is, an internal port named br0. This is the access port for the hypervisor host.
- A bonded port named br0-up. The bonded port aggregates all the physical interfaces available on the node. For example, if the node has two 10 GbE interfaces and two 1 GbE interfaces, all four interfaces are aggregated on br0-up. This configuration is necessary for Foundation to successfully image the node regardless of which interfaces are connected to the network.

> Note:
>
> Before you begin configuring a virtual network on a node, you must disassociate the 1 GbE interfaces from the br0-up port. This disassociation occurs when you modify the default virtual switch (vs0) and create new virtual switches. Nutanix recommends that you aggregate only the 10 GbE or faster interfaces on br0-up and use the 1 GbE interfaces on a separate OVS bridge deployed in a separate virtual switch.
>
> See Virtual Switch Management on page 69 for information about virtual switch management.

The following diagram illustrates the default factory configuration of OVS on an AHV node:

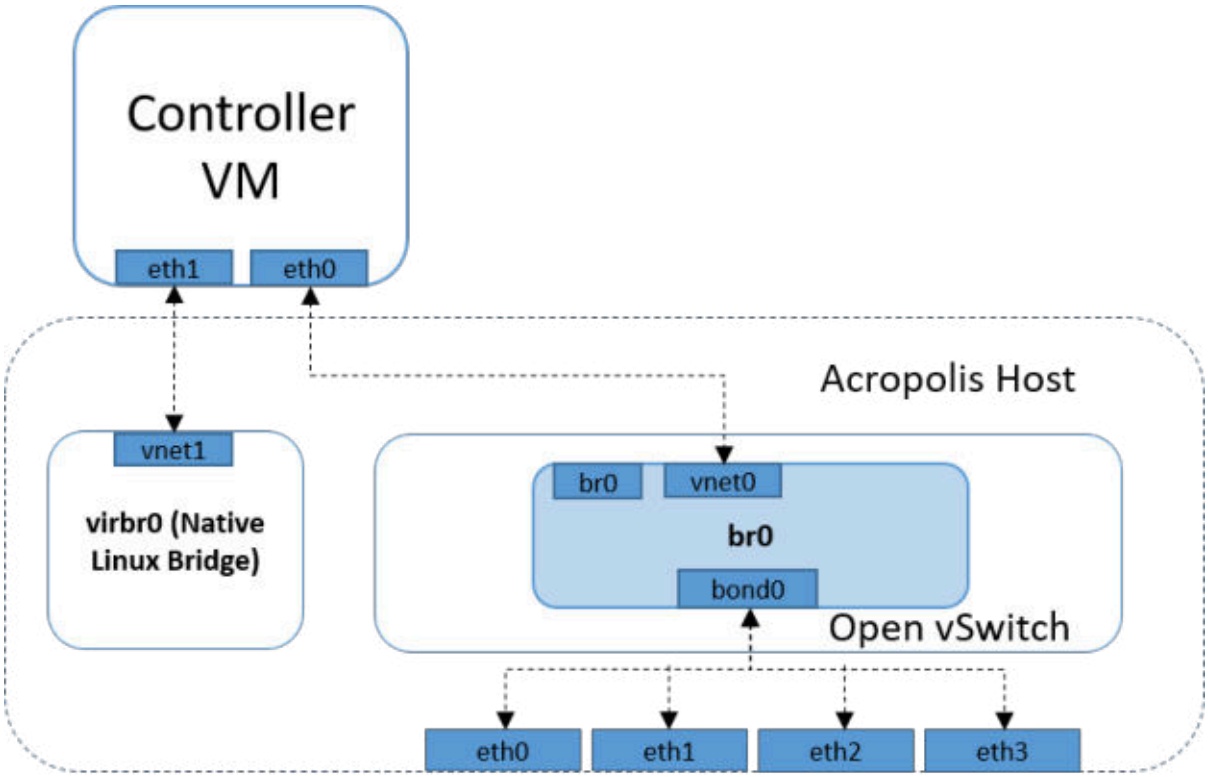


Figure 24: Default factory configuration of Open vSwitch in AHV

The Controller VM has two network interfaces by default. As shown in the diagram, one network interface connects to bridge br0. The other network interface connects to a port on virbr0. The Controller VM uses this bridge to communicate with the hypervisor host.

Virtual Switch Requirements

The requirements to deploy virtual switches are as follows:

1. Virtual switches are supported on AOS 5.19 or later with AHV 20201105.12 or later. Therefore you must install or upgrade to AOS 5.19 or later, with AHV 20201105.12 or later, to use virtual switches in your deployments.
2. Virtual bridges used for a VS on all the nodes must have the same specification such as name, MTU and uplink bond type. For example, if vs1 is mapped to br1 (virtual or OVS bridge 1) on a node, it must be mapped to br1 on all the other nodes of the same cluster.

Virtual Switch Migration Requirements

The AOS upgrade process initiates the virtual switch migration. The virtual switch migration is successful only when the following requirements are fulfilled:

- Before migrating to Virtual Switch, all bridge br0 bond interfaces should have the same bond type on all hosts in the cluster. For example, all hosts should use the Active-backup bond type or balance-tcp. If some hosts use Active-backup and other hosts use balance-tcp, virtual switch migration fails.
- Before migrating to Virtual Switch, if using LACP:
  - Confirm that all bridge br0 lacp-fallback parameters on all hosts are set to the case sensitive value `True` with `manage_ovs show_uplinks |grep lacp-fallback:`. Any host with lowercase `true` causes virtual switch migration failure.
  - Confirm that the LACP speed on the physical switch is set to `fast` or `1` second. Also ensure that the switch ports are ready to fallback to individual mode if LACP negotiation fails due to a configuration such as `no lacp suspend-individual`.
- Before migrating to the Virtual Switch, confirm that the upstream physical switch is set to `spanning-tree portfast` or `spanning-tree port type edge trunk`. Failure to do so may lead to a 30-second network timeout and the virtual switch migration may fail because it uses 20-second non-modifiable timer.
- Ensure that the pre-checks listed in *LCM Prechecks* section of the Life Cycle Manager Guide and the Always and Host Disruptive Upgrades types of pre-checks listed KB-4584 pass for Virtual Switch deployments.
- For the default virtual switch vs0,
  - All configured uplink ports must be available for connecting the network. In Active-Backup bond type, the active port is selected from any configured uplink port that is linked. Therefore, the virtual switch vs0 can use all the linked ports for communication with other CVMs/hosts.
  - All the host IP addresses in the virtual switch vs0 must be resolvable to the configured gateway using ARP.

Virtual Switch Limitations

MTU Restriction

- Nutanix recommends increasing the MTU to 9000 bytes on the virtual switch vs0 and ensure that the physical networking infrastructure supports higher MTU values (jumbo frame support). The recommended MTU range is 1600-9000 bytes.

  Nutanix CVMs use the standard Ethernet MTU (maximum transmission unit) of 1,500 bytes for all the network interfaces by default. The system advertises the MTU of 1442 bytes to guest VMs using DHCP to account for the extra 58 bytes used by Generic Network Virtualization Encapsulation (Geneve). However, some VMs ignore the MTU advertisements in the DHCP response. Therefore, to ensure that Flow Networking functions properly with such VMs, enable jumbo frame support on the physical network and the default virtual switch vs0.

  If you cannot increase the MTU of the physical network, you can decrease the MTU of every VM vNIC to 1442 bytes.

  > Note: Do not decrease the MTU of virtual switch vs0 lower than 1500 bytes without first decreasing the MTU of CVM and guest VM vNICs.

Single node and Two-node cluster configuration.

Virtual switch cannot be deployed is your single-node or two-node cluster has any instantiated user VMs. The virtual switch creation or update process involves a rolling restart, which checks for maintenance mode and whether you can migrate the VMs. On a single-node or two-node cluster, any instantiated user VMs cannot be migrated and the virtual switch operation fails.

Therefore, power down all user VMs for virtual switch operations in a single-node or two-node cluster.

Compute-only node is not supported.

Virtual switch is not compatible with Compute-only (CO) nodes. If a CO node is present in the cluster, then the virtual switches are not deployed (including the default virtual switch). You need to use the `net.disable_virtual_switch` aCLI command to disable the virtual switch workflow if you want to expand a cluster which has virtual switches and includes a CO node.

The `net.disable_virtual_switch` aCLI command cleans up all the virtual switch entries from the IDF. All the bridges mapped to the virtual switch or switches are retained as they are.

Including a storage-only node in a VS is not necessary.

Virtual switch is compatible with Storage-only (SO) nodes but you do not need to include an SO node in any virtual switch, including the default virtual switch.

Mixed-mode Clusters with AHV Storage-only Nodes

Consider that you have deployed a mixed-node cluster where the compute-only nodes are ESXi or Hyper-V nodes and the storage-only nodes are AHV nodes. In such a case, the default virtual switch deployment fails.

Without the default VS, the Prism Element, Prism Central and CLI workflows for virtual switch required to manage the bridges and bonds are not available. You need to use the `manage_ovs` command options to update the bridge and bond configurations on the AHV hosts.

Virtual Switch Management

Virtual Switch can be viewed, created, updated or deleted from both Prism Web Console as well as Prism Central.

Virtual Switch Views and Visualization

For information on the virtual switch network visualization in Prism Element Web Console, see the *Network Visualization* topic in the *Prism Web Console Guide*.

Virtual Switch Create, Update and Delete Operations

For information about the procedures to create, update and delete a virtual switch in Prism Element Web Console, see the *Configuring a Virtual Network for Guest VMs* section in the *Prism Web Console Guide*.

For information about the procedures to create, update and delete a virtual switch in Prism Central, see the *Network Connections* section in the *Prism Central Guide*.

Enabling LAG and LACP on the ToR Switch (AHV Only)

Active-Active with MAC pinning does not require link aggregation since each source MAC address is balanced to only one adapter at a time. If you select the **Active-Active** NIC-teaming policy, you must enable LAG and LACP on the corresponding ToR switch for each node in the cluster one after the other.

About this task

In an **Active-Active** NIC-teaming configuration, network traffic is balanced among the members of the team based on source and destination IP addresses and TCP and UDP ports. With link aggregation negotiated by LACP, the uplinks appear as a single layer-2 link so a VM MAC address can appear on multiple links and use the bandwidth of all uplinks. If you do not enable LAG and LACP in this **Active-Active** configuration, network traffic might be dropped by the switch.

Perform the following procedure to enable LAG and LACP on a TOR switch after enabling the Active-Active policy by using the Prism Element web console.

Procedure

1. Put the node in maintenance mode. This is in addition to the previous maintenance mode that enabled Active-Active on the node.
2. Enable LAG and LACP on the ToR switch connected to that node.
3. Exit maintenance mode after LAG and LACP is successfully enabled.
4. Repeat steps 1 to 3 for every node in the cluster.

VLAN Configuration

You can set up a VLAN-based segmented virtual network on an AHV node by assigning the ports on virtual bridges managed by virtual switches to different VLANs. VLAN port assignments are configured from the Controller VM that runs on each node.

For best practices associated with VLAN assignments, see AHV Networking Recommendations on page 49. For information about assigning guest VMs to a virtual switch and VLAN, see *Network Connections* section in *Prism Central Guide*..

Assigning an AHV Host to a VLAN

About this task

> Note: Perform the following procedure during a scheduled downtime. Before you begin, stop the cluster. Once the process begins, hosts and CVMs partially lose network access to each other and VM data or storage containers become unavailable until the process completes.

To assign an AHV host to a VLAN, do the following on every AHV host in the cluster:

Procedure

1. Log on to the AHV host with SSH.
2. Put the AHV host and the CVM in maintenance mode.
   See Putting a Node into Maintenance Mode using CLI on page 34 for instructions about how to put a node into maintenance mode.
3. Assign port br0 (the internal port on the default OVS bridge, br0 on defaul virtual switch vs0) to the VLAN that you want the host be on.

   ```
   root@ahv# ovs-vsctl set port br0 tag=host_vlan_tag
   ```

   Replace *host_vlan_tag* with the VLAN tag for hosts.
4. Confirm VLAN tagging on port br0.

   ```
   root@ahv# ovs-vsctl list port br0
   ```

5. Check the value of the tag parameter that is shown.
6. Verify connectivity to the IP address of the AHV host by performing a ping test.
7. Exit the AHV host and the CVM from the maintenance mode.
   See Exiting a Node from the Maintenance Mode Using CLI on page 36 for more information.

Assigning the Controller VM to a VLAN

By default, the public interface of a Controller VM is assigned to VLAN 0. To assign the Controller VM to a different VLAN, change the VLAN ID of its public interface. After the change, you can access the public interface from a device that is on the new VLAN.

About this task

> Note: Perform the following procedure during a scheduled downtime. Before you begin, stop the cluster. Once the process begins, hosts and CVMs partially lose network access to each other and VM data or storage containers become unavailable until the process completes.

> Note: To avoid losing connectivity to the Controller VM, do not change the VLAN ID when you are logged on to the Controller VM through its public interface. To change the VLAN ID, log on to the internal interface that has IP address 192.168.5.254.

Perform these steps on every Controller VM in the cluster. To assign the Controller VM to a VLAN, do the following:

Procedure

1. Log on to the AHV host with SSH.

2. Put the AHV host and the Controller VM in maintenance mode.

   See Putting a Node into Maintenance Mode using CLI on page 34 for instructions about how to put a node into maintenance mode.

3. Check the Controller VM status on the host.

```
root@host# virsh list
```

   An output similar to the following is displayed:

```
root@host# virsh list
 Id    Name                           State
----------------------------------------------------
 1     NTNX-CLUSTER_NAME-3-CVM             running
 3     3197bf4a-5e9c-4d87-915e-59d4aff3096a running
 4     c624da77-945e-41fd-a6be-80abf06527b9 running

root@host# logout
```

4. Log on to the Controller VM.

```
root@host# ssh nutanix@192.168.5.254
```

   Accept the host authenticity warning if prompted, and enter the Controller VM nutanix password.

5. Assign the public interface of the Controller VM to a VLAN.

```
nutanix@cvm$ change_cvm_vlan vlan_id
```

   Replace *vlan_id* with the ID of the VLAN to which you want to assign the Controller VM.

   For example, add the Controller VM to VLAN 201.

```
nutanix@cvm$ change_cvm_vlan 201
```

6. Confirm VLAN tagging on the Controller VM.

```
root@host# virsh dumpxml cvm_name
```

   Replace *cvm_name* with the CVM name or CVM ID to view the VLAN tagging information.

   > Note: Refer to step 3 for Controller VM name and Controller VM ID.

   An output similar to the following is displayed:

```
root@host# virsh dumpxml 1 | grep "tag id" -C10 --color
      <target dev='vnet2'/>
      <model type='virtio'/>
      <driver name='vhost' queues='4'/>
      <alias name='net2'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x05' function='0x0'/>
    </interface>
    <interface type='bridge'>
      <mac address='50:6b:8d:b9:0a:18'/>
      <source bridge='br0'/>
      <vlan>
           <tag id='201'/>
      </vlan>
      <virtualport type='openvswitch'>
        <parameters interfaceid='c46374e4-c5b3-4e6b-86c6-bfd6408178b5'/>
      </virtualport>
      <target dev='vnet0'/>
      <model type='virtio'/>
      <driver name='vhost' queues='4'/>
```

```
      <alias name='net3'/>
      <address type='pci' domain='0x0000' bus='0x00' slot='0x03' function='0x0'/>
    </interface>
root@host#
```

7. Check the value of the tag parameter that is shown.

8. Restart the network service.

```
nutanix@cvm$ sudo service network restart
```

9. Verify connectivity to the Controller VMs external IP address by performing a ping test from the same subnet. For example, perform a ping from another Controller VM or directly from the host itself.

10. Exit the AHV host and the Controller VM from the maintenance mode.
    See Exiting a Node from the Maintenance Mode Using CLI on page 36 for more information.

IGMP Snooping

On an AHV host, when multicast traffic flows to a virtual switch the host floods the Mcast traffic to all the VMs on the specific VLAN. This mechanism is inefficient when many of the VMs on the VLAN do not need that multicast traffic. IGMP snooping allows the host to track which VMs on the VLAN need the multicast traffic and send the multicast traffic to only those VMs. For example, assume there are 50 VMs on VLAN 100 on virtual switch vs1 and only 25 VMs need to receive (hence, receiver VMs) the multicast traffic. Turn on IGMP snooping to help the AHV host track the 25 receiver VMs and deliver the multicast traffic to only the 25 receiver VMs instead of pushing the multicast traffic to all 50 VMs.

When IGMP snooping is enabled in a virtual switch on a VLAN, the ToR switch or router queries the VMs about the Mcast traffic that the VMs are interested in. When the switch receives a join request from a VM in response to the query, it adds the VM to the multicast list for that source entry as a receiver VM. When the switch sends a query, only the VMs that require the multicast traffic respond to the switch. The VMs that do not need the traffic do not respond at all. So, the switch does not add a VM to a multicast group or list unless it receives a response from that VM for the query.

Typically, in a multicast scenario, there is a source entity that casts the multicast traffic. This source may be another VM in this target cluster (that contains the target VMs that need to receive the multicast traffic) or another cluster connected to the target cluster. The host in the target cluster acts as the multicast router. Enable IGMP snooping in the virtual switch that hosts the VLAN connecting the VMs. You must also enable either the native Acropolis IGMP querier on the host or a separate third party querier that you install on the host. The native Acropolis IGMP querier sends IGMP v2 query packets to the VMs.

The IGMP Querier sends out queries periodically to keep the multicast groups or lists updated. The periodicity of the query is determined by the IGMP snooping timeout value that you must specify when you enable IGMP snooping. For example, if you have configured the IGMP snooping timeout as 30 seconds then the IGMP Querier sends out a query every 15 seconds.

When you enable IGMP snooping and are using the native Acropolis IGMP querier, you must configure the IGMP VLAN list. The IGMP VLAN list is a list of VLANs that the native IGMP Querier must send the query out to. This list value is a comma-separated list of the VLAN IDs that the query needs to be sent to. If you do not provide a list of VLANs, then the native IGMP Querier sends the query to all the VLANs in the switch.

When a VM needs to receive the multicast traffic from specific multicast source, configure the multicast application on the VM to listen to the queries received by the VM from the IGMP Querier. Also, configure The multicast application on the VM to respond to the relevant query, that is, the query for the specific multicast source. The response that the application sends

is logged by the virtual switch which then sends the multicast traffic to that VM instead of flooding it to all the VMs on the VLAN.

A multicast source always sends multicast traffic to a multicast group or list that is indicated by a multicast group IP address.

Enabling or Disabling IGMP Snooping

IGMP snooping helps you manage multicast traffic to specific VMs configured on a VLAN.

About this task

You can only enable IGMP snooping using aCLI.

Procedure

- Run the following command:

```
net.update_virtual_switch virtual-switch-name enable_igmp_snooping=true
 enable_igmp_querier=[true | false] igmp_query_vlan_list=VLAN IDs
 igmp_snooping_timeout=timeout
```

  Provide:

  - `virtual-switch-name`—The name of the virtual switch in which the VLANs are configured. For example, the name of the default virtual switch is `vs0`. Provide the name of the virtual switch exactly as it is configured.
  - `enable_igmp_snooping=[true | false]`—`true` to enable IGMP snooping. Provide `false` to disable IGMP snooping. The default setting is `false`.
  - `enable_igmp_querier=[true | false]`—`true` to enable the native IGMP querier. Provide `false` to disable the native IGMP querier. The default setting is `false`.
  - `igmp_query_vlan_list=VLAN IDs`—List of VLAN IDs mapped to the virtual switch for which IGMP querier is enabled. When it's not set or set as an empty list, querier is enabled for all VLANs of the virtual switch.
  - `igmp_snooping_timeout=timeout`—An integer indicating time in seconds. For example, you can provide `30` to indicate IGMP snooping timeout of 30 seconds.

    The default timeout is `300` seconds.

    You can set the timeout in the range of `15` - `3600` seconds.

What to do next

You can verify whether IGMP snooping is enabled or disabled by running the following command:

```
net.get_virtual_switch virtual-switch-name
```

The output of this command includes the following sample configuration:

```
igmp_config {
  enable_querier: True
  enable_snooping: True
 }
```

The above sample shows that IGMP snooping and the native acropolis IGMP querier are enabled.

Switch Port ANalyzer on AHV Hosts

Switch Port ANalyzer (SPAN) or port mirroring enables you to mirror traffic from interfaces of the AHV hosts to the VNIC of guest VMs. SPAN mirrors some or all packets from a set of source ports to a set of destination ports. You can mirror inbound, outbound, or bidirectional traffic on a set of source ports. You can then use the mirrored traffic for security analysis and gain visibility of traffic flowing through the set of source ports. SPAN is a useful tool for troubleshooting packets and can prove to be necessary for compliance reasons.

AHV supports the following types of source ports in a SPAN session:

1. A bond port that is already mapped to a Virtual Switch (VS) such as vs0, vs1, or any other VS you have created.
2. A non-bond port that is already mapped to a VS such as vs0, vs1, or any other VS you have created.
3. An uplink port that is not assigned to any VS or bridge on the host.

Important Considerations

Consider the following before you configure SPAN on AHV hosts:

- In this release, AHV supports mirroring of traffic only from physical interfaces.
- The SPAN destination VM or guest VM must be running on the same AHV host where the source ports are located.
- Delete the SPAN session before you delete the SPAN destination VM or VNIC. Otherwise, the state of the SPAN session is displayed as error.
- AHV does not support SPAN from a member of a bond port. For example, if you have mapped br0-up to bridge br0 with members eth0 and eth1, you cannot create a SPAN session with either eth0 or eth1 as the source port. You must use only br0-up as the source port.
- AHV supports different types of source ports in one session. For example, you can create a session with br0-up (bond port) and eth5 (single uplink port) on the same host as two different source ports in the same session. You can even have two different bond ports in the same session.
- One SPAN session supports up to two source and two destination ports.
- One host supports up to two SPAN sessions.
- You cannot create a SPAN session on an AHV host that is in the maintenance mode.
- If you move the uplink interface to another Virtual Switch, the SPAN session fails. Note that the system does not generate an alert in this situation.
- With TCP Segmentation Offload, multiple packets belonging to the same stream can be coalesced into a single one before being delivered to the SPAN destination VM. With TCP Segmentation Offload enabled, there can be a difference between the number of packets received on the uplink interface and packets forwarded to the SPAN destination VM (session packet count <= uplink interface packet count). However, the byte count at the SPAN destination VM is closer to the number at the uplink interface.

Configuring SPAN on an AHV Host

To configure SPAN on an AHV host, create a SPAN destination VNIC where you assign that VNIC to a guest VM (SPAN destination VM). After you create the VNIC, create a SPAN session specifying the source and destination ports between which you want to run the SPAN session.

Before you begin

Ensure that you have created the guest VM that you want to configure as the SPAN destination VM.

> Note: The SPAN destination VM must run on the same AHV host where the source ports are located. Therefore, Nutanix highly recommends that you create or modify the guest VM as an agent VM so that the VM is not migrated from the host.

Command and example for modifying a guest VM as an agent VM: (Recommended)

```
nutanix@cvm$ acli vm.update vm-name agent_vm=true
```

```
nutanix@cvm$ acli vm.update span-dest-VM agent_vm=true
```

In this example, **span-dest-VM** is the name of the guest VM that you are modifying as an agent VM.

About this task

Perform the following procedure to configure SPAN on an AHV host:

Procedure

1. Log on to a Controller VM in the cluster with SSH.
2. Determine the name and UUID of the guest VM that you want to configure as the SPAN destination VM.

   ```
   nutanix@cvm$ acli vm.list
   ```

   Example:

   ```
   nutanix@cvm$ acli vm.list
   VM name       VM UUID
   span-dest-VM  85abfdd5-7419-4f7c-bffa-8f961660e516
   ```

   In this example, **span-dest-VM** is the name and **85abfdd5-7419-4f7c-bffa-8f961660e516** is the UUID of the guest VM.

   > Note: If you delete the SPAN destination VM without deleting the SPAN session you create with this SPAN destination VM, the SPAN session `State` displays **kError**.

3. Create a SPAN destination VNIC for the guest VM.

   ```
   nutanix@cvm$ acli vm.nic_create vm-name type=kSpanDestinationNic
   ```

   Replace *`vm-name`* with the name of the guest VM on which you want to configure SPAN.

   > Note: Do not include any other parameter when you are creating a SPAN destination VNIC.

   Example:

   ```
   nutanix@cvm$ acli vm.nic_create span-dest-VM type=kSpanDestinationNic
   NicCreate: complete
   ```

   > Note: If you delete the SPAN destination VNIC without deleting the SPAN session you create with this SPAN destination VNIC, the SPAN session `State` displays **kError**.

4. Determine the MAC address of the VNIC.

```
nutanix@cvm$ acli vm.nic_get vm-name
```

Replace *vm-name* with the name of the guest VM to which you assigned the VNIC.

Example:

```
nutanix@cvm$ acli vm.nic_get span-dest-VM
x.x.x.x {
  connected: True
  ip_address: "x.x.x.x"
  mac_addr: "50:6b:8d:8b:2c:94"
  network_name: "mgmt"
  network_type: "kNativeNetwork"
  network_uuid: "c14b0092-877e-489b-a399-2749a60b3206"
  type: "kNormalNic"
  uuid: "9dd4f307-2506-4354-86a3-0b99abdeba6c"
  vlan_mode: "kAccess"
}
50:6b:8d:de:c6:44 {
  mac_addr: "50:6b:8d:de:c6:44"
  network_type: "kNativeNetwork"
  type: "kSpanDestinationNic"
  uuid: "b59e99bc-6bc7-4fab-ac35-543695c300d1"
}
```

Note the MAC address (value of **mac_addr**) of the VNIC whose **type** is set to **kSpanDestinationNic**.

5. Determine the UUID of the host whose traffic you want to monitor by using SPAN.

```
nutanix@cvm$ acli host.list
```

6. Create a SPAN session.

```
nutanix@cvm$ acli net.create_span_session span-session-name description="description-text"
 source_list=\{uuid=host-uuid,type=kHostNic,identifier=source-port-name,direction=traffic-
type} dest_list=\{uuid=vm-uuid,type=kVmNic,identifier=vnic-mac-address}
```

Replace the variables mentioned in the command for the following parameters with their appropriate values as follows:

- **span-session-name**: Replace *span-session-name* with a name for the session.
- **description** (Optional): Replace *description-text* with a description for the session. This is an optional parameter.

> Note:
>
> All *source_list* and *dest_list* parameters are mandatory inputs. The parameters do not have default values. Provide an appropriate value for each parameter.

Source list parameters:

- **uuid**: Replace *host-uuid* with the UUID of the host whose traffic you want to monitor by using SPAN. (determined in step 5).
- **type**: Specify **kHostNic** as the type. Only the kHostNic type is supported in this release.
- **identifier**: Replace *source-port-name* with name of the source port whose traffic you want to mirror. For example, br0-up, eth0, or eth1.
- **direction**: Replace *traffic-type* with **kIngress** if you want to mirror inbound traffic, **kEgress** for outbound traffic, or **kBiDir** for bidirectional traffic.

Destination list parameters:

- **uuid**: Replace *vm-uuid* with the UUID of the guest VM that you want to configure as the SPAN destination VM. (determined in step 2).
- **type**: Specify **kVmNic** as the type. Only the **kVmNic** type is supported in this release.
- **identifier**: Replace *vnic-mac-address* with the MAC address of the destination port where you want to mirror the traffic (determined in step 4).

> Note: The syntax for **source_list** and **dest_list** is as follows:
>
> source_list/dest_list=[{key1=value1,key2=value2,..}]
>
> Each pair of curly brackets includes the details of one source or destination port with a comma-separated list of the key-value pairs. There must not be any space between two key-value pairs.
>
> One SPAN session supports up to two source and two destination ports. If you want to include an extra port, separate the curly brackets with a semicolon (no space) and list the key-value pairs of the second port in the other curly bracket.

Example:

```
nutanix@cvm$ acli net.create_span_session span1 description="span session 1"
 source_list=\{uuid=492a2bda-ffc0-486a-8bc0-8ae929471714,type=kHostNic,identifier=br0-
up,direction=kBiDir} dest_list=\{uuid=85abfdd5-7419-4f7c-
bffa-8f961660e516,type=kVmNic,identifier=50:6b:8d:de:c6:44}
SpanCreate: complete
```

7. Display the list of all SPAN sessions running on a host.

```
nutanix@cvm$ acli net.list_span_session
```

Example:

```
nutanix@cvm$ acli net.list_span_session
Name   UUID                                  State
span1  69252eb5-8047-4e3a-8adc-91664a7104af  kActive
```

Possible values for `State` are:

- **kActive**: Denotes that the SPAN session is active.
- **kError**: denotes that there is an error and the configuration is not working. For example, if there are two surces and one source is down, the `State` of the session is diplayed as **kError**.

8. Display the details of a SPAN session.

```
nutanix@cvm$ acli net.get_span_session span-session-name
```

Replace span-session-name with the name of the SPAN session whose details you want to view.

Example:

```
nutanix@cvm$ acli net.get_span_session span1
span1 {
  config {
    datapath_name: "s6925"
    description: "span session 1"
    destination_list {
      nic_type: "kVmNic"
      port_identifier: "50:6b:8d:de:c6:44"
      uuid: "85abfdd5-7419-4f7c-bffa-8f961660e516"
    }
    name: "span1"
    session_uuid: "69252eb5-8047-4e3a-8adc-91664a7104af"
    source_list {
      direction: "kBiDir"
      nic_type: "kHostNic"
      port_identifier: "br0-up"
      uuid: "492a2bda-ffc0-486a-8bc0-8ae929471714"
    }
  }
  stats {
    name: "span1"
    session_uuid: "69252eb5-8047-4e3a-8adc-91664a7104af"
    state: "kActive"
    stats_list {
      tx_byte_cnt: 67498
      tx_pkt_cnt: 436
    }
  }
}
```

Note the value of the **datapath_name** field in the SPAN session configuration, which is a unique key that identifies the SPAN session. You might need the unique key to correctly identify the SPAN session for troubleshooting reasons.

Updating a SPAN Session

You can update any of the details of a SPAN session. When you are updating a SPAN session, specify the values of the parameters you want to update and then specify the rest of the

parameters again as you specified them when you created the SPAN session. For example, if you want to change only the name and description, specify the updated name and description and then include the complete details of the source and destination ports again even though you are not updating those details.

About this task

Perform the following procedure to update a SPAN session:

Procedure

1. Log on to a Controller VM in the cluster with SSH.
2. Update the SPAN session.

```
nutanix@cvm$ acli net.update_span_session span-session-name description="description-text"
 source_list=\{uuid=host-uuid,type=kHostNic,identifier=source-port-name,direction=traffic-
type} dest_list=\{uuid=vm-UUID,type=kVmNic,identifier=vNIC-mac-address}
```

The update command includes the same parameters as the create command. See Configuring SPAN on an AHV Host on page 74 for more information.

Example:

```
nutanix@cvm$ acli net.update_span_session span1 name=span_br0_to_span_dest
 description="span from br0-up to span-dest VM" source_list=\{uuid=492a2bda-
ffc0-486a-8bc0-8ae929471714,type=kHostNic,identifier=br0-up,direction=kBiDir} dest_list=
\{uuid=85abfdd5-7419-4f7c-bffa-8f961660e516,type=kVmNic,identifier=50:6b:8d:de:c6:44}
SpanUpdate: complete

nutanix@cvm$ acli net.list_span_session
Name                   UUID                                   State
span_br0_to_span_dest  69252eb5-8047-4e3a-8adc-91664a7104af  kActive

nutanix@cvm$ acli net.get_span_session span_br0_to_span_dest
span_br0_to_span_dest {
  config {
    datapath_name: "s6925"
    description: "span from br0-up to span-dest VM"
    destination_list {
      nic_type: "kVmNic"
      port_identifier: "50:6b:8d:de:c6:44"
      uuid: "85abfdd5-7419-4f7c-bffa-8f961660e516"
    }
    name: "span_br0_to_span_dest"
    session_uuid: "69252eb5-8047-4e3a-8adc-91664a7104af"
    source_list {
      direction: "kBiDir"
      nic_type: "kHostNic"
      port_identifier: "br0-up"
      uuid: "492a2bda-ffc0-486a-8bc0-8ae929471714"
    }
  }
  stats {
    name: "span_br0_to_span_dest"
    session_uuid: "69252eb5-8047-4e3a-8adc-91664a7104af"
    state: "kActive"
    stats_list {
      tx_byte_cnt: 805705
      tx_pkt_cnt: 4792
    }
  }
```

```
}
```

In this example, only the name and description were updated. However, complete details of the source and destation ports were included in the command again.

If you want to change the name of a SPAN session, specify the existing name first and then include the new name by using the “name=” parameter as shown in this example.

Deleting a SPAN Session

Delete the SPAN session if you want to disable SPAN on an AHV host. Nutanix recommends that you delete the SPAN session associated with a SPAN destination VM or SPAN destination VNIC.

About this task

Perform the following procedure to delete a SPAN session:

Procedure

1. Log on to a Controller VM in the cluster with SSH.
2. Delete the SPAN session.

   ```
   nutanix@cvm$ acli net.delete_span_session span-session-name
   ```

   Replace *span-session-name* with the name of the SPAN session you want to delete.

Enabling RSS Virtio-Net Multi-Queue by Increasing the Number of VNIC Queues

Multi-Queue in VirtIO-net enables you to improve network performance for network I/O-intensive guest VMs or applications running on AHV hosts.

About this task

You can enable VirtIO-net multi-queue by increasing the number of VNIC queues. If an application uses many distinct streams of traffic, Receive Side Scaling (RSS) can distribute the streams across multiple VNIC DMA rings. This increases the amount of RX buffer space by the number of VNIC queues (N). Also, most guest operating systems pin each ring to a particular vCPU, handling the interrupts and ring-walking on that vCPU, by that means achieving N-way parallelism in RX processing. However, if you increase the number of queues beyond the number of vCPUs, you cannot achieve extra parallelism.

Following workloads have the greatest performance benefit of VirtIO-net multi-queue:

- VMs where traffic packets are relatively large
- VMs with many concurrent connections
- VMs with network traffic moving:
  - Among VMs on the same host
  - Among VMs across hosts
  - From VMs to the hosts
  - From VMs to an external system
- VMs with high VNIC RX packet drop rate if CPU contention is not the cause

You can increase the number of queues of the AHV VM VNIC to allow the guest OS to use multi-queue VirtIO-net on guest VMs with intensive network I/O. Multi-Queue VirtIO-net scales

the network performance by transferring packets through more than one Tx/Rx queue pair at a time as the number of vCPUs increases.

Nutanix recommends that you be conservative when increasing the number of queues. Do not set the number of queues larger than the total number of vCPUs assigned to a VM. Packet reordering and TCP retransmissions increase if the number of queues is larger than the number vCPUs assigned to a VM. For this reason, start by increasing the queue size to 2. The default queue size is 1. After making this change, monitor the guest VM and network performance. Before you increase the queue size further, verify that the vCPU usage has not dramatically or unreasonably increased.

Perform the following steps to make more VNIC queues available to a guest VM. See your guest OS documentation to verify if you must perform extra steps on the guest OS to apply the additional VNIC queues.

> Note: You must shut down the guest VM to change the number of queues. Therefore, make this change during a planned maintenance window. The VNIC status might change from Up->Down->Up or a restart of the guest OS might be required to finalize the settings depending on the guest OS implementation requirements.

Procedure

1. (Optional) Nutanix recommends that you ensure the following:
   a. AHV and AOS are running the latest version.
   b. AHV guest VMs are running the latest version of the Nutanix VirtIO driver package.
      For RSS support, ensure you are running Nutanix VirtIO 1.1.6 or later. See Nutanix VirtIO for Windows on page 110 for more information about Nutanix VirtIO.
2. Determine the exact name of the guest VM for which you want to change the number of VNIC queues.

   ```
   nutanix@cvm$ acli vm.list
   ```

   An output similar to the following is displayed:

   ```
   nutanix@cvm$ acli vm.list
   VM name          VM UUID
   ExampleVM1       a91a683a-4440-45d9-8dbe-xxxxxxxxxxxx
   ExampleVM2       fda89db5-4695-4055-a3d4-xxxxxxxxxxxx
   ...
   ```

3. Determine the MAC address of the VNIC and confirm the current number of VNIC queues.

   ```
   nutanix@cvm$ acli vm.nic_get VM-name
   ```

   Replace *VM-name* with the name of the VM.

   An output similar to the following is displayed:

   ```
   nutanix@cvm$ acli vm.nic_get VM-name
   ...
   mac_addr: "50:6b:8d:2f:zz:zz"
   ...
   (queues: 2)    <- If there is no output of 'queues', the setting is default (1 queue).
   ```

   > Note: AOS defines queues as the maximum number of Tx/Rx queue pairs (default is 1).

4. Check the number of vCPUs assigned to the VM.

```
nutanix@cvm$ acli vm.get VM-name | grep num_vcpus
```

An output similar to the following is displayed:

```
nutanix@cvm$ acli vm.get VM-name | grep num_vcpus
num_vcpus: 1
```

5. Shut down the guest VM.

```
nutanix@cvm$ acli vm.shutdown VM-name
```

Replace *VM-name* with the name of the VM.

6. Increase the number of VNIC queues.

```
nutanix@cvm$acli vm.nic_update VM-name vNIC-MAC-address queues=N
```

Replace *VM-name* with the name of the guest VM, *vNIC-MAC-address* with the MAC address of the VNIC, and *N* with the number of queues.

> Note: N must be less than or equal to the vCPUs assigned to the guest VM.

7. Start the guest VM.

```
nutanix@cvm$ acli vm.on VM-name
```

Replace *VM-name* with the name of the VM.

8. Confirm in the guest OS documentation if any additional steps are required to enable multi-queue in VirtIO-net.

> Note: Microsoft Windows has RSS enabled by default.

For example, for RHEL and CentOS VMs, do the following:

a. Log on to the guest VM.

b. Confirm if irqbalance.service is active or not.

```
uservm# systemctl status irqbalance.service
```

An output similar to the following is displayed:

```
irqbalance.service - irqbalance daemon
   Loaded: loaded (/usr/lib/systemd/system/irqbalance.service; enabled; vendor preset:
 enabled)
```

```
   Active: active (running) since Tue 2020-04-07 10:28:29 AEST; Ns ago
```

c. Start irqbalance.service if it is not active.

> Note: It is active by default on CentOS VMs. You might have to start it on RHEL VMs.

```
uservm# systemctl start irqbalance.service
```

d. Run the following command:

```
uservm$ ethtool -L ethX combined M
```

Replace M with the number of VNIC queues.

Note the following caveat from the RHEL 7 *Virtualization Tuning and Optimization Guide : 5.4. NETWORK TUNING TECHNIQUES* document:

"Currently, setting up a multi-queue virtio-net connection can have a negative effect on the performance of outgoing traffic. Specifically, this may occur when sending packets under 1,500 bytes over the Transmission Control Protocol (TCP) stream."

9. Monitor the VM performance to make sure that the expected network performance increase is observed and that the guest VM vCPU usage is not dramatically increased to impact the application on the guest VM.

   For assistance with the steps described in this document, or if these steps do not resolve your guest VM network performance issues, contact Nutanix Support.

Changing the IP Address of an AHV Host

Change the IP address, netmask, or gateway of an AHV host.

Before you begin

Perform the following tasks before you change the IP address, netmask, or gateway of an AHV host:

> CAUTION: All Controller VMs and hypervisor hosts must be on the same subnet.

> Warning: Ensure that you perform the steps in the exact order as indicated in this document.

1. Verify the cluster health by following the instructions in KB-2852.

   Do not proceed if the cluster cannot tolerate failure of at least one node.
2. Put the AHV host into the maintenance mode.

   See Putting a Node into Maintenance Mode using CLI on page 34 for instructions about how to put a node into maintenance mode.

About this task

Perform the following procedure to change the IP address, netmask, or gateway of an AHV host.

Procedure

1. Edit the settings of port br0, which is the internal port on the default bridge br0.

   a. Log on to the host console as root.

      You can access the hypervisor host console either through IPMI or by attaching a keyboard and monitor to the node.

   b. Open the network interface configuration file for port br0 in a text editor.

      ```
      root@ahv# vi /etc/sysconfig/network-scripts/ifcfg-br0
      ```

   c. Update entries for host IP address, netmask, and gateway.

      The block of configuration information that includes these entries is similar to the following:

      ```
      ONBOOT="yes"
      NM_CONTROLLED="no"
      PERSISTENT_DHCLIENT=1
      NETMASK="subnet_mask"
      IPADDR="host_ip_addr"
      DEVICE="br0"
      TYPE="ethernet"
      GATEWAY="gateway_ip_addr"
      BOOTPROTO="none"
      ```

      - Replace `host_ip_addr` with the IP address for the hypervisor host.
      - Replace `subnet_mask` with the subnet mask for host_ip_addr.
      - Replace `gateway_ip_addr` with the gateway address for host_ip_addr.

   d. Save your changes.

   e. Restart network services.

      ```
      /etc/init.d/network restart
      ```

   f. Assign the host to a VLAN. For information about how to add a host to a VLAN, see Assigning an AHV Host to a VLAN on page 70.

   g. Verify network connectivity by pinging the gateway, other CVMs, and AHV hosts.

2. Log on to the Controller VM that is running on the AHV host whose IP address you changed and restart genesis.

   ```
   nutanix@cvm$ genesis restart
   ```

   If the restart is successful, output similar to the following is displayed:

   ```
   Stopping Genesis pids [1933, 30217, 30218, 30219, 30241]
   Genesis started on pids [30378, 30379, 30380, 30381, 30403]
   ```

   See Controller VM Access on page 12 for information about how to log on to a Controller VM.

   Genesis takes a few minutes to restart.

3. Verify if the IP address of the hypervisor host has changed. Run the following nCLI command from any CVM other than the one in the maintenance mode.

   ```
   nutanix@cvm$ ncli host list
   ```

   An output similar to the following is displayed:

   ```
   nutanix@cvm$ ncli host list
   ```

```
    Id                          : aaaaaaaa-bbbb-cccc-dddd-eeeeeeeeeeee::1234
    Uuid                        : ffffffff-gggg-hhhh-iiii-jjjjjjjjjjjj
    Name                        : XXXXXXXXXXX-X
    IPMI Address                : X.X.Z.3
    Controller VM Address       : X.X.X.1
    Hypervisor Address          : X.X.Y.4 <- New IP Address
...
```

4. Stop the Acropolis service on all the CVMs.

   a. Stop the Acropolis service on each CVM in the cluster.

   ```
   nutanix@cvm$ genesis stop acropolis
   ```

   > Note: You cannot manage your guest VMs after the Acropolis service is stopped.

   b. Verify if the Acropolis service is DOWN on all the CVMs, except the one in the maintenance mode.

   ```
   nutanix@cvm$ cluster status | grep -v UP
   ```

   An output similar to the following is displayed:

   ```
   nutanix@cvm$ cluster status | grep -v UP

   2019-09-04 14:43:18 INFO zookeeper_session.py:143 cluster is attempting to connect to
    Zookeeper

   2019-09-04 14:43:18 INFO cluster:2774 Executing action status on SVMs X.X.X.1, X.X.X.2,
    X.X.X.3

   The state of the cluster: start

   Lockdown mode: Disabled
           CVM: X.X.X.1 Up
                                Acropolis DOWN       []
           CVM: X.X.X.2 Up, ZeusLeader
                                Acropolis DOWN       []
           CVM: X.X.X.3 Maintenance
   ```

5. From any CVM in the cluster, start the Acropolis service.

   ```
   nutanix@cvm$ cluster start
   ```

6. Verify if all processes on all the CVMs, except the one in the maintenance mode, are in the UP state.

   ```
   nutanix@cvm$ cluster status | grep -v UP
   ```

7. Exit the AHV host and the CVM from the maintenance mode.

   See Exiting a Node from the Maintenance Mode Using CLI on page 36 for more information.

4

VIRTUAL MACHINE MANAGEMENT

The following topics describe various aspects of virtual machine management in an AHV cluster.

Supported Guest VM Types for AHV

The compatibility matrix available on the Nutanix Support portal includes the latest supported AHV guest VM OSes.

AHV Configuration Maximums

The Nutanix configuration maximums available on the Nutanix support portal includes all the latest configuration limits applicable to AHV. Select the appropriate AHV version to view version specific information.

Creating a VM (AHV)

In AHV clusters, you can create a new virtual machine (VM) through the Prism Element web console.

About this task

When creating a VM, you can configure all of its components, such as number of vCPUs and memory, but you cannot attach a volume group to the VM. Attaching a volume group is possible only when you are modifying a VM.

To create a VM, do the following:

Procedure

1. In the VM dashboard, click the **Create VM** button.

   Note: This option does not appear in clusters that do not support this feature.

   The Create VM dialog box appears.

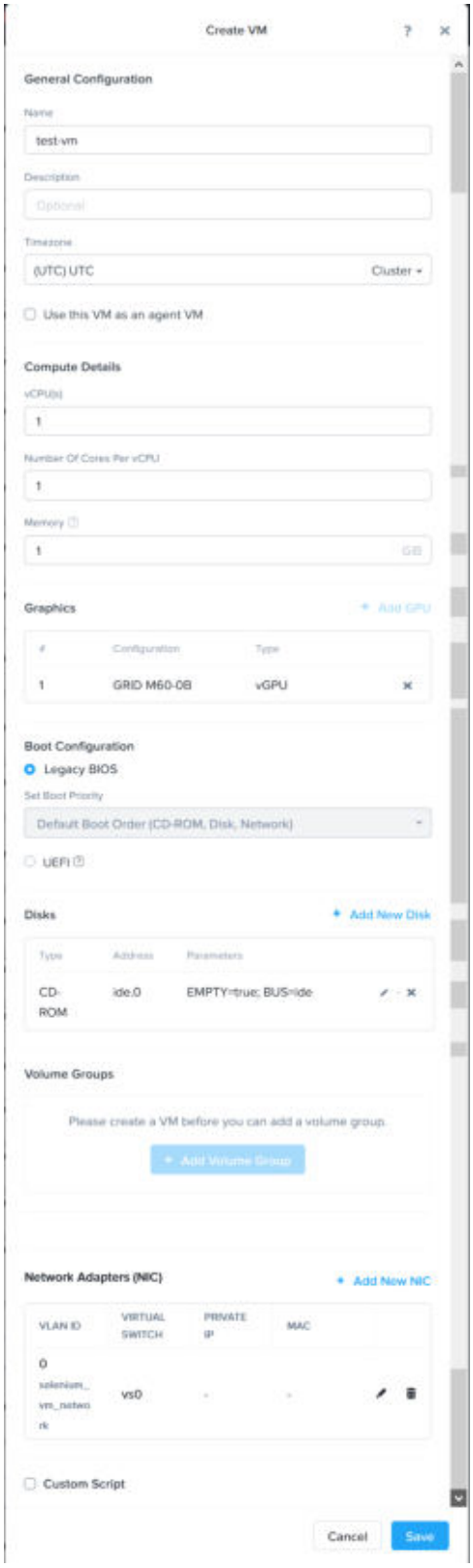


Figure 25: Create VM Dialog Box

2. Do the following in the indicated fields:

   a. **Name**: Enter a name for the VM.

   b. **Description** (optional): Enter a description for the VM.

   c. **Timezone**: Select the timezone that you want the VM to use. If you are creating a Linux VM, select **(UTC) UTC**.

      > Note:
      >
      > The RTC of Linux VMs must be in UTC, so select the UTC timezone if you are creating a Linux VM.
      >
      > Windows VMs preserve the RTC in the local timezone, so set up the Windows VM with the hardware clock pointing to the desired timezone.

   d. **Use this VM as an agent VM**: Select this option to make this VM as an agent VM.

      You can use this option for the VMs that must be powered on before the rest of the VMs (for example, to provide network functions before the rest of the VMs are powered on on the host) and must be powered off after the rest of the VMs are powered off (for example, during maintenance mode operations). Agent VMs are never migrated to any other host in the cluster. If an HA event occurs or the host is put in maintenance mode, agent VMs are powered off and are powered on on the same host once that host comes back to a normal state.

      If an agent VM is powered off, you can manually start that agent VM on another host and the agent VM now permanently resides on the new host. The agent VM is never migrated back to the original host. Note that you cannot migrate an agent VM to another host while the agent VM is powered on.

   e. **vCPU(s)**: Enter the number of virtual CPUs to allocate to this VM.

   f. **Number of Cores per vCPU**: Enter the number of cores assigned to each virtual CPU.

   g. **Memory**: Enter the amount of memory (in GiB) to allocate to this VM.

3. (For GPU-enabled AHV clusters only) To configure GPU access, click **Add GPU** in the **Graphics** section, and then do the following in the Add GPU dialog box:

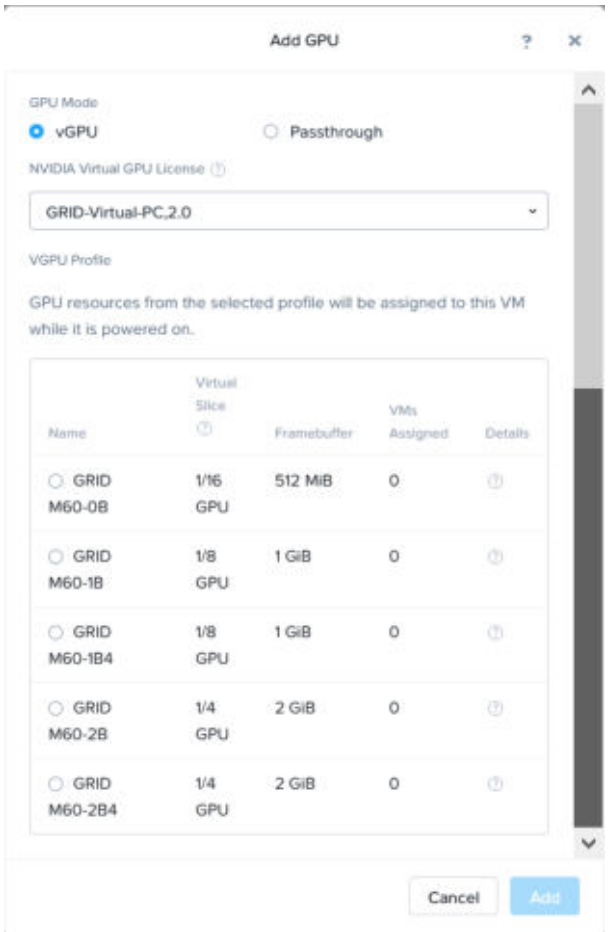


Figure 26: Add GPU Dialog Box

See *GPU and vGPU Support* in *AHV Administration Guide* for more information.

a. To configure GPU pass-through, in **GPU Mode**, click **Passthrough**, select the GPU that you want to allocate, and then click **Add**.

   If you want to allocate additional GPUs to the VM, repeat the procedure as many times as you need to. Make sure that all the allocated pass-through GPUs are on the same host. If all specified GPUs of the type that you want to allocate are in use, you can

proceed to allocate the GPU to the VM, but you cannot power on the VM until a VM that is using the specified GPU type is powered off.

See *GPU and vGPU Support* in *AHV Administration Guide* for more information.

b. To configure virtual GPU access, in **GPU Mode**, click **virtual GPU**, select a GRID license, and then select a virtual GPU profile from the list.

> Note: This option is available only if you have installed the GRID host driver on the GPU hosts in the cluster.
>
> See the NVIDIA Grid Host Driver for Nutanix AHV Installation Guide for the NVIDIA GRID host driver installation instructions.

You can assign multiple virtual GPU to a VM. A vGPU is assigned to the VM only if a vGPU is available when the VM is starting up.

Before you add multiple vGPUs to the VM, see *Multiple Virtual GPU Support* and *Restrictions for Multiple vGPU Support* in the *AHV Admin Guide*.

> Note:
>
> Multiple vGPUs are supported on the same VM only if you select the highest vGPU profile type.

After you add the first vGPU, to add multiple vGPUs, see *Adding Multiple vGPUs to the Same VM* in the *AHV Admin Guide*.

4. Select one of the following firmware to boot the VM.

» **Legacy BIOS**: Select legacy BIOS to boot the VM with legacy BIOS firmware.

» **UEFI**: Select UEFI to boot the VM with UEFI firmware. UEFI firmware supports larger hard drives, faster boot time, and provides more security features. For more information about UEFI firmware, see *UEFI Support for VM* section in the *AHV Administration Guide*.

» Secure Boot is supported with AOS 5.16. The current support to Secure Boot is limited to the aCLI. For more information about Secure Boot, see *Secure Boot Support for VMs* section in the *AHV Administration Guide*. To enable Secure Boot, do the following:

- Select UEFI.
- Power-off the VM.
- Log on to the aCLI and update the VM to enable Secure Boot. For more information, see *Updating a VM to Enable Secure Boot* in the *AHV Administration Guide*.

5. To attach a disk to the VM, click the **Add New Disk** button.
   The Add Disk dialog box appears.

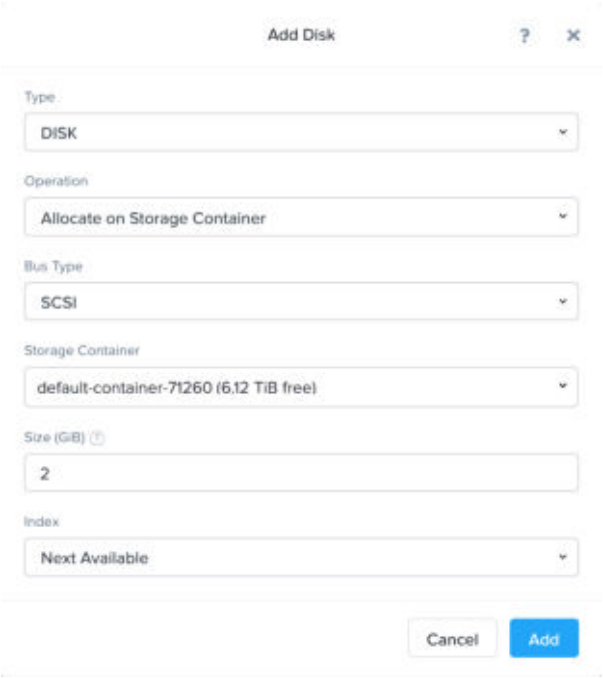


Figure 27: Add Disk Dialog Box

Do the following in the indicated fields:

a. **Type**: Select the type of storage device, **DISK** or **CD-ROM**, from the drop-down list.
   The following fields and options vary depending on whether you choose **DISK** or **CD-ROM**.

b. **Operation**: Specify the device contents from the drop-down list.

   - Select **Clone from ADSF file** to copy any file from the cluster that can be used as an image onto the disk.
   - Select **Empty CD-ROM** to create a blank CD-ROM device. (This option appears only when **CD-ROM** is selected in the previous field.) A CD-ROM device is needed when you intend to provide a system image from CD-ROM.
   - Select **Allocate on Storage Container** to allocate space without specifying an image. (This option appears only when **DISK** is selected in the previous field.) Selecting this option means you are allocating space only. You have to provide a system image later from a CD-ROM or other source.
   - Select **Clone from Image Service** to copy an image that you have imported by using image service feature onto the disk. For more information on the Image Service

feature, see the *Configuring Images* section of *Prism Web Console Guide* and *Image Management* section in the *Prism Central Guide*.

c. **Bus Type**: Select the bus type from the pull-down list. The choices are **IDE**, **SCSI**, or **SATA**.

   The options displayed in the **Bus Type** drop-down list varies based on the storage device Type selected in Step a.

   - For device **Disk**, select from **SCSI**, **SATA**, **PCI**, or **IDE** bus type.
   - For device **CD-ROM**, you can select either **IDE**,or **SATA** bus type.

   > Note: SCSI bus is the preferred bus type and it is used in most cases. Ensure you have installed the VirtIO drivers in the guest OS.

   > CAUTION: Use SATA, PCI, IDE for compatibility purpose when the guest OS does not have VirtIO drivers to support SCSI devices. This may have performance implications.

   > Note: For AHV 5.16 and later, you cannot use an IDE device if **Secured Boot** is enabled for **UEFI Mode** boot configuration.

d. **ADSF Path**: Enter the path to the desired system image.

   This field appears only when **Clone from ADSF file** is selected. It specifies the image to copy. Enter the path name as `/storage_container_name/iso_name.iso`. For example to clone an image from `myos.iso` in a storage container named `crt1`, enter `/crt1/myos.iso`. When a user types the storage container name (`/storage_container_name/`), a list appears of the ISO files in that storage container (assuming one or more ISO files had previously been copied to that storage container).

e. **Image**: Select the image that you have created by using the image service feature.

   This field appears only when **Clone from Image Service** is selected. It specifies the image to copy.

f. **Storage Container**: Select the storage container to use from the drop-down list.

   This field appears only when **Allocate on Storage Container** is selected. The list includes all storage containers created for this cluster.

g. **Size**: Enter the disk size in GiB.

h. **Index**: Displays **Next Available** by default.

i. When all the field entries are correct, click the **Add** button to attach the disk to the VM and return to the Create VM dialog box.

j. Repeat this step to attach additional devices to the VM.

6. To create a network interface for the VM, click the **Add New NIC** button.
   The Create NIC dialog box appears.

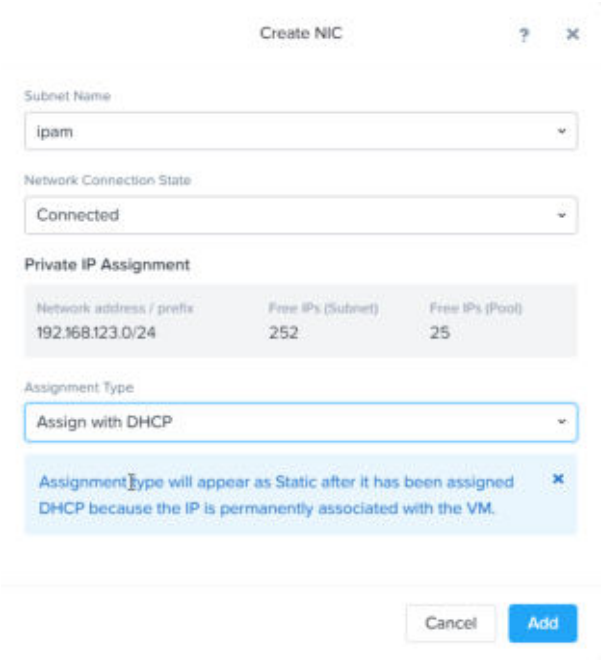


Figure 28: Create NIC Dialog Box

Do the following in the indicated fields:

a. **Subnet Name**: Select the target virtual LAN from the drop-down list.
   The list includes all defined networks (see the *Network Configuration For VM Interfaces* in the *Prism Web Console Guide*).

   > Note: Selecting IPAM enabled subnet from the drop-down list displays the **Private IP Assignment** information that provides information about the number of free IP addresses available in the subnet and in the IP pool.

b. **Network Connection State**: Select the state for the network that you want it to operate in after VM creation. The options are `Connected` or `Disconnected`.

c. **Private IP Assignment**: This is a read-only field and displays the following:

   - **Network Address/Prefix**: The network IP address and prefix.
   - **Free IPs (Subnet)**: The number of free IP addresses in the subnet.
   - **Free IPs (Pool)**: The number of free IP addresses available in the IP pools for the subnet.

d. **Assignment Type**: This is for IPAM enabled network. Select **Assign with DHCP** to assign IP address automatically to the VM using DHCP. For more information, see *IP Address Management* in the *AHV Administration Guide*.

e. When all the field entries are correct, click the **Add** button to create a network interface for the VM and return to the Create VM dialog box.

f. Repeat this step to create additional network interfaces for the VM.

> Note: Nutanix guarantees a unique VM MAC address in a cluster. You can come across scenarios where two VM in different clusters can have the same MAC address.

> Note: Acropolis leader generates MAC address for the VM on AHV. The first 24 bits of the MAC address is set to 50-6b-8d (`0101 0000 0110 1101 1000 1101`) and are reserved by Nutanix, the 25th bit is set to `1` (reserved by Acropolis leader), the 26th bit to 48th bits are auto generated random numbers.

7. To configure affinity policy for this VM, click **Set Affinity**.

   The Set VM Host Affinity dialog box appears.

   a. Select the host or hosts on which you want configure the affinity for this VM.

   b. Click **Save**.

      The selected host or hosts are listed. This configuration is permanent. The VM will not be moved from this host or hosts even in case of HA event and will take effect once the VM starts.

8. To customize the VM by using Cloud-init (for Linux VMs) or Sysprep (for Windows VMs), select the **Custom Script** check box.
   Fields required for configuring Cloud-init and Sysprep, such as options for specifying a configuration script or answer file and text boxes for specifying paths to required files, appear below the check box.

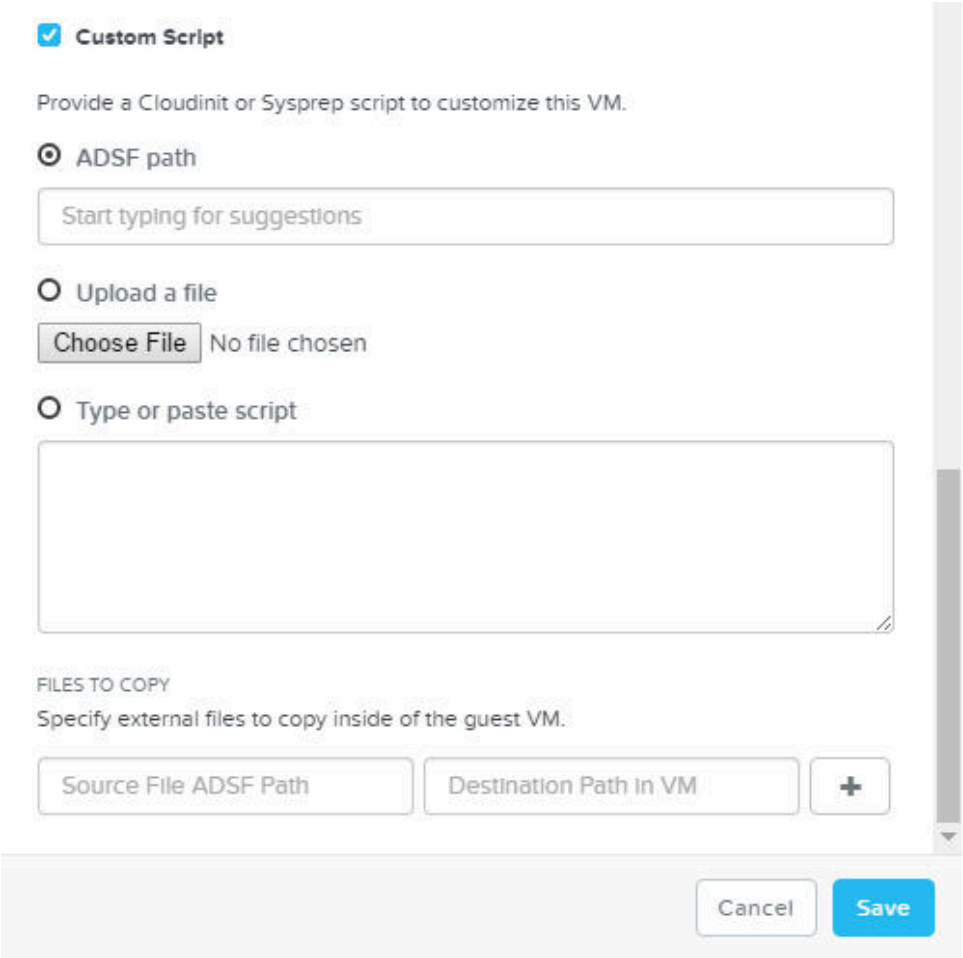


Figure 29: Create VM Dialog Box (custom script fields)

9. To specify a user data file (Linux VMs) or answer file (Windows VMs) for unattended provisioning, do one of the following:

   » If you uploaded the file to a storage container on the cluster, click **ADSF path**, and then enter the path to the file.

     Enter the ADSF prefix (`adsf://`) followed by the absolute path to the file. For example, if the user data is in `/home/my_dir/cloud.cfg`, enter `adsf:///home/my_dir/cloud.cfg`. Note the use of three slashes.

   » If the file is available on your local computer, click **Upload a file**, click **Choose File**, and then upload the file.

   » If you want to create or paste the contents of the file, click **Type or paste script**, and then use the text box that is provided.

10. To copy one or more files to a location on the VM (Linux VMs) or to a location in the ISO file (Windows VMs) during initialization, do the following:

    a. In **Source File ADSF Path**, enter the absolute path to the file.

    b. In **Destination Path in VM**, enter the absolute path to the target directory and the file name.

       For example, if the source file entry is `/home/my_dir/myfile.txt` then the entry for the **Destination Path in VM** should be `/<directory_name>/copy_desitation>` i.e. `/mnt/myfile.txt`.

    c. To add another file or directory, click the button beside the destination path field. In the new row that appears, specify the source and target details.

11. When all the field entries are correct, click the **Save** button to create the VM and close the Create VM dialog box.

    The new VM appears in the VM table view.

Managing a VM (AHV)

You can use the web console to manage virtual machines (VMs) in Acropolis managed clusters.

About this task

After creating a VM (see Creating a VM (AHV) on page 86), you can use the web console to start or shut down the VM, launch a console window, update the VM configuration, take a snapshot, attach a volume group, migrate the VM, clone the VM, or delete the VM.

> Note: Your available options depend on the VM status, type, and permissions. Unavailable options are grayed out.

To accomplish one or more of these tasks, do the following:

Procedure

1. In the VM dashboard, click the **Table** view.

2. Select the target VM in the table (top section of screen).

   The Summary line (middle of screen) displays the VM name with a set of relevant action links on the right. You can also right-click on a VM to select a relevant action.

   The possible actions are **Manage Guest Tools**, **Launch Console**, **Power on** (or **Power off**), **Take Snapshot**, **Migrate**, **Clone**, **Update**, and **Delete**.

   > Note: VM pause and resume feature is not supported on AHV.

   The following steps describe how to perform each action.

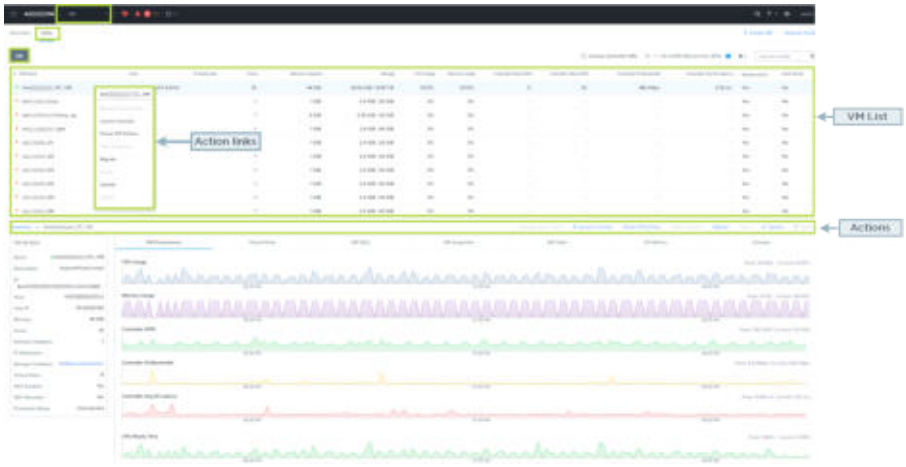


   Figure 30: VM Action Links

3. To manage guest tools as follows, click **Manage Guest Tools**.

   You can also enable NGT applications (self-service restore, Volume Snapshot Service and application-consistent snapshots) also as part of manage guest tools.

   a. Select **Enable Nutanix Guest Tools** check box to enable NGT on the selected VM.

   b. Select **Mount Nutanix Guest Tools** to mount NGT on the selected VM.

      Ensure that VM must have at least one empty IDE CD-ROM slot to attach the ISO.

      The VM is registered with the NGT service. NGT is enabled and mounted on the selected virtual machine. A CD with volume label NUTANIX_TOOLS gets attached to the VM.

   c. To enable self-service restore feature for Windows VMs, click **Self Service Restore (SSR)** check box.

      The Self-Service Restore feature is enabled of the VM. The guest VM administrator can restore the desired file or files from the VM. For more information on self-service restore

feature, see Self-Service Restore in the guide titled *Data Protection and Recovery with Prism Element*.

d. After you select **Enable Nutanix Guest Tools** check box the VSS snapshot feature is enabled by default.

   After this feature is enabled, Nutanix native in-guest VmQuiesced Snapshot Service (VSS) agent takes snapshots for VMs that support VSS.

   > Note:
   >
   > The AHV VM snapshots are not application consistent. The AHV snapshots are taken from the **VM** entity menu by selecting a VM and clicking **Take Snapshot**.
   >
   > The application consistent snapshots feature is available with Protection Domain based snapshots and Recovery Points in Prism Central. For more information, see *Implementation Guidelines for Asynchronous Disaster Recovery* section in the guide titled *Data Protection and Recovery with Prism Element*.

e. Click **Submit**.

   The VM is registered with the NGT service. NGT is enabled and mounted on the selected virtual machine. A CD with volume label NUTANIX_TOOLS gets attached to the VM.

   > Note:
   >
   > - If you clone a VM, by default NGT is not enabled on the cloned VM. If the cloned VM is powered off, enable NGT from the UI and power on the VM. If cloned VM is powered on, enable NGT from the UI and restart the nutanix guest agent service.
   > - If you want to enable NGT on multiple VMs simultaneously, see *Enabling NGT and Mounting the NGT Installer Simultaneously on Multiple Cloned VMs* in the Prism Web Console Guide.

   If you eject the CD, you can mount the CD back again by logging into the Controller VM and running the following nCLI command.

   ```
   nutanix@cvm$ ncli ngt mount vm-id=virtual_machine_id
   ```

   For example, to mount the NGT on the VM with VM_ID=00051a34-066f-72ed-0000-000000005400::38dc7bf2-a345-4e52-9af6-c1601e759987, type the following command.

   ```
   nutanix@cvm$ ncli ngt mount vm-id=00051a34-066f-72ed-0000-000000005400::38dc7bf2-
   a345-4e52-9af6-
   c1601e759987
   ```

4. To launch a console window, click the **Launch Console** action link.

   This opens a Virtual Network Computing (VNC) client and displays the console in a new tab or window. This option is available only when the VM is powered on. The console window includes four menu options (top right):

   - Clicking the **Mount ISO** button displays the following window that allows you to mount an ISO image to the VM. To mount an image, select the desired image and CD-ROM drive from the drop-down lists and then click the **Mount** button.

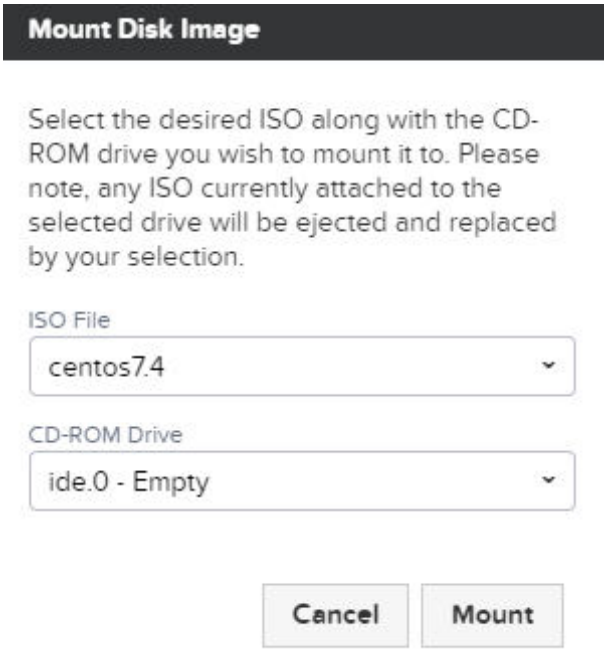


Figure 31: Mount Disk Image Window

> Note: See *Add New Disk* in Creating a VM (AHV) on page 86 section for information about how to select CD-ROM as the storage device when you intent to provide a system image from CD-ROM.

   - Clicking the **C-A-D** icon button sends a **CtrlAltDel** command to the VM.
   - Clicking the camera icon button takes a screenshot of the console window.
   - Clicking the power icon button allows you to power on/off the VM. These are the same options that you can access from the **Power On Actions** or **Power Off Actions** action link below the VM table (see next step).

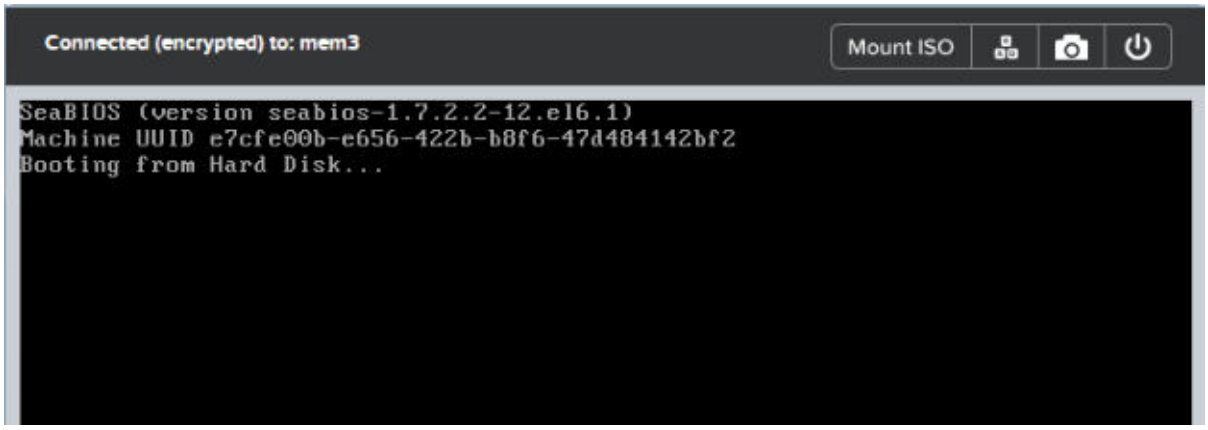


Figure 32: Virtual Network Computing (VNC) Window

5. To start or shut down the VM, click the **Power on** (or **Power off**) action link.

   Power on begins immediately. If you want to power off the VMs, you are prompted to select one of the following options:

   - **Power Off**. Hypervisor performs a hard power off action on the VM.
   - **Power Cycle**. Hypervisor performs a hard restart action on the VM.
   - **Reset**. Hypervisor performs an ACPI reset action through the BIOS on the VM.
   - **Guest Shutdown**. Operating system of the VM performs a graceful shutdown.
   - **Guest Reboot**. Operating system of the VM performs a graceful restart.

   Note: If you perform power operations such as Guest Reboot or Guest Shutdown by using the Prism Element web console or API on Windows VMs, these operations might silently fail without any error messages if at that time a screen saver is running in the Windows VM. Perform the same power operations again immediately, so that they succeed.

6. To make a snapshot of the VM, click the **Take Snapshot** action link.

   See Virtual Machine Snapshots on page 109 for more information.

7. To migrate the VM to another host, click the **Migrate** action link.

   This displays the Migrate VM dialog box. Select the target host from the drop-down list (or select the **System will automatically select a host** option to let the system choose the host) and then click the **Migrate** button to start the migration.

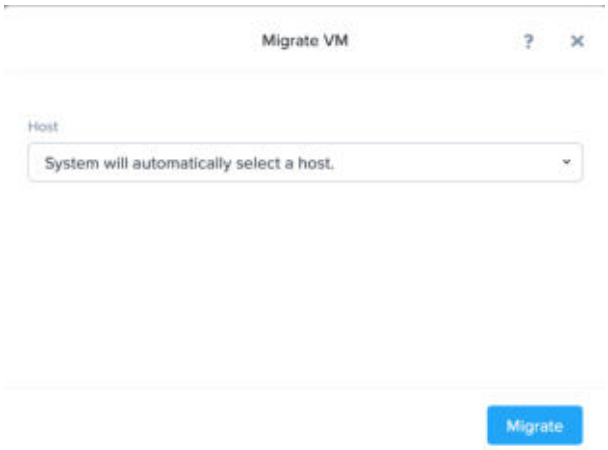


Figure 33: Migrate VM Dialog Box

> Note: Nutanix recommends to live migrate VMs when they are under light load. If they are migrated while heavily utilized, migration may fail because of limited bandwidth.

8. To clone the VM, click the **Clone** action link.

   This displays the Clone VM dialog box, which includes the same fields as the Create VM dialog box. A cloned VM inherits the most the configurations (except the name) of the source VM. Enter a name for the clone and then click the **Save** button to create the clone. You can optionally override some of the configurations before clicking the **Save** button. For

example, you can override the number of vCPUs, memory size, boot priority, NICs, or the guest customization.

> Note:
>
> - You can clone up to 250 VMs at a time.
> - You cannot override the secure boot setting while cloning a VM, unless the source VM already had secure boot setting enabled.

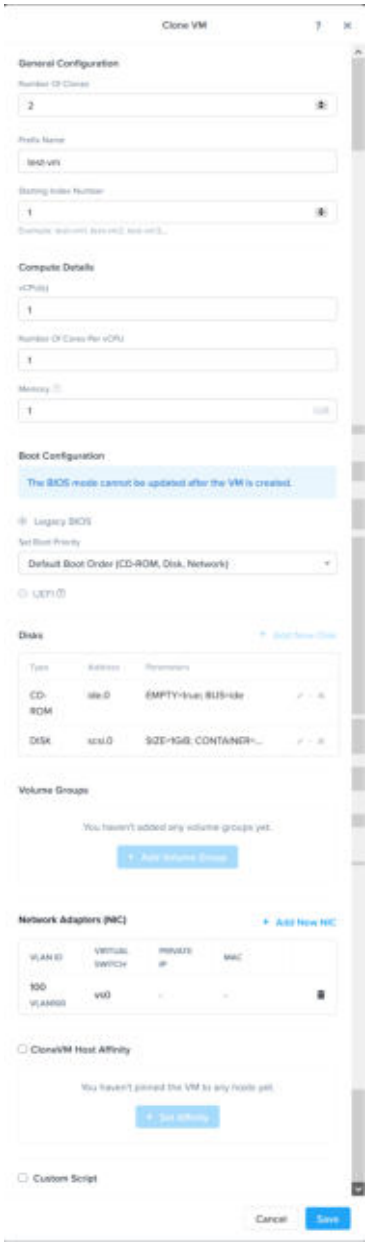


Figure 34: Clone VM Window

9. To modify the VM configuration, click the **Update** action link.

   The Update VM dialog box appears, which includes the same fields as the Create VM dialog box. Modify the configuration as needed, and then save the configuration. In addition

to modifying the configuration, you can attach a volume group to the VM and enable flash mode on the VM. If you attach a volume group to a VM that is part of a protection domain, the VM is not protected automatically. Add the VM to the same Consistency Group manually.

(For GPU-enabled AHV clusters only) You can add pass-through GPUs if a VM is already using GPU pass-through. You can also change the GPU configuration from pass-through to vGPU or vGPU to pass-through, change the vGPU profile, add more vGPUs, and change the

specified vGPU license. However, you need to power off the VM before you perform these operations.

- Before you add multiple vGPUs to the VM, see *Multiple Virtual GPU Support* and *Restrictions for Multiple vGPU Support* in the *AHV Admin Guide*.
- Multiple vGPUs are supported on the same VM only if you select the highest vGPU profile type.
- For more information on vGPU profile selection, see:
  - *Virtual GPU Types for Supported GPUs* in the *NVIDIA Virtual GPU Software User Guide* in the NVIDIA's *Virtual GPU Software Documentation* webpage, and
  - *GPU and vGPU Support* in the *AHV Administration Guide*.
- After you add the first vGPU, to add multiple vGPUs, see *Adding Multiple vGPUs to the Same VM* in the *AHV Admin Guide*.

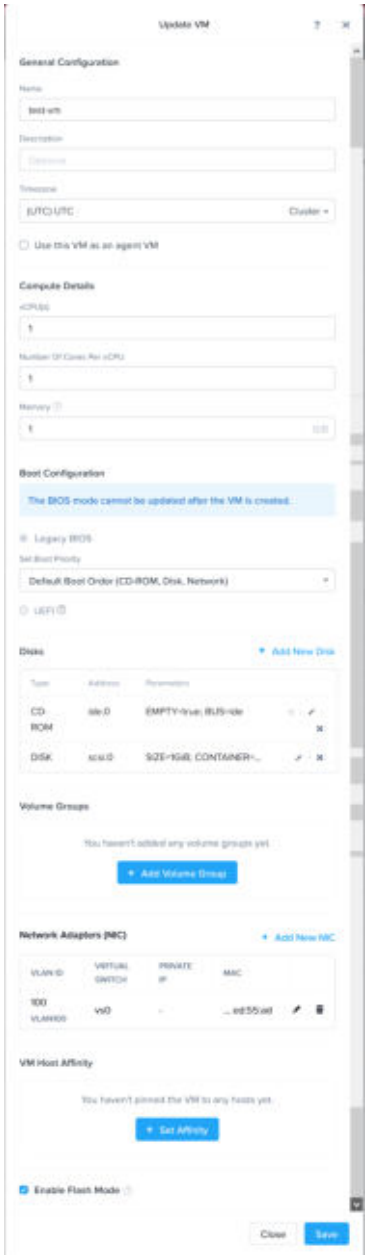


Figure 35: VM Update Dialog Box

> Note: If you delete a vDisk attached to a VM and snapshots associated with this VM exist, space associated with that vDisk is not reclaimed unless you also delete the VM snapshots.

To increase the memory allocation and the number of vCPUs on your VMs while the VMs are powered on (hot-pluggable), do the following:

a. In the **vCPUs** field, you can increase the number of vCPUs on your VMs while the VMs are powered on.

b. In the **Number of Cores Per vCPU** field, you can change the number of cores per vCPU only if the VMs are powered off.

> Note: This is not a hot-pluggable feature.

c. In the **Memory** field, you can increase the memory allocation on your VMs while the VMs are powered on.

For more information about hot-pluggable vCPUs and memory, see the *Virtual Machine Memory and CPU Hot-Plug Configurations* topic in the *AHV Administration Guide*.

To attach a volume group to the VM, do the following:

a. In the **Volume Groups** section, click **Add volume group**, and then do one of the following:

   » From the **Available Volume Groups** list, select the volume group that you want to attach to the VM.

   » Click **Create new volume group**, and then, in the **Create Volume Group** dialog box, create a volume group (see *Creating a Volume Group* in the *Prism Web Console Guide*). After you create a volume group, select it from the **Available Volume Groups** list.

   Repeat these steps until you have added all the volume groups that you want to attach to the VM.

b. Click **Add**.

10. To enable flash mode on the VM, click the **Enable Flash Mode** check box.

    » After you enable this feature on the VM, the status is updated in the VM table view. To view the status of individual virtual disks (disks that are flashed to the SSD), click the update disk icon in the **Disks** pane in the Update VM window.

    » You can disable the flash mode feature for individual virtual disks. To update the flash mode for individual virtual disks, click the update disk icon in the **Disks** pane and deselect the **Enable Flash Mode** check box.

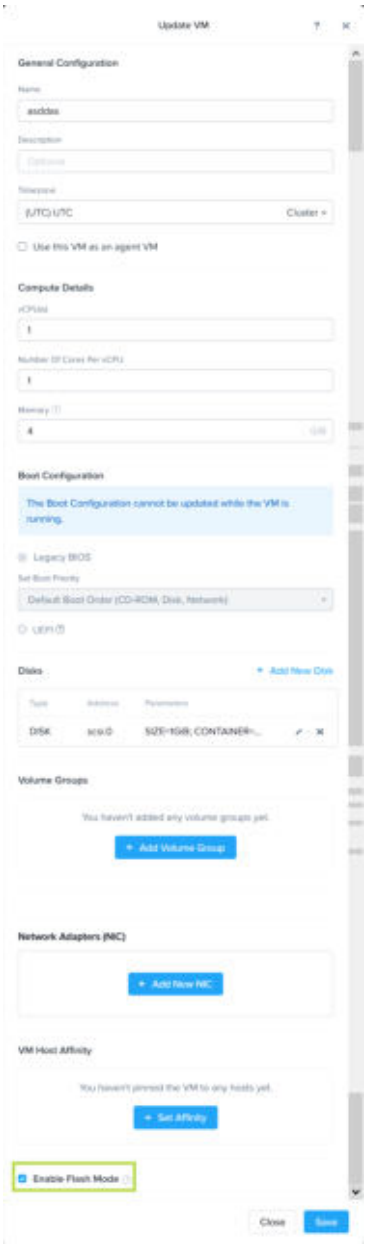


Figure 36: Update VM Resources

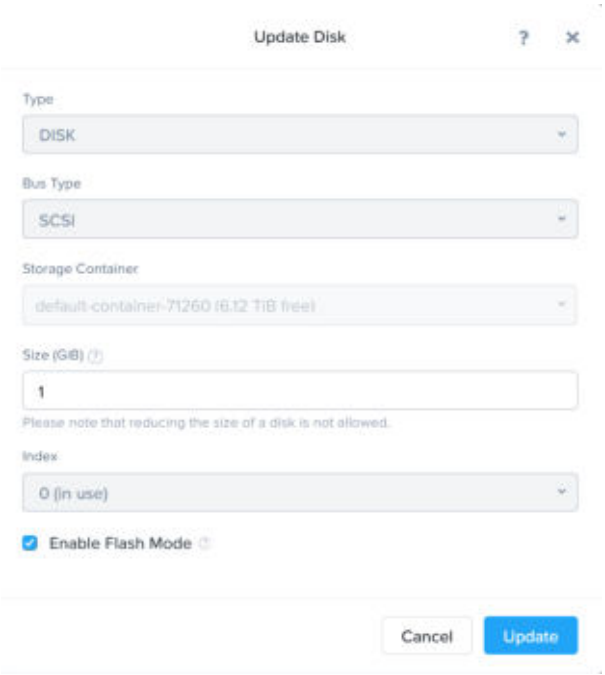


Figure 37: Update VM Resources - VM Disk Flash Mode

11. To delete the VM, click the **Delete** action link. A window prompt appears; click the **OK** button to delete the VM.
    The deleted VM disappears from the list of VMs in the table.

Virtual Machine Snapshots

You can generate snapshots of virtual machines or VMs. You can generate snapshots of VMs manually or automatically. Some of the purposes that VM snapshots serve are as follows:

- Disaster recovery
- Testing - as a safe restoration point in case something went wrong during testing.
- Migrate VMs
- Create multiple instances of a VM.

Snapshot is a point-in-time reference of the system's data, used for restoration and replication of data.. You can generate snapshots and store them locally or remotely. Snapshots are mechanism to capture the delta changes that has occurred over time. Snapshots are primarily used for data protection and disaster recovery. Snapshots are not autonomous like backup, in the sense that they depend on the underlying VM infrastructure and other snapshots to restore the VM. Snapshots consume less resources compared to a full autonomous backup. Typically, a VM snapshot captures the following:

- The state including the power state (for example, powered-on, powered-off, suspended) of the VMs.
- The data includes all the files that make up the VM. This data also includes the data from disks, memory, and other devices, such as virtual network interface cards.

VM Snapshots and Snapshots for Disaster Recovery

The VM Dashboard only allows you to generate VM snapshots manually. You cannot select VMs and schedule snapshots of the VMs using the VM dashboard. The snapshots generated manually have very limited utility.

> Note: These snapshots (stored locally) cannot be replicated to other sites.

You can schedule and generate snapshots as a part of the disaster recovery process using Nutanix DR solutions. AOS generates snapshots when you protect a VM with a protection domain using the Data Protection dashboard in Prism Web Console (see the *Data Protection and Recovery with Prism Element* guide). Similarly, Recovery Points (snapshots are called Recovery Points in Prism Central) when you protect a VM with a protection policy using Data Protection dashboard in Prism Central (see the *Leap Administration Guide*).

For example, in the Data Protection dashboard in Prism Web Console, you can create schedules to generate snapshots using various RPO schemes such as asynchronous replication with frequency intervals of 60 minutes or more, or NearSync replication with frequency intervals of as less as 20 seconds up to 15 minutes. These schemes create snapshots in addition to the ones generated by the schedules, for example, asynchronous replication schedules generate snapshots according to the configured schedule and, in addition, an extra snapshot every 6 hours. Similarly, NearSync generates snapshots according to the configured schedule and also generates one extra snapshot every hour.

Similarly, you can use the options in the Data Protection section of Prism Central to generate Recovery Points using the same RPO schemes.

Windows VM Provisioning

Nutanix VirtIO for Windows

Nutanix VirtIO is a collection of drivers for paravirtual devices that enhance the stability and performance of virtual machines on AHV.

Nutanix VirtIO is available in two formats:

- To install Windows in a VM on AHV, use the VirtIO ISO.
- To update VirtIO for Windows, use the VirtIO MSI installer file.

Use Nutanix Guest Tools (NGT) to install the Nutanix VirtIO Package. For more information about installing the Nutanix VirtIO package using the NGT, see *NGT Installation* in the *Prism Web Console Guide*.

VirtIO Requirements

Requirements for Nutanix VirtIO for Windows.

VirtIO supports the following operating systems:

- Microsoft Windows server version: Windows 2008 R2 or later
- Microsoft Windows client version: Windows 7 or later

> Note: On Windows 7 and Windows Server 2008 R2, install Microsoft KB3033929 or update the operating system with the latest Windows Update to enable support for SHA2 certificates.

Installing or Upgrading Nutanix VirtIO for Windows

Download Nutanix VirtIO and the Nutanix VirtIO Microsoft installer (MSI). The MSI installs and upgrades the Nutanix VirtIO drivers.

Before you begin

Make sure that your system meets the VirtIO requirements described in VirtIO Requirements on page 110.

About this task

If you have already installed Nutanix VirtIO, use the following procedure to upgrade VirtIO to a latest version. If you have not yet installed Nutanix VirtIO, use the following procedure to install Nutanix VirtIO.

Procedure

1. Go to the Nutanix Support portal, select **Downloads** > **AHV**, and click **VirtIO**.
2. Select the appropriate VirtIO package.
   - » If you are creating a new Windows VM, download the ISO file. The installer is available on the ISO if your VM does not have Internet access.
   - » If you are updating drivers in a Windows VM, download the MSI installer file.

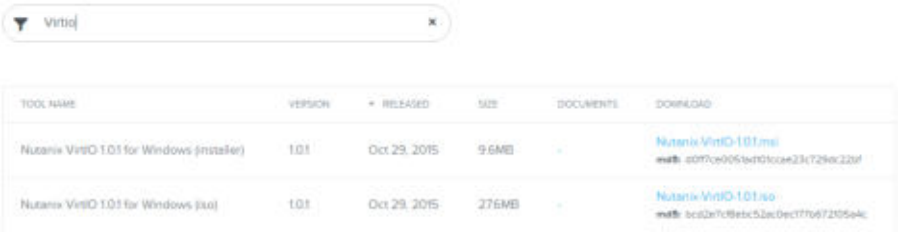


Figure 38: Search filter and VirtIO options

3. Run the selected package.
   - » For the ISO: Upload the ISO to the cluster, as described in *Prism Web Console Guide: Configuring Images*.
   - » For the MSI: open the download file to run the MSI.

4. Read and accept the Nutanix VirtIO license agreement. Click **Install**.

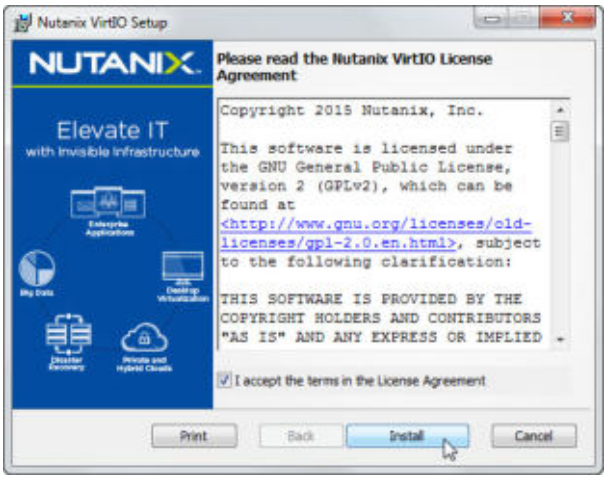


Figure 39: Nutanix VirtIO Windows Setup Wizard

The Nutanix VirtIO setup wizard shows a status bar and completes installation.

Manually Installing or Upgrading Nutanix VirtIO

Manually install or upgrade Nutanix VirtIO.

Before you begin

Make sure that your system meets the VirtIO requirements described in VirtIO Requirements on page 110.

About this task

> Note: To automatically install Nutanix VirtIO, see Installing or Upgrading Nutanix VirtIO for Windows on page 110.

If you have already installed Nutanix VirtIO, use the following procedure to upgrade VirtIO to a latest version. If you have not yet installed Nutanix VirtIO, use the following procedure to install Nutanix VirtIO.

Procedure

1. Go to the Nutanix Support portal, select **Downloads** > **AHV**, and click **VirtIO**.
2. Do one of the following:
   - » Extract the VirtIO ISO into the same VM where you load Nutanix VirtIO, for easier installation.

     If you choose this option, proceed directly to step 7.
   - » Download the VirtIO ISO for Windows to your local machine.

     If you choose this option, proceed to step 3.
3. Upload the ISO to the cluster, as described in the Configuring Images topic of Prism Web Console Guide.
4. Locate the VM where you want to install the Nutanix VirtIO ISO and update the VM.

5. Add the Nutanix VirtIO ISO by clicking **Add New Disk** and complete the indicated fields.
   - **TYPE**: **CD-ROM**
   - **OPERATION**: **CLONE FROM IMAGE SERVICE**
   - **BUS TYPE**: **IDE**
   - **IMAGE**: Select the Nutanix VirtIO ISO
6. Click **Add**.
7. Log on to the VM and browse to **Control Panel** > **Device Manager**.

8. Note: Select the `x86` subdirectory for 32-bit Windows, or the `amd64` for 64-bit Windows.

   Open the devices and select the specific Nutanix drivers for download. For each device, right-click and **Update Driver Software** into the drive containing the VirtIO ISO. For each device, follow the wizard instructions until you receive installation confirmation.

   a. **System Devices** > **Nutanix VirtIO Balloon Drivers**

   b. **Network Adapter** > **Nutanix VirtIO Ethernet Adapter**.

   c. **Processors** > **Storage Controllers** > **Nutanix VirtIO SCSI pass through Controller**
      The Nutanix VirtIO SCSI pass-through controller prompts you to restart your system. Restart at any time to install the controller.

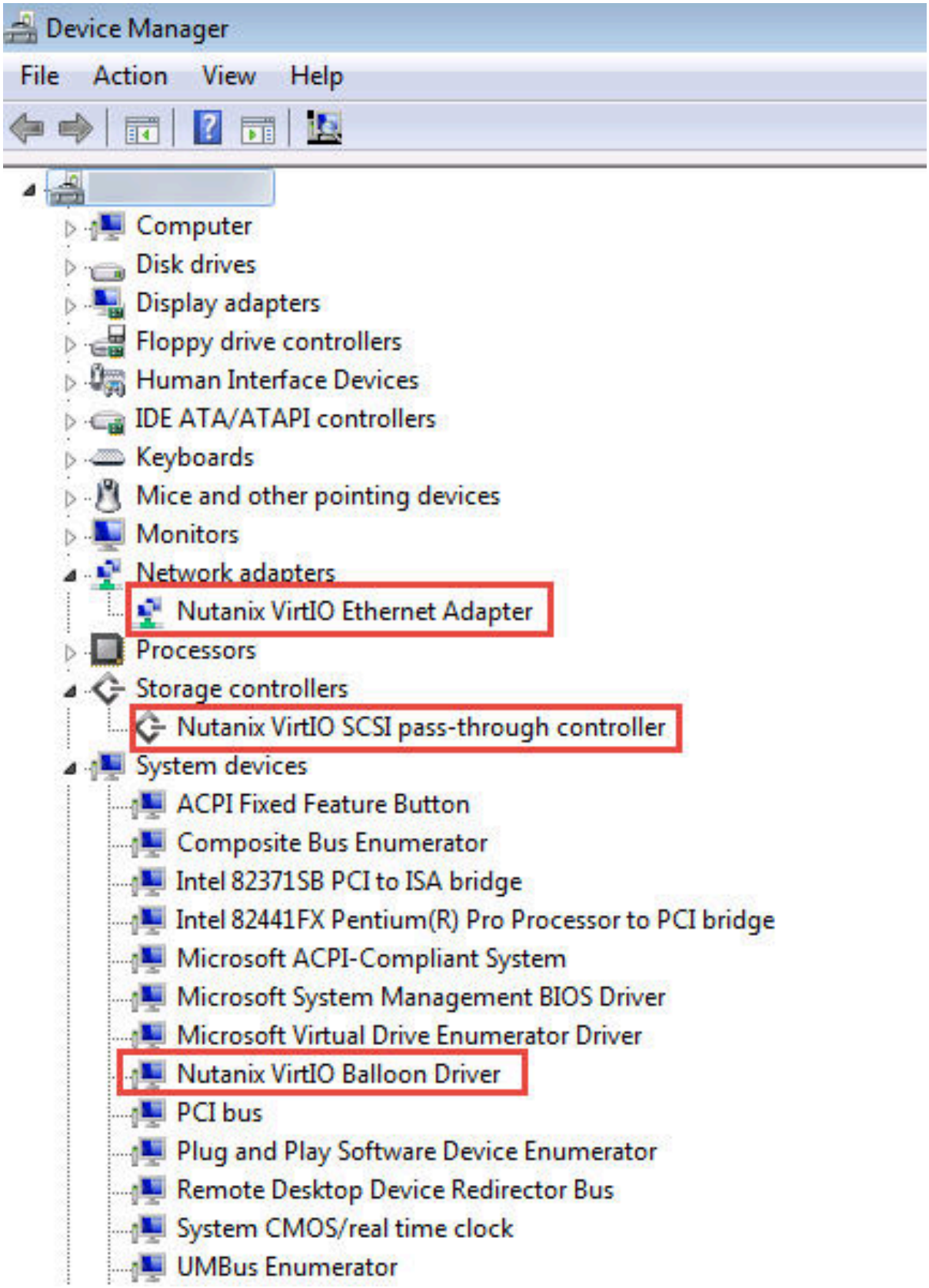


Figure 40: List of Nutanix VirtIO downloads

Creating a Windows VM on AHV with Nutanix VirtIO

Create a Windows VM in AHV, or migrate a Windows VM from a non-Nutanix source to AHV, with the Nutanix VirtIO drivers.

Before you begin

- Upload the Windows installer ISO to your cluster as described in the *Web Console Guide: Configuring Images*.
- Upload the Nutanix VirtIO ISO to your cluster as described in the *Web Console Guide: Configuring Images*.

About this task

To install a new or migrated Windows VM with Nutanix VirtIO, complete the following.

Procedure

1. Log on to the Prism web console using your Nutanix credentials.
2. At the top-left corner, click **Home** > **VM**.
   The VM page appears.

3. Click **+ Create VM** in the corner of the page.
   The Create VM dialog box appears.

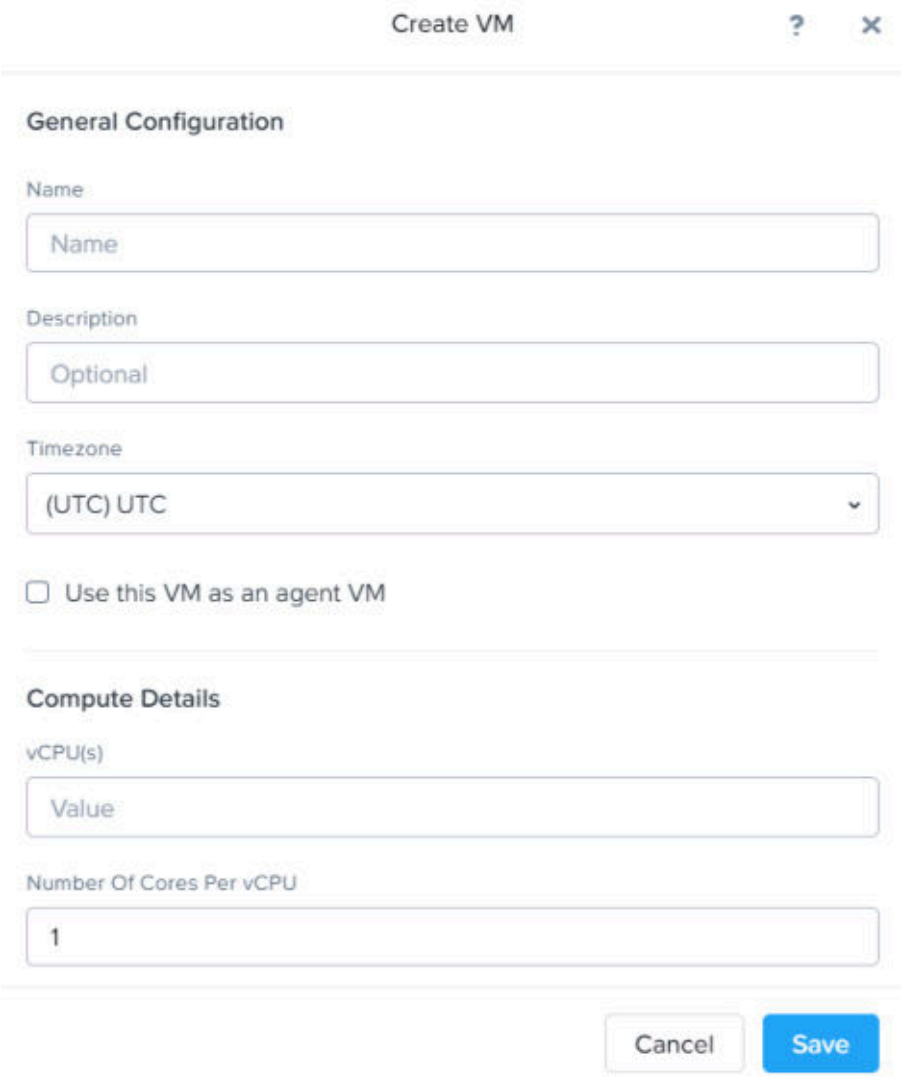


Figure 41: Create VM dialog box

4. Complete the indicated fields.
   a. **NAME**: Enter a name for the VM.
   b. **Description** (optional): Enter a description for the VM.
   c. **Timezone**: Select the timezone that you want the VM to use. If you are creating a Linux VM, select **(UTC) UTC**.

      > Note:
      >
      > The RTC of Linux VMs must be in UTC, so select the UTC timezone if you are creating a Linux VM.
      >
      > Windows VMs preserve the RTC in the local timezone, so set up the Windows VM with the hardware clock pointing to the desired timezone.

   d. **Number of Cores per vCPU**: Enter the number of cores assigned to each virtual CPU.
   e. **MEMORY**: Enter the amount of memory for the VM (in GiB).
5. If you are creating a Windows VM, add a Windows CD-ROM to the VM.
   a. Click the pencil icon next to the CD-ROM that is already present and fill out the indicated fields.
      - **OPERATION**: **CLONE FROM IMAGE SERVICE**
      - **BUS TYPE**: **IDE**
      - **IMAGE**: Select the Windows OS install ISO.
   b. Click **Update**.
      The current CD-ROM opens in a new window.
6. Add the Nutanix VirtIO ISO.
   a. Click **Add New Disk** and complete the indicated fields.
      - **TYPE**: **CD-ROM**
      - **OPERATION**: **CLONE FROM IMAGE SERVICE**
      - **BUS TYPE**: **IDE**
      - **IMAGE**: Select the Nutanix VirtIO ISO.
   b. Click **Add**.
7. Add a new disk for the hard drive.
   a. Click **Add New Disk** and complete the indicated fields.
      - **TYPE**: **DISK**
      - **OPERATION**: **ALLOCATE ON STORAGE CONTAINER**
      - **BUS TYPE**: **SCSI**
      - **STORAGE CONTAINER**: Select the appropriate storage container.
      - **SIZE**: Enter the number for the size of the hard drive (in GiB).
   b. Click **Add** to add the disk driver.

8. If you are migrating a VM, create a disk from the disk image.
   a. Click **Add New Disk** and complete the indicated fields.
      - **TYPE**: **DISK**
      - **OPERATION**: **CLONE FROM IMAGE**
      - **BUS TYPE**: **SCSI**
      - **CLONE FROM IMAGE SERVICE**: Click the drop-down menu and choose the image you created previously.
   b. Click **Add** to add the disk driver.
9. Optionally, after you have migrated or created a VM, add a network interface card (NIC).
   a. Click **Add New NIC**.
   b. In the **VLAN ID** field, choose the VLAN ID according to network requirements and enter the IP address, if necessary.
   c. Click **Add**.
10. Click **Save**.

What to do next

Install Windows by following Installing Windows on a VM on page 119.

Installing Windows on a VM

Install a Windows virtual machine.

Before you begin

Create a Windows VM as described in the *Migration Guide: Creating a Windows VM on AHV after Migration*.

Procedure

1. Log on to the web console.
2. Click **Home** > **VM** to open the VM dashboard.
3. Select the Windows VM.
4. In the center of the VM page, click **Power On**.
5. Click **Launch Console**.
   The Windows console opens in a new window.
6. Select the desired language, time and currency format, and keyboard information.
7. Click **Next** > **Install Now**.
   The Windows setup dialog box shows the operating systems to install.
8. Select the Windows OS you want to install.
9. Click **Next** and accept the license terms.
10. Click **Next** > **Custom: Install Windows only (advanced)** > **Load Driver** > **OK** > **Browse**.

11. Choose the Nutanix VirtIO driver.
    a. Select the Nutanix VirtIO CD drive.
    b. Expand the Windows OS folder and click **OK**.

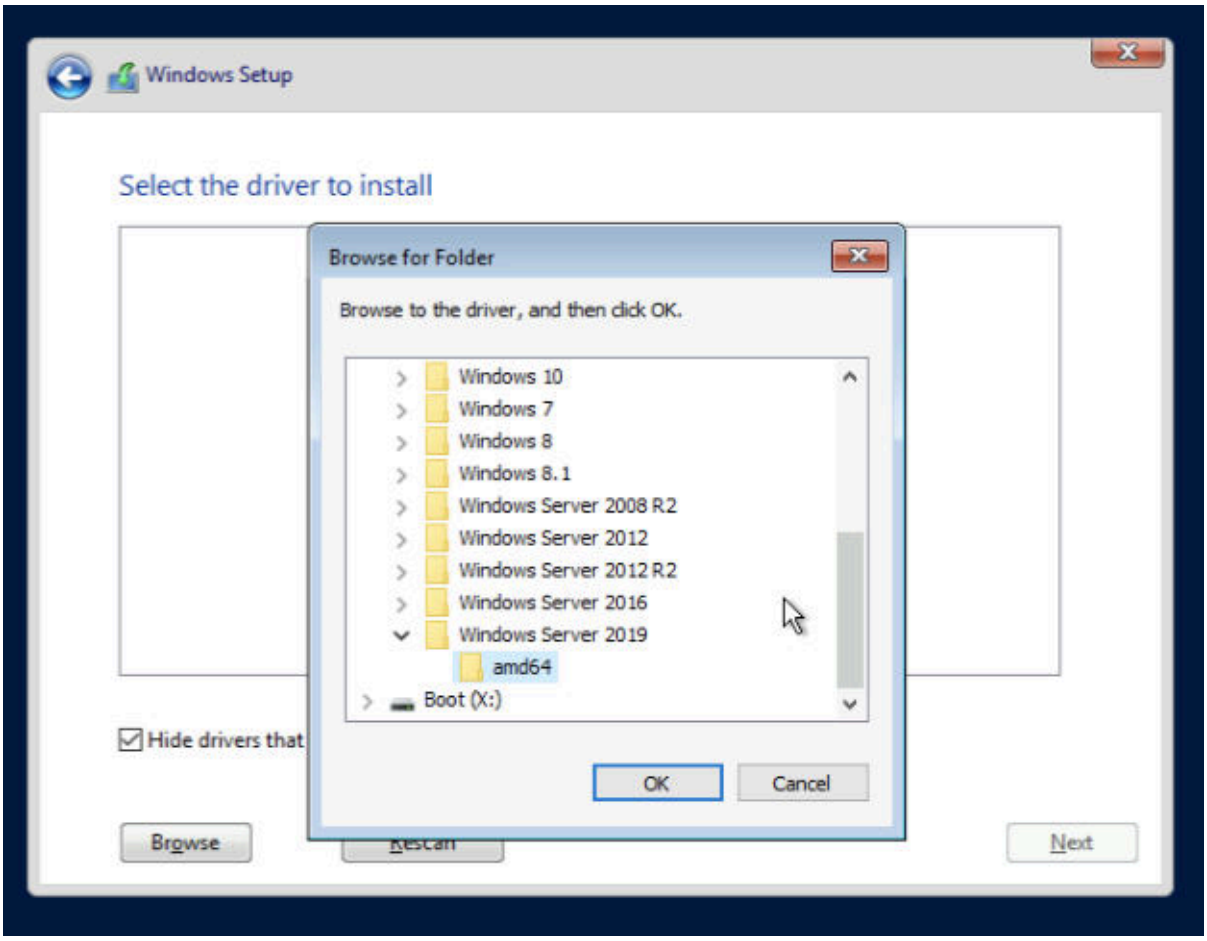


Figure 42: Select the Nutanix VirtIO drivers for your OS

The Select the driver to install window appears.

12. Select the VirtIO SCSI driver (`vioscsi.inf`) and click **Next**.

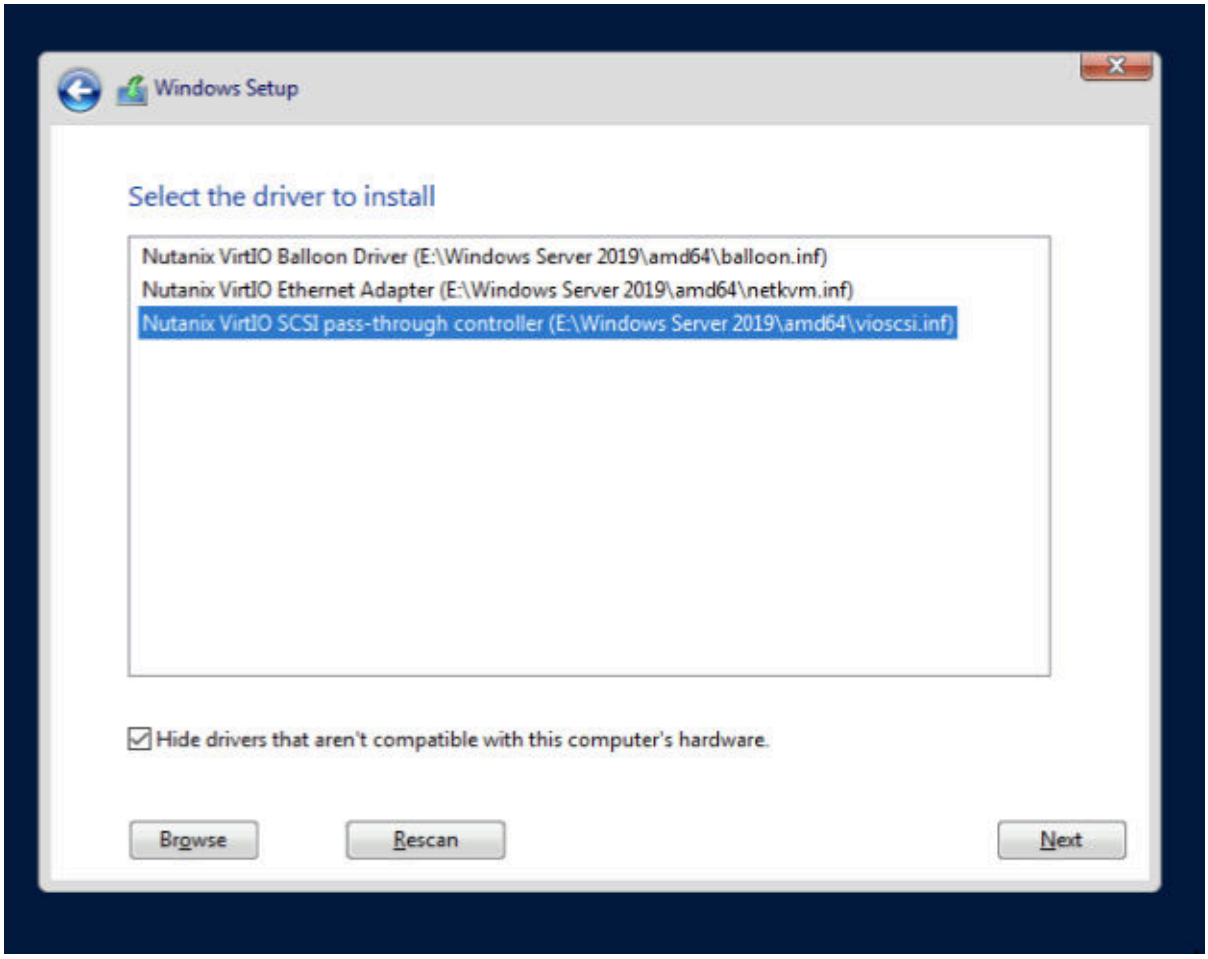


Figure 43: Select the Driver for Installing Windows on a VM

The `amd64` folder contains drivers for 64-bit operating systems. The `x86` folder contains drivers for 32-bit operating systems.

> Note: From Nutanix VirtIO driver version 1.1.5, the driver package contains Windows Hardware Quality Lab (WHQL) certified driver for Windows.

13. Select the allocated disk space for the VM and click **Next**.
    Windows shows the installation progress, which can take several minutes.

14. Enter your user name and password information and click **Finish**.

    Installation can take several minutes.

    Once you complete the logon information, Windows setup completes installation.

15. Follow the instructions in Installing or Upgrading Nutanix VirtIO for Windows on page 110 to install other drivers which are part of Nutanix VirtIO package.

Windows Defender Credential Guard Support in AHV

AHV enables you to use the Windows Defender Credential Guard security feature on Windows guest VMs.

Windows Defender Credential Guard feature of Microsoft Windows operating systems allows you to securely isolate user credentials from the rest of the operating system. By that means, you can protect guest VMs from credential theft attacks such as Pass-the-Hash or Pass-The-Ticket.

See the Microsoft documentation for more information about the Windows Defender Credential Guard security feature.

Windows Defender Credential Guard Architecture in AHV

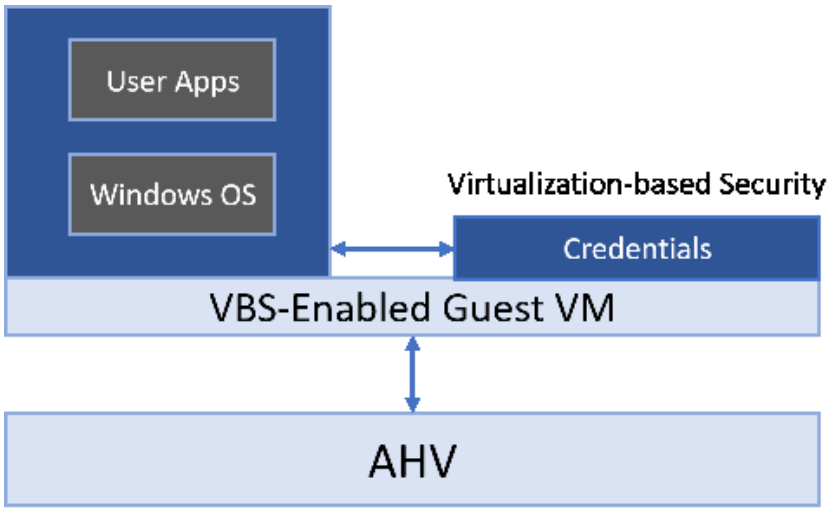


Figure 44: Architecture

Windows Defender Credential Guard uses Microsoft virtualization-based security to isolate user credentials in the virtualization-based security (VBS) module in AHV. When you enable Windows Defender Credential Guard on an AHV guest VM, the guest VM runs on top of AHV running both the Windows OS and VBS. Each Windows OS guest VM, which has credential guard enabled, has a VBS to securely store credentials.

Windows Defender Credential Guard Requirements

Ensure the following to enable Windows Defender Credential Guard:

1. AOS, AHV, and Windows Server versions support Windows Defender Credential Guard:
   - AOS version must be 5.19 or later
   - AHV version must be AHV 20201007.1 or later
   - Windows Server version must be Windows server 2016 or later, Windows 10 Enterprise or later and Windows Server 2019 or later
2. UEFI, Secure Boot, and machine type q35 are enabled in the Windows VM from AOS.

   The Prism Element workflow to enable Windows Defender Credential Guard includes the workflow to enable these features.

Limitations

- Windows Defender Credential guard is not supported on hosts with AMD CPUs.

- If you enable Windows Defender Credential Guard for your AHV guest VMs, the following optional configurations are not supported:
  - vTPM (Virtual Trusted Platform Modules) to store MS policies.
  - DMA protection (vIOMMU).
  - Nutanix Live Migration.
  - Cross hypervisor DR of Credential Guard VMs.

  CAUTION: Use of Windows Defender Credential Guard in your AHV clusters impacts VM performance. If you enable Windows Defender Credential Guard on AHV guest VMs, VM density drops by ~15–20%. This expected performance impact is due to nested virtualization overhead added as a result of enabling credential guard.

Enabling Windows Defender Credential Guard Support in AHV Guest VMs

You can enable Windows Defender Credential Guard when you are either creating a VM or updating a VM.

About this task

Perform the following procedure to enable Windows Defender Credential Guard:

Procedure

1. Enable Windows Defender Credential Guard when you are either creating a VM or updating a VM. Do one of the following:
   » If you are creating a VM, see step 2.
   » If you are updating a VM, see step 3.

2. If you are creating a Windows VM, do the following:
   a. Log on to the Prism Element web console.
   b. In the VM dashboard, click **Create VM**.
   c. Fill in the mandatory fields to configure a VM.
   d. Under **Boot Configuration**, select **UEFI**, and then select the **Secure Boot** and **Windows Defender Credential Guard** options.

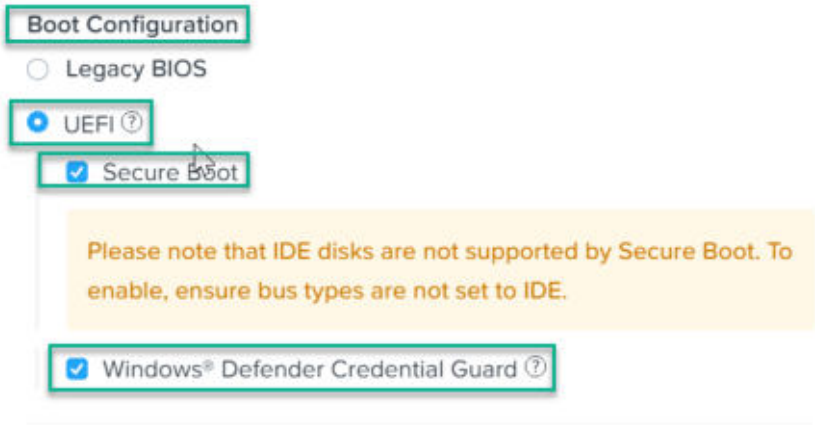


Figure 45: Enable Windows Defender Credential Guard

   See UEFI Support for VM on page 135 and Secure Boot Support for VMs on page 140 for more information about these features.

   e. Proceed to configure other attributes for your Windows VM.
   See Creating a Windows VM on AHV with Nutanix VirtIO on page 116 for more information.
   f. Click **Save**.
   g. Turn on the VM.
3. If you are updating an existing VM, do the following:
   a. Log on to the Prism Element web console.
   b. In the VM dashboard, click the **Table** view, select the VM, and click **Update**.
   c. Under **Boot Configuration**, select **UEFI**, and then select the **Secure Boot** and **Windows Defender Credential Guard** options.

   > Note:
   > If the VM is configured to use BIOS, install the guest OS again.
   > If the VM is already configured to use UEFI, skip the step to select Secure Boot.

   See UEFI Support for VM on page 135 and Secure Boot Support for VMs on page 140 for more information about these features.

   d. Click **Save**.
   e. Turn on the VM.

4. Enable Windows Defender Credential Guard in the Windows VM by using group policy.

   See the *Enable Windows Defender Credential Guard by using the Group Policy* procedure of the *Manage Windows Defender Credential Guard* topic in the Microsoft documentation to enable VBS, Secure Boot, and Windows Defender Credential Guard for the Windows VM.

5. Open command prompt in the Windows VM and apply the **Group Policy** settings:

   ```
   > gpupdate /force
   ```

   If you have not enabled Windows Defender Credential Guard (step 4) and perform this step (step 5), a warning similar to the following is displayed:

   ```
   Updating policy...

   Computer Policy update has completed successfully.

   The following warnings were encountered during computer policy processing:

   Windows failed to apply the {F312195E-3D9D-447A-A3F5-08DFFA24735E} settings.
    {F312195E-3D9D-447A-A3F5-08DFFA24735E} settings might have its own log file. Please click on
    the "More information" link.
   User Policy update has completed successfully.

   For more detailed information, review the event log or run GPRESULT /H GPReport.html from the
    command line to access information about Group Policy results.
   ```

   Event Viewer displays a warning for the group policy with an error message that indicates Secure Boot is not enabled on the VM.

   To view the warning message in Event Viewer, do the following:

   - In the Windows VM, open **Event Viewer**.
   - Go to **Windows Logs** -> **System** and click the warning with the **Source** as **GroupPolicy (Microsoft-Windows-GroupPolicy)** and **Event ID** as **1085**.

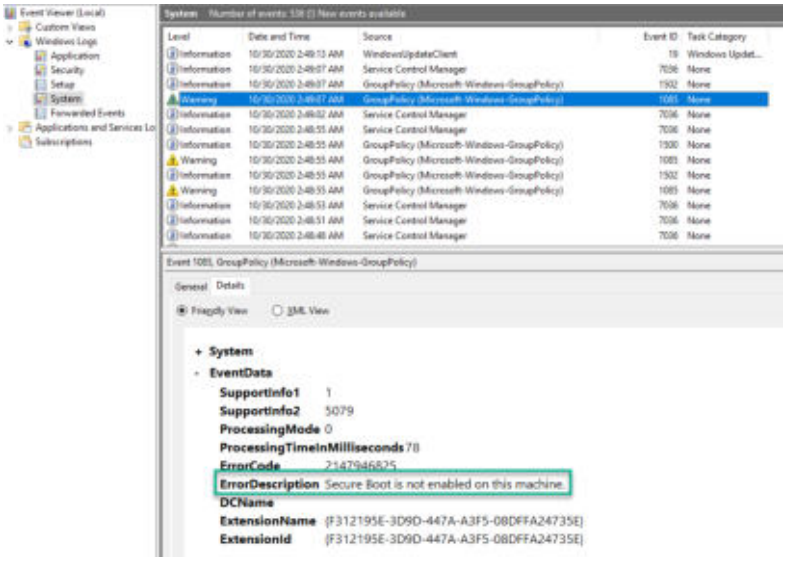


   Figure 46: Warning in Event Viewer

   Note: Ensure that you follow the steps in the order that is stated in this document to successfully enable Windows Defender Credential Guard.

6. Restart the VM.
7. Verify if Windows Defender Credential Guard is enabled in your Windows VM.
   a. Start a Windows PowerShell terminal.
   b. Run the following command.

```
PS > Get-CimInstance -ClassName Win32_DeviceGu
ard -Namespace 'root\Microsoft\Windows\DeviceGuard'
```

   An output similar to the following is displayed.

```
PS > Get-CimInstance -ClassName Win32_DeviceGuard -Namespace 'root\Microsoft\Windows
\DeviceGuard'
AvailableSecurityProperties                : {1, 2, 3, 5}
CodeIntegrityPolicyEnforcementStatus       : 0
InstanceIdentifier                         : 4ff40742-2649-41b8-bdd1-e80fad1cce80
RequiredSecurityProperties                 : {1, 2}
SecurityServicesConfigured                 : {1}
SecurityServicesRunning                    : {1}
UsermodeCodeIntegrityPolicyEnforcementStatus : 0
Version                                    : 1.0
VirtualizationBasedSecurityStatus          : 2
PSComputerName
```

   Confirm that both **SecurityServicesConfigured** and **SecurityServicesRunning** have the value **{ 1 }**.

   Alternatively, you can verify if Windows Defender Credential Guard is enabled by using System Information (msinfo32):

   a. In the Windows VM, open **System Information** by typing **msinfo32** in the search field next to the **Start** menu.
   b. Verify if the values of the parameters are as indicated in the following screen shot:

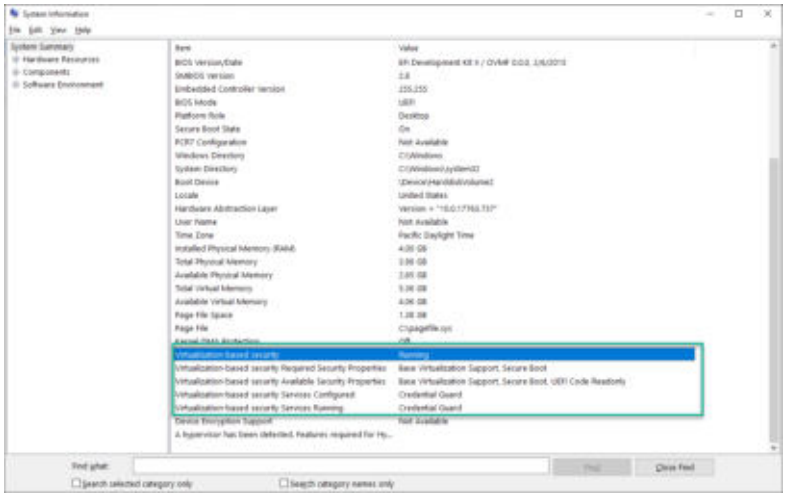


Figure 47: Verify Windows Defender Credential Guard

Affinity Policies for AHV

As an administrator of an AHV cluster, you can create VM-Host affinity policies for virtual machines on an AHV cluster. By defining these policies, you can control the placement of virtual machines on the hosts within a cluster.

> Note: VMs with Host affinity policies can only be migrated to the hosts specified in the affinity policy. If only one host is specified, the VM cannot be migrated or started on another host during an HA event. For more information, see Non-Migratable Hosts on page 134.

You can define affinity policies for a VM at two levels:

Affinity Policies defined in Prism Element

In Prism Element, you can define affinity policies at VM level during the VM create or update operation. You can use an affinity policy to specify that a particular VM can only run on the members of the affinity host list.

Affinity Policies defined in Prism Central

In Prism Central, you can define category-based VM-Host affinity policies, where a set of VMs can be affined to run only on a particular set of hosts. Category-based affinity policy enables you to easily manage affinities for a large number of VMs.

Affinity Policies Defined in Prism Element

In Prism Element, you can define scheduling policies for virtual machines on an AHV cluster at a VM level. By defining these policies, you can control the placement of a virtual machine on specific hosts within a cluster.

You can define two types of affinity policies in Prism Element.

VM-Host Affinity Policy

The VM-host affinity policy controls the placement of a VM. You can use this policy to specify that a selected VM can only run on the members of the affinity host list. This policy checks and enforces where a VM can be hosted when you power on or migrate the VM.

> Note:
>
> - If you choose to apply the VM-host affinity policy, it limits Acropolis HA and Acropolis Dynamic Scheduling (ADS) in such a way that a virtual machine cannot be powered on or migrated to a host that does not conform to requirements of the affinity policy as this policy is enforced mandatorily.
> - The VM-host anti-affinity policy is not supported.
> - VMs configured with host affinity settings retain these settings if the VM is migrated to a new cluster. Remove the VM-host affinity policies applied to a VM that you want to migrate to another cluster, as the UUID of the host is retained by the VM and it does not allow VM restart on the destination cluster. When you attempt to protect such VMs, it is successful. However, some disaster recovery operations like migration fail and attempts to power on these VMs also fail.

You can define the VM-host affinity policies by using Prism Element during the VM create or update operation. For more information, see *Creating a VM (AHV)* in *Prism Web Console Guide* or *AHV Administration Guide*.

VM-VM Anti-Affinity Policy

You can use this policy to specify anti-affinity between the virtual machines. The VM-VM anti-affinity policy keeps the specified virtual machines apart in such a way that when a problem occurs with one host, you should not lose both the virtual machines. However, this is a preferential policy. This policy does not limit the Acropolis Dynamic Scheduling (ADS) feature to take necessary action in case of resource constraints.

Note:

- Currently, you can only define VM-VM anti-affinity policy by using aCLI. For more information, see Configuring VM-VM Anti-Affinity Policy on page 128.
- The VM-VM affinity policy is not supported.

Note: If a VM is cloned that has the affinity policies configured, then the policies are not automatically applied to the cloned VM. However, if a VM is restored from a DR snapshot, the policies are automatically applied to the VM.

Limitations of Affinity Rules

Even though if a host is removed from a cluster, the host UUID is not removed from the host-affinity list for a VM.

Configuring VM-VM Anti-Affinity Policy

To configure VM-VM anti-affinity policies, you must first define a group and then add all the VMs on which you want to define VM-VM anti-affinity policy.

About this task

Note: Currently, the VM-VM affinity policy is not supported.

Perform the following procedure to configure the VM-VM anti-affinity policy.

Procedure

1. Log on to the Controller VM with SSH session.
2. Create a group.

   ```
   nutanix@cvm$ acli vm_group.create group_name
   ```

   Replace *group_name* with the name of the group.
3. Add the VMs on which you want to define anti-affinity to the group.

   ```
   nutanix@cvm$ acli vm_group.add_vms group_name vm_list=vm_name
   ```

   Replace *group_name* with the name of the group. Replace *vm_name* with the name of the VMs that you want to define anti-affinity on.
4. Configure VM-VM anti-affinity policy.

   ```
   nutanix@cvm$ acli vm_group.antiaffinity_set group_name
   ```

   Replace *group_name* with the name of the group.

   After you configure the group and then power on the VMs, the VMs that are part of the group are started (attempt to start) on the different hosts. However, this is a preferential policy. This policy does not limit the Acropolis Dynamic Scheduling (ADS) feature to take necessary action in case of resource constraints.

Removing VM-VM Anti-Affinity Policy

Perform the following procedure to remove the VM-VM anti-affinity policy.

Procedure

1. Log on to the Controller VM with SSH session.

2. Remove the VM-VM anti-affinity policy.

```
nutanix@cvm$ acli vm_group.antiaffinity_unset group_name
```

Replace `group_name` with the name of the group.

The VM-VM anti-affinity policy is removed for the VMs that are present in the group, and they can start on any host during the next power on operation (as necessitated by the ADS feature).

Affinity Policies Defined in Prism Central

In Prism Central, you can define category-based VM-Host affinity policies, where a set of VMs can be affined to run only on a particular set of hosts. Category-based affinity policy enables you to easily manage affinities for a large number of VMs. In case of any changes to the affined hosts, you only need to update the category of the host, and it updates the affinity policy for all the affected VMs.

This policy checks and enforces where a VM can be hosted when you start or migrate the VM. If there are no resources available on any of the affined hosts, the VM does not get started.

> Note:
>
> If you create a VM-Host affinity policy for a VM that is configured for asynchronous replication, you must create similar categories and corresponding policies on the remote site as well. If you define similar categories and policies on the remote site, affinity policies will be applied when the VMs are migrated to the remote site.

Limitations of Affinity Policies

Affinity policies created in Prism Central have the following limitations:

- Only a super admin can create, modify, or delete affinity policies.
- The minimum supported versions for VM-Host affinity policies are version 6.1 for Prism Element and version 2022.1 for Prism Central.
- Host category attach or detach takes around five minutes to get reflected in the applicable affinity policies.
- You cannot apply VM-Host affinity policy on a VM that is enabled for synchronous replication. Also, you cannot enable synchronous replication on a VM that is associated with a VM-Host affinity policy.

Affinity Policy Configuration Workflow

About this task

To set up an affinity policy, do the following:

Procedure

1. Create categories for the following entities:

   a. VMs

   b. Hosts

   For information about creating a category, see *Creating a Category* topic in the *Prism Central Guide* (see Prism).

2. Apply the VM categories to the VMs and host categories to the hosts.

   For information about associating categories with VMs, see *Associating VMs with Categories* topic in the *Prism Central Guide* (see Prism). For information about associating categories with hosts, see Associating hosts with Categories.

3. Create the affinity policy. See *Creating an Affinity Policy* topic in the *Prism Central Guide* (see Prism).

Associating VMs with Categories

About this task

To associate categories with VMs, do the following:

Procedure

1. In Prism Central, in the **Entities** menu, go to **Compute & Storage** > **VMs**.
2. Select the VMs that you want to associate with a category.
3. In the **Actions** menu that is displayed, click **Manage Categories**.
4. In the **Search for a category** field, type the name of the category or value that you want to add, and then select the category and value from the list that is displayed.
5. Click the add button to add as many search fields and repeat this step for each category that you want to add. To remove a category, click the remove button beside the field.
6. Click **Save**.

Associating Hosts with Categories

About this task

To associate categories with hosts, do the following:

Procedure

1. In Prism Central, in the Entities menu, go to **Hardware** > **Hosts**.
2. Select the hosts that you want to associate with a category.
3. In the **Actions** menu that is displayed, click **Manage Categories**.
4. In the **Search for a category** field, type the name of the category or value that you want to add, and then select the category and value from the list that is displayed.
5. Click the add button to add as many search fields and repeat this step for each category that you want to add. To remove a category, click the remove button beside the field.
6. Click **Save**.

Creating an Affinity Policy

About this task

To create an affinity policy, do the following:

Before you begin

Complete the following tasks:

- Configure the categories that you need to associate with the VMs. Associate this category with all the relevant VMs. Alternatively, you can associate the category with the VMs after creation of the affinity policy.
- Configure the categories that you need to associate with the hosts. Associate this category with all the relevant hosts. Alternatively, you can associate the category with the hosts after creation of the affinity policy.

> Note: If you have configured any legacy affinity policy (non-category-based affinity policy) associated with the VMs, you must first remove those legacy affinity policies to allow the creation of category-based affinity policies associated with the same VMs.

Procedure

1. In the VMs Summary View, go to **Policies** > **Affinity Policies**, and then click **Create**.
2. In the **Create Affinity Policy** page, enter a name and description (optional) for the policy.

Figure 48: Create Affinity Policy

3. Click inside the **VM Categories** search field and select the category you want to associate with the VMs.
4. Click inside the **Host Categories** search field and select the category you associated with the hosts.
5. Click **Create**.

Updating an Affinity Policy

You can update an affinity policy. Updates to an affinity policy can result in a policy violation. Prism Central attempts to correct the violation by executing a series of actions.

About this task

To update an affinity policy, do the following:

Procedure

1. In the entity menu, go to **Policies** > **Affinity Policies**, and then click the affinity policy that you want to update.
2. From the **Actions** menu that is displayed, click **Update**.

3. The **Update Affinity Policy** page that is displayed includes the same options and settings as the **Create Affinity Policy** page.

Figure 49: Update Affinity Policy

4. Update the settings that you want, and then click **Save**.

Affinity Policies Summary View

To access the affinity policies dashboard, select **Compute & Storage** > **VMs** > **Policies** > **Affinity Policies** from the **Entities** menu in Prism Central .

> Note: This section describes the information and options that appear in the affinity policies dashboard.
>
> - See *Entity Exploring* chapter in the *Prism Central Guide* (see Prism) for instructions on how to view and organize that information in various ways.
> - See *Affinity Policies Defined in Prism Central* topic in the *Prism Central Guide* (see Prism) for information on how to create or modify the affinity policies.

The affinity policies dashboard displays a list of current policies that includes the name and type for each policy.

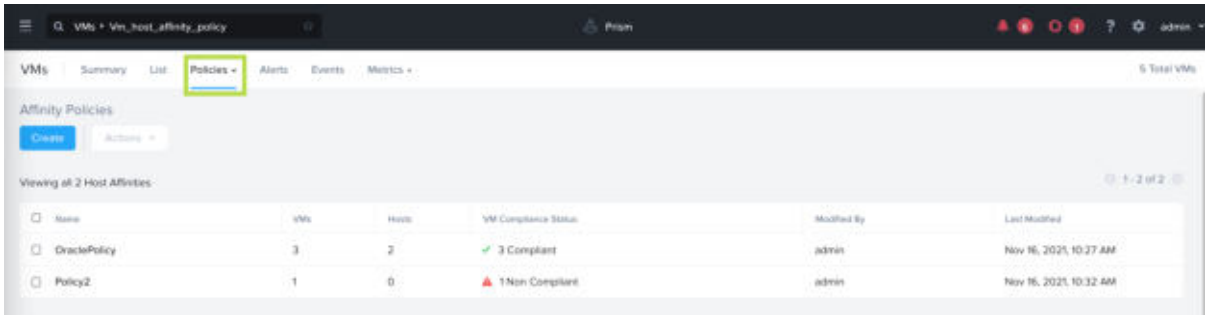


Figure 50: Affinity Policies Dashboard

The following table describes the fields that appear in the affinity policies list. A dash (-) is displayed in a field when a value is not available or not applicable.

Table 7: Affinity Policies List Fields

Parameter	Description	Values
Name	Displays the policy name.	(name)

Parameter	Description	Values
VMs	Displays the count of VMs associated with this policy.	(number of VMs)
Hosts	Displays the count of hosts associated with this policy.	(number of hosts)
VM Compliance Status	Displays the compliance status of the VMs associated with this policy. If the policy is being applied and the compliance status is not yet known, the status is displayed as Pending. If a VM is part of multiple VM-Host affinity policies, the oldest policy is applied on the VM. For rest of the policies, the VM is displayed as non-compliant.	(number of VMs Compliant/Non Compliant/Pending)
Modified By	Displays the name of the user who modified the policy last time.	(user)
Last Modified	Displays the date and time when the policy was modified last time.	(date & time)

Affinity Policies Details View

To access the details page for an Affinity policy, click the desired policy name in the list (see Affinity Policies Summary View on page 132). The affinity policy details page includes the following tabs:

- **Summary**: On the **Summary** tab, you can view the **Overview**, **Associations**, and **Compliance Status** sections. All the three sections display information related to the policy.

  The **Summary** tab view also includes options to **Update**, **Delete**, and **Re-Enforce** the policy. If any of the VMs become non-compliant, you can use this option to re-enforce the policy after fixing the issue for the non-compliance.

- **Entities**: On the **Entities** tab, you can view the details of the **VMs** and **Hosts** entities that are associated with the policy. The details displayed on the **Hosts** tab include the host name, cluster name, and the category used to associate, while the details displayed on the **VMs** tab include the VM name, host name, cluster name, category used to associate, and VM compliance status. If a VM is non-compliant, the cause of non-compliance is also displayed along with the status.

Figure 51: Affinity Policies Details view

Non-Migratable Hosts

VMs with GPU, CPU passthrough, PCI passthrough, and host affinity policies are not migrated to other hosts in the cluster. Such VMs are treated in a different manner in scenarios where VMs are required to migrate to other hosts in the cluster.

Table 8: Scenarios Where VMs Are Required to Migrate to Other Hosts

Scenario	Behavior
One-click upgrade	VM is powered off.
Life-cycle management (LCM)	Pre-check for LCM fails and the VMs are not migrated.
Rolling restart	VM is powered off.
AHV host maintenance mode	Use the tunable option to shut down the VMs while putting the node in maintenance mode. For more information, see Putting a Node into Maintenance Mode using CLI on page 34.

Performing Power Operations on VMs by Using Nutanix Guest Tools (aCLI)

You can initiate safe and graceful power operations such as soft shutdown and restart of the VMs running on the AHV hosts by using the aCLI. Nutanix Guest Tools (NGT) initiates and performs the soft shutdown and restart operations within the VM. This workflow ensures a safe and graceful shutdown or restart of the VM. You can create a pre-shutdown script that you can choose to run before a shutdown or restart of the VM. In the pre-shutdown script, include any tasks or checks that you want to run before a VM is shut down or restarted. You can choose to cancel the power operation if the pre-shutdown script fails. If the script fails, an alert (guest_agent_alert) is generated in the Prism web console.

Before you begin

Ensure that you have met the following prerequisites before you initiate the power operations:

1. NGT is enabled on the VM. All operating systems that NGT supports are supported for this feature.
2. NGT version running on the Controller VM and guest VM is the same.
3. (Optional) If you want to run a pre-shutdown script, place the script in the following locations depending on your VMs:

   - Windows VMs: installed_dir\scripts\power_off.bat

     The file name of the script must be **power_off.bat**.

   - Linux VMs: installed_dir/scripts/power_off

     The file name of the script must be **power_off**.

About this task

> Note: You can also perform these power operations by using the V3 API calls. For more information, see developer.nutanix.com.

Perform the following steps to initiate the power operations:

Procedure

1. Log on to a Controller VM with SSH.
2. Do one of the following:
   - » Soft shut down the VM.

     ```
     nutanix@cvm$ acli vm.guest_shutdown vm_name enable_script_exec=[true or false]
      fail_on_script_failure=[true or false]
     ```

     Replace *vm_name* with the name of the VM.
   - » Restart the VM.

     ```
     nutanix@cvm$ acli vm.guest_reboot vm_name enable_script_exec=[true or false]
      fail_on_script_failure=[true or false]
     ```

     Replace *vm_name* with the name of the VM.

   Set the value of enable_script_exec to **true** to run your pre-shutdown script and set the value of fail_on_script_failure to **true** to cancel the power operation if the pre-shutdown script fails.

UEFI Support for VM

UEFI firmware is a successor to legacy BIOS firmware that supports larger hard drives, faster boot time, and provides more security features.

VMs with UEFI firmware have the following advantages:

- Boot faster
- Avoid legacy option ROM address constraints
- Include robust reliability and fault management
- Use UEFI drivers

> Note:
>
> - Nutanix supports the starting of VMs with UEFI firmware in an AHV cluster. However, if a VM is added to a protection domain and later restored on a different cluster, the VM loses boot configuration. To restore the lost boot configuration, see Setting up Boot Device.
> - Nutanix also provides limited support for VMs migrated from a Hyper-V cluster.

You can create or update VMs with UEFI firmware by using `acli` commands, Prism Element web console, or Prism Central web console. For more information about creating a VM by using the Prism Element web console or Prism Central web console, see Creating a VM (AHV) on page 86. For information about creating a VM by using aCLI, see Creating UEFI VMs by Using aCLI on page 136.

> Note: If you are creating a VM by using aCLI commands, you can define the location of the storage container for UEFI firmware and variables. Prism Element web console or Prism Central web console does not provide the option to define the storage container to store UEFI firmware and variables.

For more information about the supported OSes for the guest VMs, see the *AHV Guest OS* section in the *Compatibility Matrix* document.

Creating UEFI VMs by Using aCLI

In AHV clusters, you can create a virtual machine (VM) to start with UEFI firmware by using Acropolis CLI (aCLI). This topic describes the procedure to create a VM by using aCLI. See the "Creating a VM (AHV)" topic for information about how to create a VM by using the Prism Element web console.

Before you begin

Ensure that the VM has an empty vDisk.

About this task

Perform the following procedure to create a UEFI VM by using aCLI:

Procedure

1. Log on to any Controller VM in the cluster with SSH.
2. Create a UEFI VM.

   ```
   nutanix@cvm$ acli vm.create vm-name uefi_boot=true
   ```

   A VM is created with UEFI firmware. Replace `vm-name` with a name of your choice for the VM. By default, the UEFI firmware and variables are stored in an NVRAM container. If you would like to specify a location of the NVRAM storage container to store the UEFI firmware and variables, do so by running the following command.

   ```
   nutanix@cvm$ acli vm.create vm-name uefi_boot=true nvram_container=NutanixManagementShare
   ```

   Replace `NutanixManagementShare` with a storage container in which you want to store the UEFI variables.

   The UEFI variables are stored in a default NVRAM container. Nutanix recommends you to choose a storage container with at least RF2 storage policy to ensure the VM high availability for node failure scenarios. For more information about RF2 storage policy, see *Failure and Recovery Scenarios* in the *Prism Web Console Guide* document.

   > Note: When you update the location of the storage container, clear the UEFI configuration and update the location of `nvram_container` to a container of your choice.

What to do next

Go to the UEFI BIOS menu and configure the UEFI firmware settings. For more information about accessing and setting the UEFI firmware, see Getting Familiar with UEFI Firmware Menu on page 136.

Getting Familiar with UEFI Firmware Menu

After you launch a VM console from the Prism Element web console, the UEFI firmware menu allows you to do the following tasks for the VM.

- Changing default boot resolution
- Setting up boot device
- Changing boot-time value

Changing Boot Resolution

You can change the default boot resolution of your Windows VM from the UEFI firmware menu.

Before you begin

Ensure that the VM is in powered on state.

About this task

Perform the following procedure to change the default boot resolution of your Windows VM by using the UEFI firmware menu.

Procedure

1. Log on to the Prism Element web console.
2. Launch the console for the VM.
   For more details about launching console for the VM, see *Managing a VM (AHV)* section in *Prism Web Console Guide*.
3. To go to the UEFI firmware menu, press the **F2** keys on your keyboard.

   Tip: To enter UEFI menu, open the VM console, select **Reset** in the **Power off/Reset VM** dialog box, and immediately press **F2** when the VM starts to boot.

   Important: Resetting the VM causes a downtime. We suggest that you reset the VM only during off-production hours or during a maintenance period.

   Figure 52: UEFI Firmware Menu

4. Use the up or down arrow key to go to **Device Manager** and press **Enter**.
   The **Device Manager** page appears.

5. In the **Device Manager** screen, use the up or down arrow key to go to **OVMF Platform Configuration** and press **Enter**.

Figure 53: OVMF Settings

The **OVMF Settings** page appears.

6. In the **OVMF Settings** page, use the up or down arrow key to go to the **Change Preferred** field and use the right or left arrow key to increase or decrease the boot resolution.

   The default boot resolution is 1280X1024.

7. Do one of the following.

   » To save the changed resolution, press the **F10** key.

   » To go back to the previous screen, press the **Esc** key.

8. Select **Reset** and click **Submit** in the Power off/Reset dialog box to restart the VM.
   After you restart the VM, the OS displays the changed resolution.

Setting up Boot Device

You cannot set the boot order for UEFI VMs by using the aCLI, Prism Central web console, or Prism Element web console. You can change the boot device for a UEFI VM by using the UEFI firmware menu.

Before you begin

Ensure that the VM is in powered on state.

Procedure

1. Log on to the Prism Element web console.

2. Launch the console for the VM.

   For more details about launching console for the VM, see *Managing a VM (AHV)* section in *Prism Web Console Guide*.

3. To go to the UEFI firmware menu, press the **F2** keys on your keyboard.

> Tip: To enter UEFI menu, open the VM console, select **Reset** in the **Power off/Reset VM** dialog box, and immediately press **F2** when the VM starts to boot.

> Important: Resetting the VM causes a downtime. We suggest that you reset the VM only during off-production hours or during a maintenance period.

4. Use the up or down arrow key to go to **Boot Manager** and press **Enter**.
   The **Boot Manager** screen displays the list of available boot devices in the cluster.

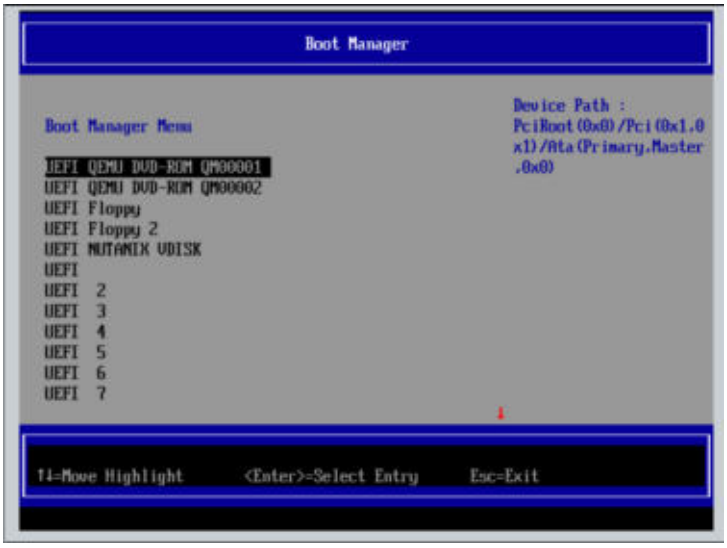


Figure 54: Boot Manager

5. In the **Boot Manager** screen, use the up or down arrow key to select the boot device and press **Enter**.
   The boot device is saved. After you select and save the boot device, the VM boots up with the new boot device.

6. To go back to the previous screen, press **Esc**.

Changing Boot Time-Out Value

The boot time-out value determines how long the boot menu is displayed (in seconds) before the default boot entry is loaded to the VM. This topic describes the procedure to change the default boot-time value of 0 seconds.

About this task

Ensure that the VM is in powered on state.

Procedure

1. Log on to the Prism Element web console.

2. Launch the console for the VM.
   For more details about launching console for the VM, see *Managing a VM (AHV)* section in *Prism Web Console Guide*.

3. To go to the UEFI firmware menu, press the **F2** keys on your keyboard.

> Tip: To enter UEFI menu, open the VM console, select **Reset** in the **Power off/Reset VM** dialog box, and immediately press **F2** when the VM starts to boot.

> Important: Resetting the VM causes a downtime. We suggest that you reset the VM only during off-production hours or during a maintenance period.

4. Use the up or down arrow key to go to **Boot Maintenance Manager** and press **Enter**.

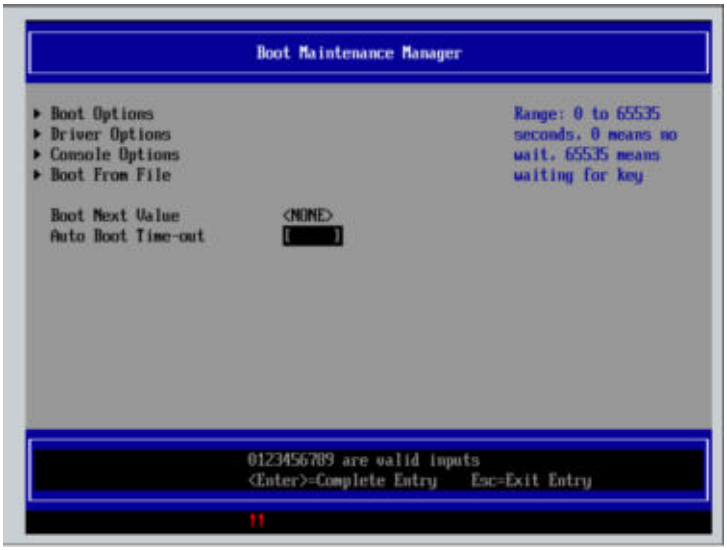


Figure 55: Boot Maintenance Manager

5. In the **Boot Maintenance Manager** screen, use the up or down arrow key to go to the **Auto Boot Time-out** field.
   The default boot-time value is 0 seconds.

6. In the **Auto Boot Time-out** field, enter the boot-time value and press **Enter**.

   > Note: The valid boot-time value ranges from 1 second to 9 seconds.

   The boot-time value is changed. The VM starts after the defined boot-time value.

7. To go back to the previous screen, press **Esc**.

Secure Boot Support for VMs

The pre-operating system environment is vulnerable to attacks by possible malicious loaders. Secure boot addresses this vulnerability with UEFI secure boot using policies present in the firmware along with certificates, to ensure that only properly signed and authenticated components are allowed to execute.

Supported Operating Systems

For more information about the supported OSes for the guest VMs, see the *AHV Guest OS* section in the *Compatibility Matrix* document.

Secure Boot Considerations

This section provides the limitations and requirements to use Secure Boot.

Limitations

Secure Boot for guest VMs has the following limitation:

- Nutanix does not support converting a VM that uses IDE disks or legacy BIOS to VMs that use Secure Boot.
- The minimum supported version of the Nutanix VirtIO package for Secure boot-enabled VMs is 1.1.6.
- vNIC hotplug of a VM is not supported.

Requirements

Following are the requirements for Secure Boot:

- Secure Boot is supported only on the Q35 machine type.

Creating/Updating a VM with Secure Boot Enabled

You can enable Secure Boot with UEFI firmware, either while creating a VM or while updating a VM by using aCLI commands or Prism Element web console.

See Creating a VM (AHV) on page 86 for instructions about how to enable Secure Boot by using the Prism Element web console.

Creating a VM with Secure Boot Enabled

About this task

To create a VM with Secure Boot enabled:

Procedure

1. Log on to any Controller VM in the cluster with SSH.
2. To create a VM with Secure Boot enabled:

```
nutanix@cvm$ acli vm.create  <vm_name> secure_boot=true machine_type=q35
```

> Note: Specifying the machine type is required to enable the secure boot feature. UEFI is enabled by default when the Secure Boot feature is enabled.

Updating a VM to Enable Secure Boot

About this task

To update a VM to enable Secure Boot:

Procedure

1. Log on to any Controller VM in the cluster with SSH.

2. To update a VM to enable Secure Boot, ensure that the VM is powered off.

```
nutanix@cvm$ acli vm.update <vm_name> secure_boot=true machine_type=q35
```

> Note:
>
> - If you disable the secure boot flag alone, the machine type remains q35, unless you disable that flag explicitly.
> - UEFI is enabled by default when the Secure Boot feature is enabled. Disabling Secure Boot does not revert the UEFI flags.

Virtual Machine Network Management

Virtual machine network management involves configuring connectivity for guest VMs through virtual switches, VLANs, and VPCs.

For information about creating or updating a virtual switch and other VM network options, see *Network*in the *Prism Central Guide*. Virtual switch creation and updates are also covered in *Prism Web Console Guide*.

Virtual Machine Memory and CPU Hot-Plug Configurations

Memory and CPUs are hot-pluggable on guest VMs running on AHV. You can increase the memory allocation and the number of CPUs on your VMs while the VMs are powered on. You can change the number of vCPUs (sockets) while the VMs are powered on. However, you cannot change the number of cores per socket while the VMs are powered on.

> Note: You cannot decrease the memory allocation and the number of CPUs on your VMs while the VMs are powered on.

You can change the memory and CPU configuration of your VMs by using the Acropolis CLI (aCLI), Prism Element (see *Managing a VM (AHV)* in the Prism Web Console Guide), or Prism Central (see *Managing a VM (AHV and Self Service)* in the Prism Central Guide).

See the AHV Guest OS Compatibility Matrix for information about operating systems on which you can hot plug memory and CPUs.

Memory OS Limitations

1. On Linux operating systems, the Linux kernel might not make the hot-plugged memory online. If the memory is not online, you cannot use the new memory. Perform the following procedure to make the memory online.
   1. Identify the memory block that is offline.

      Display the status of all of the memory.

      ```
      $ cat /sys/devices/system/memory/memoryXXX/state
      ```

      Display the state of a specific memory block.

      ```
      $ grep line /sys/devices/system/memory/*/state
      ```

   2. Make the memory online.

      ```
      $ echo online > /sys/devices/system/memory/memoryXXX/state
      ```

2. If your VM has CentoOS 7.2 as the guest OS and less than 3 GB memory, hot plugging more memory to that VM so that the final memory is greater than 3 GB, results in a memory-

overflow condition. To resolve the issue, restart the guest OS (CentOS 7.2) with the following setting:

```
swiotlb=force
```

CPU OS Limitation

On CentOS operating systems, if the hot-plugged CPUs are not displayed in `/proc/cpuinfo`, you might have to bring the CPUs online. For each hot-plugged CPU, run the following command to bring the CPU online.

```
$ echo 1 > /sys/devices/system/cpu/cpu<n>/online
```

Replace `<n>` with the number of the hot plugged CPU.

Hot-Plugging the Memory and CPUs on Virtual Machines (AHV)

About this task

Perform the following procedure to hot plug the memory and CPUs on the AHV VMs.

Procedure

1. Log on the Controller VM with SSH.
2. Update the memory allocation for the VM.

   ```
   nutanix@cvm$ acli vm.update vm-name memory=new_memory_size
   ```

   Replace `vm-name` with the name of the VM and `new_memory_size` with the memory size.
3. Update the number of CPUs on the VM.

   ```
   nutanix@cvm$ acli vm.update vm-name num_vcpus=n
   ```

   Replace `vm-name` with the name of the VM and `n` with the number of CPUs.

Virtual Machine Memory Management (vNUMA)

AHV hosts support Virtual Non-uniform Memory Access (vNUMA) on virtual machines. You can enable vNUMA on VMs when you create or modify the VMs to optimize memory performance.

Non-uniform Memory Access (NUMA)

In a NUMA topology, memory access times of a VM are dependent on the memory location relative to a processor. A VM accesses memory local to a processor faster than the non-local memory. You can achieve optimal resource utilization if both CPU and memory from the same physical NUMA node is used. Memory latency is introduced if you are running the CPU on one NUMA node (for example, node 0) and the VM accesses the memory from another node (node 1). Ensure that the virtual topology of VMs matches the physical hardware topology to achieve minimum memory latency.

Virtual Non-uniform Memory Access (vNUMA)

vNUMA optimizes memory performance of virtual machines that require more vCPUs or memory than the capacity of a single physical NUMA node. In a vNUMA topology, you can create multiple vNUMA nodes where each vNUMA node includes vCPUs and virtual RAM. When you assign a vNUMA node to a physical NUMA node, the vCPUs can intelligently determine the memory latency (high or low). Low memory latency within a vNUMA node results in low latency within a physical NUMA node.

Enabling vNUMA on Virtual Machines

Before you begin

Before you enable vNUMA, see *AHV Best Practices Guide* under *Solutions Documentation*.

About this task

Perform the following procedure to enable vNUMA on your VMs running on the AHV hosts.

Procedure

1. Log on to a Controller VM with SSH.
2. Check how many NUMA nodes are available on each AHV host in the cluster.

```
nutanix@cvm$ hostssh "numactl --hardware"
```

An output similar to the following is displayed:

```
============= 10.x.x.x ============
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 6 7 16 17 18 19 20 21 22 23
node 0 size: 128837 MB
node 0 free: 862 MB
node 1 cpus: 8 9 10 11 12 13 14 15 24 25 26 27 28 29 30 31
node 1 size: 129021 MB
node 1 free: 352 MB
node distances:
node   0   1
  0:  10  21
  1:  21  10
============= 10.x.x.x ============
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 12 13 14 15 16 17
node 0 size: 128859 MB
node 0 free: 1076 MB
node 1 cpus: 6 7 8 9 10 11 18 19 20 21 22 23
node 1 size: 129000 MB
node 1 free: 436 MB
node distances:
node   0   1
  0:  10  21
  1:  21  10
============= 10.x.x.x ============
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 12 13 14 15 16 17
node 0 size: 128859 MB
node 0 free: 701 MB
node 1 cpus: 6 7 8 9 10 11 18 19 20 21 22 23
node 1 size: 129000 MB
node 1 free: 357 MB
node distances:
node   0   1
  0:  10  21
  1:  21  10
============= 10.x.x.x ============
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 12 13 14 15 16 17
node 0 size: 128838 MB
node 0 free: 1274 MB
node 1 cpus: 6 7 8 9 10 11 18 19 20 21 22 23
node 1 size: 129021 MB
```

```
node 1 free: 424 MB
node distances:
node   0   1
  0:  10  21
  1:  21  10
============= 10.x.x.x ============
available: 2 nodes (0-1)
node 0 cpus: 0 1 2 3 4 5 12 13 14 15 16 17
node 0 size: 128837 MB
node 0 free: 577 MB
node 1 cpus: 6 7 8 9 10 11 18 19 20 21 22 23
node 1 size: 129021 MB
node 1 free: 612 MB
node distances:
node   0   1
  0:  10  21
  1:  21  10
```

The example output shows that each AHV host has two NUMA nodes.

3. Do one of the following:

   » Enable vNUMA if you are creating a VM.

   ```
   nutanix@cvm$ acli vm.create <vm_name> num_vcpus=x \
   num_cores_per_vcpu=x memory=xG \
   num_vnuma_nodes=x
   ```

   » Enable vNUMA if you are modifying an existing VM.

   ```
   nutanix@cvm$ acli vm.update <vm_name> \
   num_vnuma_nodes=x
   ```

   Replace `<vm_name>` with the name of the VM on which you want to enable vNUMA or vUMA. Replace `x` with the values for the following indicated parameters:

   - num_vcpus: Type the number of vCPUs for the VM.
   - num_cores_per_vcpu: Type the number of cores per vCPU.
   - memory: Type the memory in GB for the VM.
   - num_vnuma_nodes: Type the number of vNUMA nodes for the VM.

   For example:

   ```
   nutanix@cvm$ acli vm.create test_vm num_vcpus=20 memory=150G num_vnuma_nodes=2
   ```

   This command creates a VM with 2 vNUMA nodes, 10 vCPUs and 75 GB memory for each vNUMA node.

GPU and vGPU Support

AHV supports GPU-accelerated computing for guest VMs. You can configure either GPU pass-through or a virtual GPU.

> Note: You can configure either pass-through or a vGPU for a guest VM but not both.

This guide describes the concepts related to the GPU and vGPU support in AHV. For the configuration procedures, see the *Prism Web Console Guide*.

For driver installation instructions, see the *NVIDIA Grid Host Driver for Nutanix AHV Installation Guide*.

> Note: VMs with GPU are not migrated to other hosts in the cluster. For more information, see Non-Migratable Hosts on page 134.

Supported GPUs

The following GPUs are supported:

> Note: These GPUs are supported only by the AHV version that is bundled with the AOS release.

- NVIDIA® A16
- NVIDIA® Tesla® M10
- NVIDIA® Tesla® M60
- NVIDIA® Tesla® P40
- NVIDIA® Tesla® P100
- NVIDIA® Tesla® P4
- NVIDIA® Tesla® V100 16 GB
- NVIDIA® Tesla® V100 32 GB
- NVIDIA® Tesla® V100S 32 GB
- NVIDIA® Tesla® T4 16 GB
- NVIDIA® Quadro® RTX 6000
- NVIDIA® Quadro® RTX 8000
- NVIDIA® Ampere® A100

GPU Pass-Through for Guest VMs

AHV hosts support GPU pass-through for guest VMs, allowing applications on VMs direct access to GPU resources. The Nutanix user interfaces provide a cluster-wide view of GPUs, allowing you to allocate any available GPU to a VM. You can also allocate multiple GPUs to a VM. However, in a pass-through configuration, only one VM can use a GPU at any given time.

Host Selection Criteria for VMs with GPU Pass-Through

When you power on a VM with GPU pass-through, the VM is started on the host that has the specified GPU, provided that the Acropolis Dynamic Scheduler determines that the host has sufficient resources to run the VM. If the specified GPU is available on more than one host, the Acropolis Dynamic Scheduler ensures that a host with sufficient resources is selected. If sufficient resources are not available on any host with the specified GPU, the VM is not powered on.

If you allocate multiple GPUs to a VM, the VM is started on a host if, in addition to satisfying Acropolis Dynamic Scheduler requirements, the host has all of the GPUs that are specified for the VM.

If you want a VM to always use a GPU on a specific host, configure host affinity for the VM.

Support for Graphics and Compute Modes

AHV supports running GPU cards in either graphics mode or compute mode. If a GPU is running in compute mode, Nutanix user interfaces indicate the mode by appending the string `compute` to the model name. No string is appended if a GPU is running in the default graphics mode.

Switching Between Graphics and Compute Modes

If you want to change the mode of the firmware on a GPU, put the host in maintenance mode, and then flash the GPU manually by logging on to the AHV host and performing standard procedures as documented for Linux VMs by the vendor of the GPU card.

Typically, you restart the host immediately after you flash the GPU. After restarting the host, redo the GPU configuration on the affected VM, and then start the VM. For example, consider that you want to re-flash an NVIDIA Tesla® M60 GPU that is running in graphics mode. The Prism web console identifies the card as an `NVIDIA Tesla M60` GPU. After you re-flash the GPU to run in compute mode and restart the host, redo the GPU configuration on the affected VMs by adding back the GPU, which is now identified as an `NVIDIA Tesla M60.compute` GPU, and then start the VM.

Supported GPU Cards

For a list of supported GPUs, see Supported GPUs on page 146.

Limitations

GPU pass-through support has the following limitations:

- Live migration of VMs with a GPU configuration is not supported. Live migration of VMs is necessary when the BIOS, BMC, and the hypervisor on the host are being upgraded. During these upgrades, VMs that have a GPU configuration are powered off and then powered on automatically when the node is back up.
- VM pause and resume are not supported.
- You cannot hot add VM memory if the VM is using a GPU.
- Hot add and hot remove support is not available for GPUs.
- You can change the GPU configuration of a VM only when the VM is turned off.
- The Prism web console does not support console access for VMs that are configured with GPU pass-through. Before you configure GPU pass-through for a VM, set up an alternative means to access the VM. For example, enable remote access over RDP.

  Removing GPU pass-through from a VM restores console access to the VM through the Prism web console.

Configuring GPU Pass-Through

For information about configuring GPU pass-through for guest VMs, see Creating a VM (AHV) in the "Virtual Machine Management" chapter of the *Prism Web Console Guide*.

NVIDIA GRID Virtual GPU Support on AHV

AHV supports NVIDIA GRID technology, which enables multiple guest VMs to use the same physical GPU concurrently. Concurrent use is made possible by dividing a physical GPU into discrete virtual GPUs (vGPUs) and allocating those vGPUs to guest VMs. Each vGPU is allocated a fixed range of the physical GPUs framebuffer and uses all the GPU processing cores in a time-sliced manner.

Virtual GPUs are of different types (vGPU types are also called vGPU profiles) and differ by the amount of physical GPU resources allocated to them and the class of workload that they target. The number of vGPUs into which a single physical GPU can be divided therefore depends on the vGPU profile that is used on a physical GPU.

Each physical GPU supports more than one vGPU profile, but a physical GPU cannot run multiple vGPU profiles concurrently. After a vGPU of a given profile is created on a physical

GPU (that is, after a vGPU is allocated to a VM that is powered on), the GPU is restricted to that vGPU profile until it is freed up completely. To understand this behavior, consider that you configure a VM to use an M60-1Q vGPU. When the VM is powering on, it is allocated an M60-1Q vGPU instance only if a physical GPU that supports M60-1Q is either unused or already running the M60-1Q profile and can accommodate the requested vGPU.

If an entire physical GPU that supports M60-1Q is free at the time the VM is powering on, an M60-1Q vGPU instance is created for the VM on the GPU, and that profile is locked on the GPU. In other words, until the physical GPU is completely freed up again, only M60-1Q vGPU instances can be created on that physical GPU (that is, only VMs configured with M60-1Q vGPUs can use that physical GPU).

> Note: NVIDIA does not support Windows Guest VMs on the C-series NVIDIA vGPU types. See the NVIDIA documentation on Virtual GPU software for more information.

NVIDIA Grid Host Drivers and License Installation

To enable guest VMs to use vGPUs on AHV, you must install NVIDIA drivers on the guest VMs, install the NVIDIA GRID host driver on the hypervisor, and set up an NVIDIA GRID License Server.

See the NVIDIA Grid Host Driver for Nutanix AHV Installation Guide for details about the workflow to enable guest VMs to use vGPUs on AHV and the NVIDIA GRID host driver installation instructions.

vGPU Profile Licensing

vGPU profiles are licensed through an NVIDIA GRID license server. The choice of license depends on the type of vGPU that the applications running on the VM require. Licenses are available in various editions, and the vGPU profile that you want might be supported by more than one license edition.

> Note: If the specified license is not available on the licensing server, the VM starts up and functions normally, but the vGPU runs with reduced capability.

You must determine the vGPU profile that the VM requires, install an appropriate license on the licensing server, and configure the VM to use that license and vGPU type. For information about licensing for different vGPU types, see the NVIDIA GRID licensing documentation.

Guest VMs check out a license over the network when starting up and return the license when shutting down. As the VM is powering on, it checks out the license from the licensing server. When a license is checked back in, the vGPU is returned to the vGPU resource pool.

When powered on, guest VMs use a vGPU in the same way that they use a physical GPU that is passed through.

Supported GPU Cards

For a list of supported GPUs, see Supported GPUs on page 146.

High Availability Support for VMs with vGPUs

Nutanix conditionally supports high availability (HA) of VMs that have NVIDIA GRID vGPUs configured. You can restart a VM with vGPUs on another (failover) host which has compatible or identical vGPU resources available. The vGPU profile available on the failover host must be identical to the vGPU profile configured on the VM that needs HA. The cluster does not reserve any specific resources to guarantee High Availability for the VMs with vGPUs. The system attempts to restart the VM after an event. If the failover host has insufficient memory and vGPU resources for the VM to start, it fails to start after failover.

The following conditions are applicable to HA of VMs with vGPUs:

- Memory is not reserved for the VM on the failover host by the HA process. When the VM fails over, if sufficient memory is not available, the VM cannot power on.
- vGPU resource is not reserved on the failover host. When the VM fails over, if the required vGPU resources are not available on the failover host, the VM cannot power on.

Limitations for vGPU Support

vGPU support on AHV has the following limitations:

- You cannot hot-add memory to VMs that have a vGPU.
- The Prism web console does not support console access for VMs that are configured with a vGPU. Before you add a vGPU to a VM, set up an alternative means to access the VM. For example, enable remote access over RDP.

  Removing a vGPU from a VM restores console access to the VM through the Prism web console.

Console Support for VMs with vGPU

Like other VMs, you can access VMs with vGPUs using the console. Enable or disable console support for a VM with only one vGPU configured. Enabling console support for a VM with multiple vGPUs is not supported. By default, console support for a VM with vGPUs is disabled.

Recovery of vGPU-enabled VMs

With AHV, you can recover vGPU console-enabled guest VMs efficiently. When you perform DR of vGPU console-enabled guest VMs, the VMs recovers with the vGPU console. The guest VMs fail to recover when you perform cross-hypervisor disaster recovery (CHDR).

For AHV with minimum AOS versions 6.1, 6.0.2.4 and 5.20:

- vGPU-enabled VMs can be recovered when protected by protection domains in PD-based DR or protection policies in Leap based solutions using asynchronous, NearSync, or Synchronous (Leap only) replications.

  > Note: GPU Passthrough is not supported.

- If both site A and site B have the same GPU boards (and the same assignable vGPU profiles), failovers work seamlessly. However, with protection domains, no additional steps are required. GPU profiles are restored correctly and vGPU console settings persist after recovery. With Leap DR, vGPU console settings do not persist after recovery.
- If site A and site B have different GPU boards and vGPU profiles, you must manually remove the vGPU profile before you power on the VM in site B.

The vGPU console settings are persistent after recovery and all failovers are supported for the following:

Table 9: Persistent vGPU Console Settings with Failover Support

Recovery using	For vGPU enabled AHV VMs
Protection domain based DR	Yes
VMware SRM with Nutanix SRA	Not applicable

For information about the behavior See the *Recovery of vGPU-enabled VMs* topic in the *Data Protection and Recovery with Prism Element* guide.

See Enabling or Disabling Console Support for vGPU VMs on page 158 for more information about configuring the support.

For SRA and SRM support, see the Nutanix SRA documentation.

ADS support for VMs with vGPUs

AHV supports Acropolis Dynamic Scheduling (ADS) for VMs with vGPUs.

> Note: ADS support requires live migration of VMs with vGPU be operational in the cluster. See *Live Migration of VMs with vGPUs* above for minimum NVIDIA and AOS versions that support live migration of VMs with vGPUs.

When a number of VMs with vGPUs are running on a host and you enable ADS support for the cluster, the Lazan manager invokes VM migration tasks to resolve resource hotspots or fragmentation in the cluster to power on incoming vGPU VMs. The Lazan manager can migrate vGPU-enabled VMs to other hosts in the cluster only if:

- The other hosts support compatible or identical vGPU resources as the source host (hosting the vGPU-enabled VMs).
- The host affinity is not set for the vGPU-enabled VM.

For more information about limitations, see Live Migration of vGPU-enabled VMs on page 156 and Limitations of Live Migration Support on page 157.

For more information about ADS, see Acropolis Dynamic Scheduling in AHV on page 7.

Multiple Virtual GPU Support

Prism Central and Prism Element Web Console can deploy VMs with multiple virtual GPU instances. This support harnesses the capabilities of NVIDIA GRID virtual GPU (vGPU) support for multiple vGPU instances for a single VM.

> Note: Multiple vGPUs on the same VM are supported on NVIDIA Virtual GPU software version 10.1 (440.53) or later.

You can deploy virtual GPUs of different types. A single physical GPU can be divided into the number of vGPUs depending on the type of vGPU profile that is used on the physical GPU. Each physical GPU on a GPU board supports more than one type of vGPU profile. For example, a Tesla® M60 GPU device provides different types of vGPU profiles like M60-0Q, M60-1Q, M60-2Q, M60-4Q, and M60-8Q.

You can only add multiple vGPUs of the same type of vGPU profile to a single VM. For example, consider that you configure a VM on a Node that has one NVidia Tesla® M60 GPU board. Tesla® M60 provides two physical GPUs, each supporting one M60-8Q (profile) vGPU, thus supporting a total of two M60-8Q vGPUs for the entire host.

For restrictions on configuring multiple vGPUs on the same VM, see Restrictions for Multiple vGPU Support on page 150.

For steps to add multiple vGPUs to the same VM, see *Creating a VM (AHV)* and *Adding Multiple vGPUs to a VM* in *Prism Web Console Guide* or *Prism Central Guide*.

Restrictions for Multiple vGPU Support

You can configure multiple vGPUs subject to the following restrictions:

- All the vGPUs that you assign to one VM must be of the same type. In the aforesaid example, with the Tesla® M60 GPU device, you can assign multiple M60-8Q vGPU profiles. You cannot assign one vGPU of the M60-1Q type and another vGPU of the M60-8Q type.

  > Note: You can configure any number of vGPUs of the same type on a VM. However, the cluster calculates a maximum number of vGPUs of the same type per VM. This number is defined as max_instances_per_vm. This number is variable and changes based on the GPU resources available in the cluster and the number of VMs deployed. If the number of vGPUs of a specific type that you configured on a VM exceeds the max_instances_per_vm number, then the VM fails to power on and the following error message is displayed:

  ```
  Operation failed: NoHostResources: No host has enough available GPU for VM <name of VM>(UUID
   of VM).
  You could try reducing the GPU allotment...
  ```

  When you configure multiple vGPUs on a VM, after you select the appropriate vGPU type for the first vGPU assignment, Prism (Prism Central and Prism Element Web Console)

automatically restricts the selection of vGPU type for subsequent vGPU assignments to the same VM.

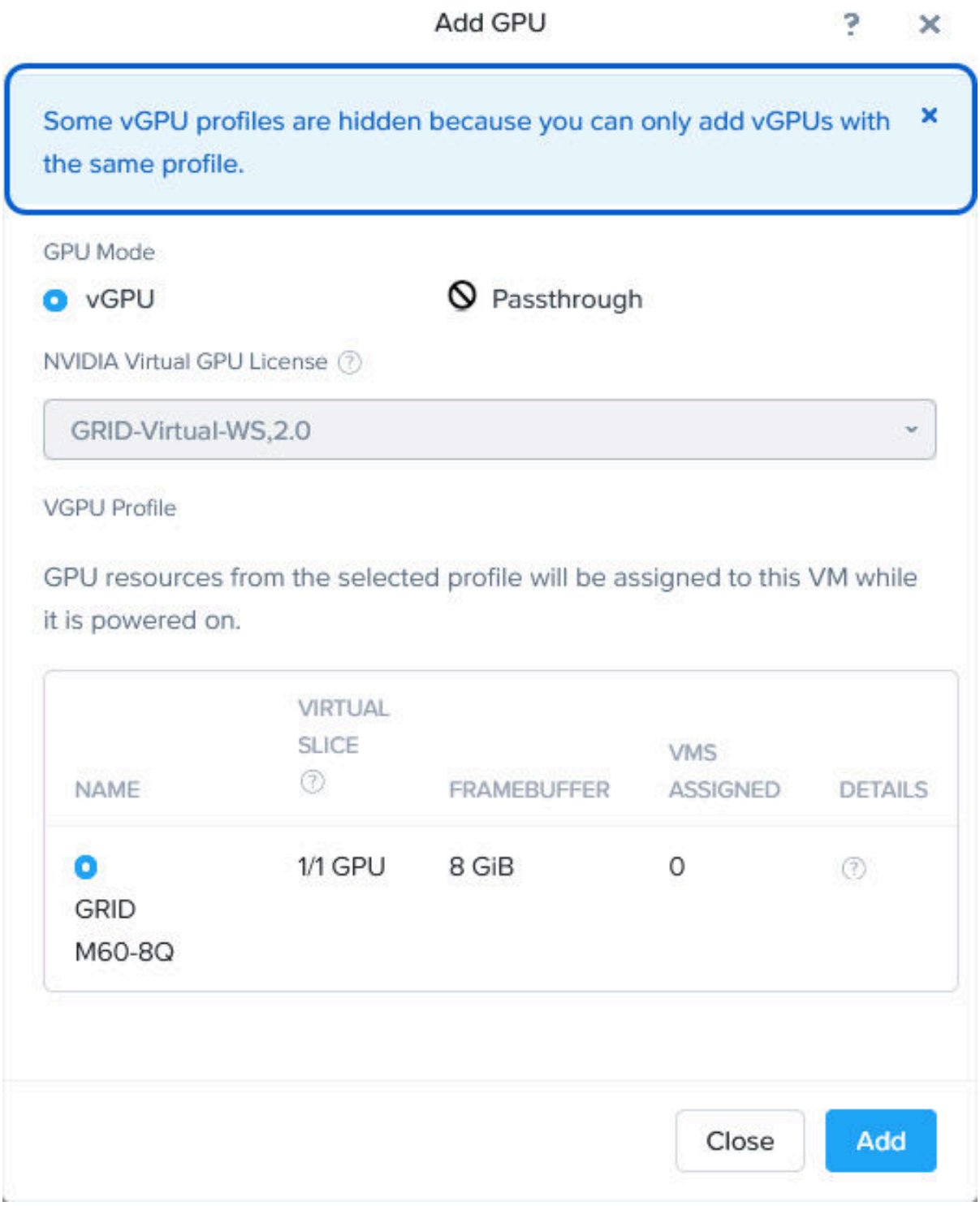


Figure 56: vGPU Type Restriction Message

Note:

You can use CLI (acli) to configure multiple vGPUs of multiple types to the same VM. See *Acropolis Command-Line Interface (aCLI)* for information about aCLI. Use the `vm.gpu_assign <vm.name> gpu=<gpu-type>` command multiple times, once for each vGPU, to add multiple vGPUs of multiple types to the same VM.

See the GPU board and software documentation for information about the combinations of the number and types of vGPUs profiles supported by the GPU resources installed in the cluster. For example, see the NVIDIA Virtual GPU Software Documentation for the vGPU type and number combinations on the Tesla® M60 board.

- Configure multiple vGPUs only of the highest type using Prism. The highest type of vGPU profile is based on the driver deployed in the cluster. In the aforesaid example, on a Tesla® M60 device, you can only configure multiple vGPUs of the M60-8Q type. Prism prevents you from configuring multiple vGPUs of any other type such as M60-2Q.

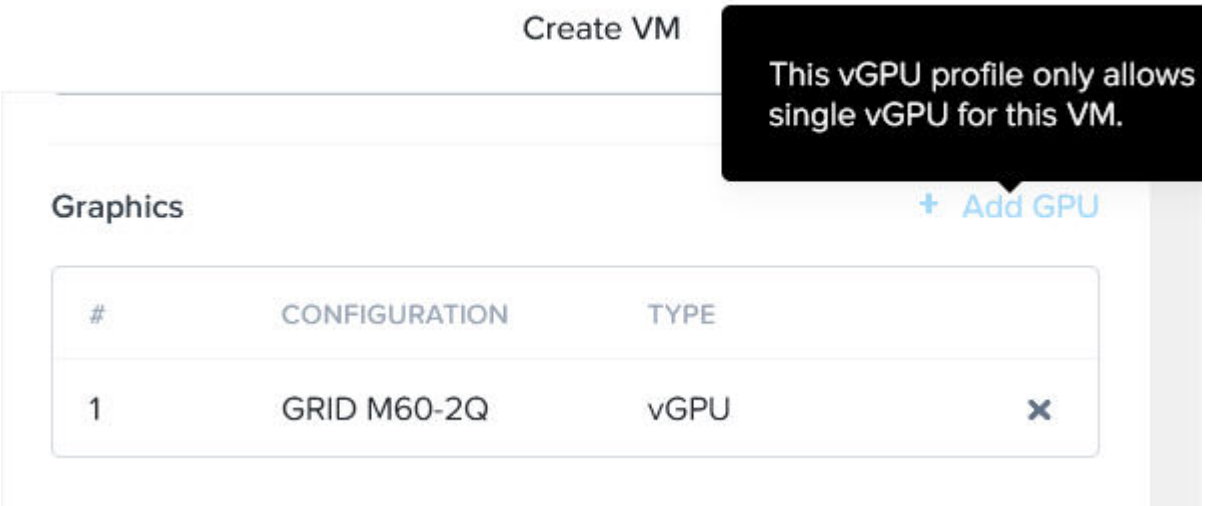


Figure 57: vGPU Type Restriction Message

Note:

You can use CLI (acli) to configure multiple vGPUs of other available types. See *Acropolis Command-Line Interface (aCLI)* for the aCLI information. Use the `vm.gpu_assign <vm.name> gpu=<gpu-type>` command multiple times, once for each vGPU, to configure multiple vGPUs of other available types.

See the GPU board and software documentation for more information.

- Configure either a passthrough GPU or vGPUs on the same VM. You cannot configure both passthrough GPU and vGPUs. Prism automatically disallows such configurations after the first GPU is configured.
- The VM powers on only if the requested type and number of vGPUs are available in the host.

  In the aforesaid example, the VM, which is configured with two M60-8Q vGPUs, fails to power on if another VM sharing the same GPU board is already using one M60-8Q vGPU. This is because the Tesla® M60 GPU board allows only two M60-8Q vGPUs. Of these, one is already used by another VM. Thus, the VM configured with two M60-8Q vGPUs fails to power on due to unavailability of required vGPUs.

- Multiple vGPUs on the same VM are supported on NVIDIA Virtual GPU software version 10.1 (440.53) or later. Ensure that the relevant GRID version license is installed and select it when you configure multiple vGPUs.

Adding Multiple vGPUs to the Same VM

About this task

You can add multiple vGPUs of the same vGPU type to:

- A new VM when you create it.
- An existing VM when you update it.

> Important:
>
> Before you add multiple vGPUs to the VM, see *Multiple Virtual GPU Support* and *Restrictions for Multiple vGPU Support* in the *AHV Admin Guide*.

After you add the first vGPU, do the following on the Create VM or Update VM dialog box (the main dialog box) to add more vGPUs:

Procedure

1. Click **Add GPU**.

2. In the Add GPU dialog box, click **Add**.

   The License field is grayed out because you cannot select a different license when you add a vGPU for the same VM.

   The **VGPU Profile** is also auto-selected because you can only select the additional vGPU of the same vGPU type as indicated by the message at the top of the dialog box.

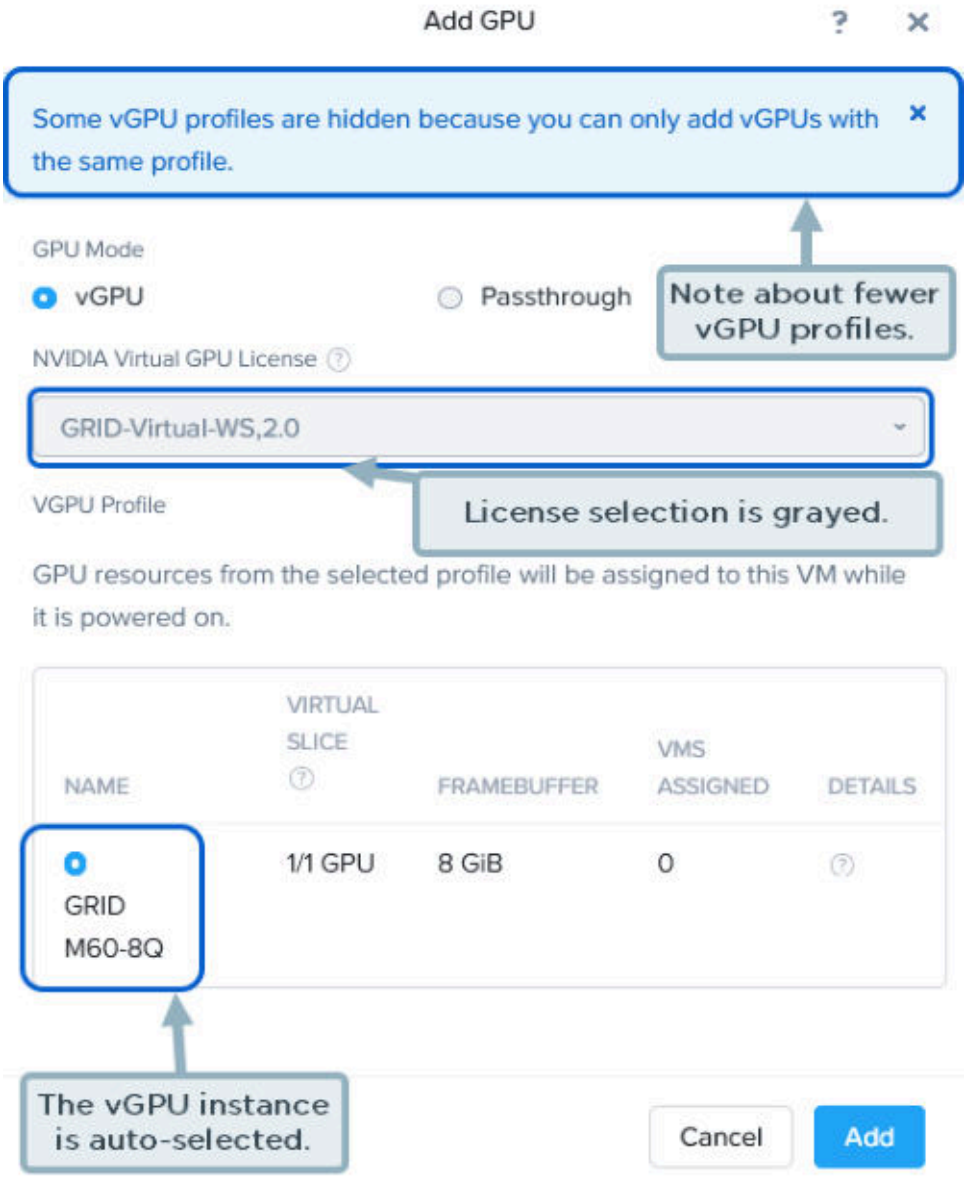


Figure 58: Add GPU for multiple vGPUs

3. In the main dialog box, you see the newly added vGPU.

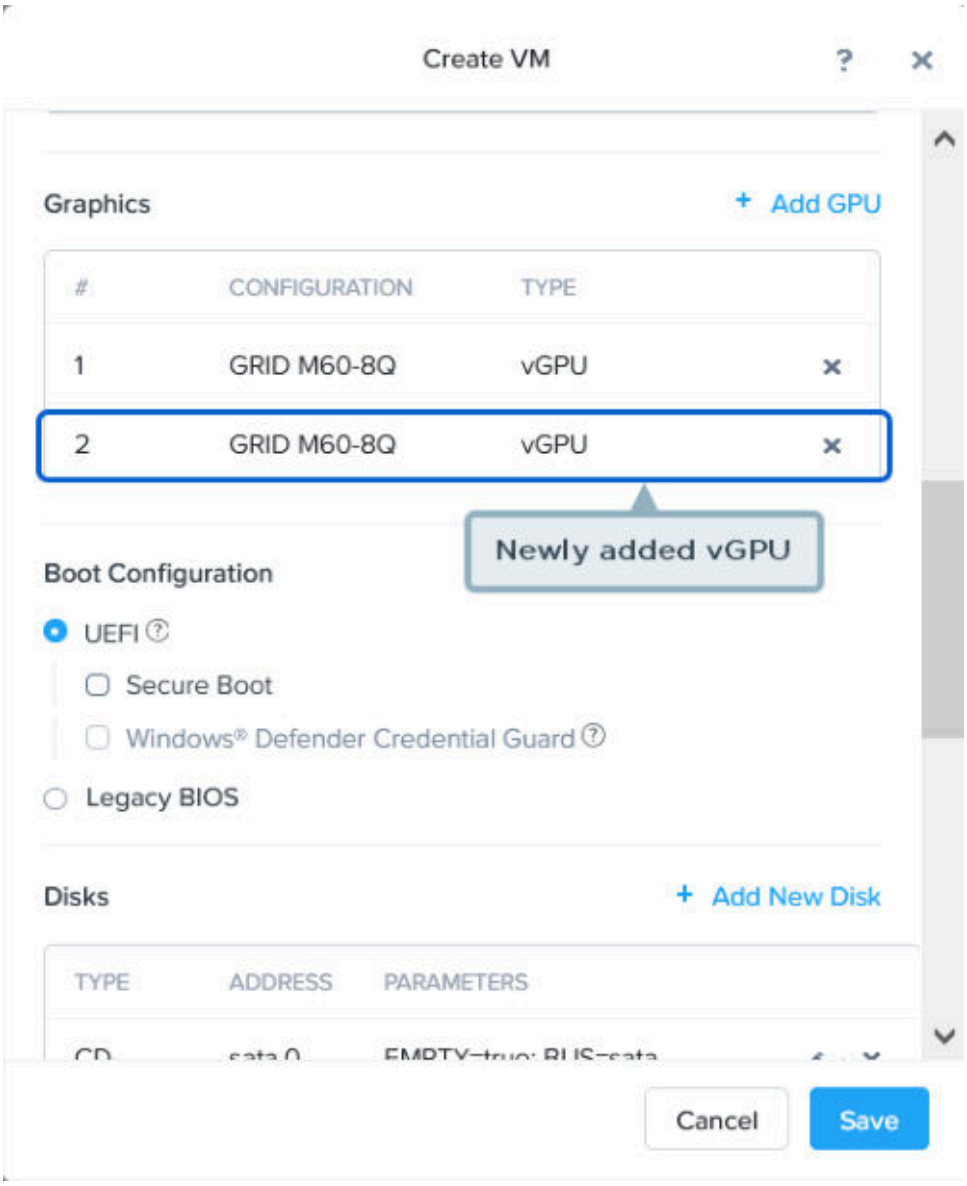


Figure 59: New vGPUs Added

4. Repeat the steps for each vGPU addition you want to make.

Live Migration of vGPU-enabled VMs

You can now perform live migration of VMs enabled with virtual GPUs (vGPU-enabled VM). The primary advantage of the live migration support is that unproductive downtime is avoided. Therefore, your vGPUs can continue to run while the VMs that are running the vGPUs are seamlessly migrated in the background. With very low stun times, as a graphics user, you barely notice the migration.

Note: Live migration of VMs with vGPUs is supported:

- For vGPUs created with minimum NVIDIA Virtual GPU software version 10.1 (440.53).
- Minimum AOS version that supports live migration within the same cluster is 5.18.1.

- Minimum AOS version that supports live migration across cluster is 6.1.

Important: In an HA event involving any GPU node, the node locality of the affected vGPU-enabled VMs is not restored after GPU node recovery. The affected vGPU-enabled VMs are not migrated back to their original GPU host intentionally to avoid extended VM stun time expected during migration. If vGPU-enabled VM node locality is required, migrate the affected vGPU-enabled VMs to the desired host manually.

Virtual GPU Limitations

Important frame buffer and VM stun time considerations are:

- The GPU board (for example, NVIDIA Tesla M60) vendor provides the information for maximum frame buffer size of vGPU types (for example, M60-8Q type) that can be configured on VMs. However, the actual frame buffer usage may be lower than the maximum sizes.
- The VM stun time depends on the number of vGPUs configured on the VM being migrated. Stun time may be longer in case of multiple vGPUs operating on the VM.

  The stun time also depends on the network factors such bandwidth available for use during the migration.

Live Migration Workflows

You can live migrate a vGPU-enabled VM to the following destinations:

- To another host within the same cluster. Both Prism Web Console (Prism Element) and Prism Central allow you to live migrate a vGPU-enabled VM to another host within the same cluster.
- To a host outside the cluster, that is, a host in another cluster. Only Prism Central allows you to migrate a vGPU-enabled VM to a host outside the cluster.)

For information about the steps to live migrate a vGPU-enabled VM, see the following:

- *Migrating Live a vGPU-enabled VM Within the Cluster* in the *Prism Web Console Guide*.
- *Migrating Live a vGPU-enabled VM Within the Cluster* in the *Prism Central Guide*.
- *Migrating Live a vGPU-enabled VM Outside the Cluster* in the *Prism Central Guide*.

Limitations of Live Migration Support

- Live migration is supported only for VMs configured with single or multiple virtual GPUs. It is not supported for VMs configured with passthrough GPUs.
- Live migration to a host in another cluster is supported only if the VM is protected by protection policy with Synchronous replication schedule.
- The target host for the migration must have adequate and available GPU resources, with the same vGPU types as configured for the VMs to be migrated, to support the vGPUs on the VMs that need to be migrated.

  See Restrictions for Multiple vGPU Support on page 150 for more details.
- The vGPU-enabled VMs that need to be migrated live cannot be protected with high availability.

- Ensure that the VM is not powered off.
- Ensure that you have the right GPU software license(an appropriate license of NVIDIA GRID software version) that supports live migration of vGPUs. The source and target hosts must have the same license type.

Enabling or Disabling Console Support for vGPU VMs

About this task

Enable or disable console support for a VM with only one vGPU configured. Enabling console support for a VM with multiple vGPUs is not supported. By default, console support for a VM with vGPUs is disabled.

To enable or disable console support for each VM with vGPUs, do the following:

Procedure

1. Run the following aCLI command to check if console support is enabled or disabled for the VM with vGPUs.

   ```
   acli> vm.get vm-name
   ```

   Where `vm-name` is the name of the VM for which you want to check the console support status.

   The step result includes the following parameter for the specified VM:

   ```
   gpu_console=False
   ```

   Where `False` indicates that console support is not enabled for the VM. This parameter is displayed as `True` when you enable console support for the VM. The default value for `gpu_console=` is `False` since console support is disabled by default.

   > Note: The console may not display the `gpu_console` parameter in the output of the `vm.get` command if the `gpu_console` parameter was not previously enabled.

2. Run the following aCLI command to enable or disable console support for the VM with vGPU:

   ```
   vm.update vm-name gpu_console=true | false
   ```

   Where:

   - `true`—indicates that you are enabling console support for the VM with vGPU.
   - `false`—indicates that you are disabling console support for the VM with vGPU.

3. Run the `vm.get` command to check if `gpu_console` value is `true` indicating that console support is enabled or `false` indicating that console support is disabled as you configured it.

   If the value indicated in the `vm.get` command output is not what is expected, then perform **Guest Shutdown** of the VM with vGPU. Next, run the `vm.on vm-name` aCLI command to turn the VM on again. Then run `vm.get` command and check the `gpu_console=` value.

4. Click a VM name in the VM table view to open the VM details page. Click **Launch Console**. The Console opens but only a black screen is displayed.

5. Click on the console screen. Click one of the following key combinations based on the operating system you are accessing the cluster from.

   » For Apple Mac OS: `Control+Command+2`

   » For MS Windows: `Ctrl+Alt+2`

   The console is fully enabled and displays the content.

PXE Configuration for AHV VMs

You can configure a VM to boot over the network in a Preboot eXecution Environment (PXE). Booting over the network is called PXE booting and does not require the use of installation media. When starting up, a PXE-enabled VM communicates with a DHCP server to obtain information about the boot file it requires.

Configuring PXE boot for an AHV VM involves performing the following steps:

- Configuring the VM to boot over the network.
- Configuring the PXE environment.

The procedure for configuring a VM to boot over the network is the same for managed and unmanaged networks. The procedure for configuring the PXE environment differs for the two network types, as follows:

- An unmanaged network does not perform IPAM functions and gives VMs direct access to an external Ethernet network. Therefore, the procedure for configuring the PXE environment for AHV VMs is the same as for a physical machine or a VM that is running on any other hypervisor. VMs obtain boot file information from the DHCP or PXE server on the external network.
- A managed network intercepts DHCP requests from AHV VMs and performs IP address management (IPAM) functions for the VMs. Therefore, you must add a TFTP server and the required boot file information to the configuration of the managed network. VMs obtain boot file information from this configuration.

A VM that is configured to use PXE boot boots over the network on subsequent restarts until the boot order of the VM is changed.

Configuring the PXE Environment for AHV VMs

The procedure for configuring the PXE environment for a VM on an unmanaged network is similar to the procedure for configuring a PXE environment for a physical machine on the external network and is beyond the scope of this document. This procedure configures a PXE environment for a VM in a managed network on an AHV host.

About this task

To configure a PXE environment for a VM on a managed network on an AHV host, do the following:

Procedure

1. Log on to the Prism web console, click the gear icon, and then click **Network Configuration** in the menu.
   The **Network Configuration** dialog box is displayed.

2. On the **Virtual Networks** tab, click the pencil icon shown for the network for which you want to configure a PXE environment.

   The VMs that require the PXE boot information must be on this network.

3. In the Update Network dialog box, do the following:

   a. Select the **Configure Domain Settings** check box and do the following in the fields shown in the domain settings sections:

      - In the **TFTP Server Name** field, specify the host name or IP address of the TFTP server.

        If you specify a host name in this field, make sure to also specify DNS settings in the **Domain Name Servers (comma separated)**, **Domain Search (comma separated)**, and **Domain Name** fields.

      - In the **Boot File Name** field, specify the boot file URL and boot file that the VMs must use. For example, tftp://`ip_address`/`boot_filename.bin`, where ip_address is the IP address (or host name, if you specify DNS settings) of the TFTP server and `boot_filename.bin` is the PXE boot file.

   b. Click **Save**.

4. Click **Close**.

Configuring a VM to Boot over a Network

To enable a VM to boot over the network, update the VM's boot device setting. Currently, the only user interface that enables you to perform this task is the Acropolis CLI (aCLI).

About this task

To configure a VM to boot from the network, do the following:

Procedure

1. Log on to any CVM in the cluster using SSH.

2. Create a VM.

   ```
   nutanix@cvm$ acli vm.create vm num_vcpus=num_vcpus memory=memory
   ```

   Replace `vm` with a name for the VM, and replace `num_vcpus` and `memory` with the number of vCPUs and amount of memory that you want to assign to the VM, respectively.

   For example, create a VM named nw-boot-vm.

   ```
   nutanix@cvm$ acli vm.create nw-boot-vm num_vcpus=1 memory=512
   ```

3. Create a virtual interface for the VM and place it on a network.

   ```
   nutanix@cvm$ acli vm.nic_create vm network=network
   ```

   Replace `vm` with the name of the VM and replace `network` with the name of the network. If the network is an unmanaged network, make sure that a DHCP server and the boot file that the VM requires are available on the network. If the network is a managed network, configure the

DHCP server to provide TFTP server and boot file information to the VM. See Configuring the PXE Environment for AHV VMs on page 159.

For example, create a virtual interface for VM nw-boot-vm and place it on a network named network1.

```
nutanix@cvm$ acli vm.nic_create nw-boot-vm network=network1
```

4. Obtain the MAC address of the virtual interface.

```
nutanix@cvm$ acli vm.nic_list vm
```

Replace *vm* with the name of the VM.

For example, obtain the MAC address of VM nw-boot-vm.

```
nutanix@cvm$ acli vm.nic_list nw-boot-vm
00-00-5E-00-53-FF
```

5. Update the boot device setting so that the VM boots over the network.

```
nutanix@cvm$ acli vm.update_boot_device vm mac_addr=mac_addr
```

Replace *vm* with the name of the VM and *mac_addr* with the MAC address of the virtual interface that the VM must use to boot over the network.

For example, update the boot device setting of the VM named nw-boot-vm so that the VM uses the virtual interface with MAC address 00-00-5E-00-53-FF.

```
nutanix@cvm$ acli vm.update_boot_device nw-boot-vm mac_addr=00-00-5E-00-53-FF
```

6. Power on the VM.

```
nutanix@cvm$ acli vm.on vm_list [host="host"]
```

Replace *vm_list* with the name of the VM. Replace *host* with the name of the host on which you want to start the VM.

For example, start the VM named nw-boot-vm on a host named host-1.

```
nutanix@cvm$ acli vm.on nw-boot-vm host="host-1"
```

Uploading Files to DSF for Microsoft Windows Users

If you are a Microsoft Windows user, you can securely upload files to DSF by using the following procedure.

Procedure

1. Use WinSCP, with SFTP selected, to connect to Controller VM through port 2222 and start browsing the DSF datastore.

> Note: The root directory displays storage containers and you cannot change it. You can only upload files to one of the storage containers and not directly to the root directory. To create or delete storage containers, you can use the Prism user interface.

2. Authenticate by using Prism username and password or, for advanced users, use the public key that is managed through the Prism cluster lockdown user interface.

Enabling Load Balancing of vDisks in a Volume Group

AHV hosts support load balancing of vDisks in a volume group for guest VMs. Load balancing of vDisks in a volume group enables IO-intensive VMs to use the storage capabilities of multiple Controller VMs (CVMs).

About this task

If you enable load balancing on a volume group, the guest VM communicates directly with each CVM hosting a vDisk. Each vDisk is served by a single CVM. Therefore, to use the storage capabilities of multiple CVMs, create more than one vDisk for a file system and use the OS-level striped volumes to spread the workload. This configuration improves performance and prevents storage bottlenecks.

> Note:
>
> - vDisk load balancing is disabled by default for volume groups that are directly attached to VMs.
>
> However, vDisk load balancing is enabled by default for volume groups that are attached to VMs by using a data services IP address.
> - You can attach a maximum number of 10 load balanced volume groups per guest VM.
> - For Linux VMs, ensure that the SCSI device timeout is 60 seconds. For information about how to check and modify the SCSI device timeout, see the Red Hat documentation at https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/5/html/online_storage_reconfiguration_guide/task_controlling-scsi-command-timer-onlining-devices.

Perform the following procedure to enable load balancing of vDisks by using aCLI.

Procedure

1. Log on to a Controller VM with SSH.
2. Do one of the following:
   - » Enable vDisk load balancing if you are creating a volume group.

     ```
     nutanix@cvm$ acli vg.create vg_name load_balance_vm_attachments=true
     ```

     Replace *vg_name* with the name of the volume group.
   - » Enable vDisk load balancing if you are updating an existing volume group.

     ```
     nutanix@cvm$ acli vg.update vg_name load_balance_vm_attachments=true
     ```

     Replace *vg_name* with the name of the volume group.

     > Note: To modify an existing volume group, you must first detach all the VMs that are attached to that volume group before you enable vDisk load balancing.
3. Verify if vDisk load balancing is enabled.

   ```
   nutanix@cvm$ acli vg.get vg_name
   ```

   An output similar to the following is displayed:

   ```
   nutanix@cvm$ acli vg.get ERA_DB_VG_xxxxxxxx
   ERA_DB_VG_xxxxxxxx {
     attachment_list {
   ```

```
    vm_uuid: "xxxxx"
.
.
.
.
iscsi_target_name: "xxxxx"
load_balance_vm_attachments: True
logical_timestamp: 4
name: "ERA_DB_VG_xxxxxxxx"
shared: True
uuid: "xxxxxx"
}
```

If vDisk load balancing is enabled on a volume group, load_balance_vm_attachments: True is displayed in the output. The output does not display the load_balance_vm_attachments: parameter at all if vDisk load balancing is disabled.

4. (Optional) Disable vDisk load balancing.

```
nutanix@cvm$ acli vg.update vg_name load_balance_vm_attachments=false
```

Replace *vg_name* with the name of the volume group.

VM High Availability in Acropolis

If you have not modified High Availability configuration from previous version of AOS releases, best effort VM availability is by default implemented in Acropolis.

If you have not enabled High Availability, in case of host failure VMs are restarted from the failed host to any available space on the other hosts in the cluster. Once the failed host joins the cluster again, VMs are migrated back to the host. This type of VM high availability is implemented without reserving any resources. Admission control is not enforced and hence there may not be sufficient capacity available to start all the VMs.

> Note: Nutanix does not support VMs that are running with 100% remote storage for High Availability. The VMs must have at least one local disk that is present on the cluster.

VM high availability can be configured for host-based reservations using segment-based reservation method that is the default.

In segment-based reservation, the cluster is divided into segments to ensure that enough space is reserved for any host failure. Each segment corresponds to the largest VM that is guaranteed to be restarted in case the failure occurs. The other factor is the number of host failures that can be tolerated. Using these inputs, the scheduler implements admission control to always have enough resources reserved so that the VMs can be restarted upon failure of any host in the cluster.

The segment size ensures that the largest VM can be powered on in HA failover when cluster is fully loaded (if the cluster is fully used except the reserved segments). The number of segments that is reserved is such a way that for each host enough resources are reserved to ensure that any host failure in the cluster is tolerated. Multiple VMs may fit into a segment. If anything changes in the cluster, the reservation is computed again. The total resources reserved for segments can be more than the resources used by running VMs. This implementation guarantees successful failover even in the case of fragmentation of segments. The actual number of reserved resources depends on the current load of the cluster, but it is typically at 1 to 1.25 times the resource usage on the most loaded host.

If the host enters maintenance mode (in case of host upgrade), you might not be protected against further host failures. Maintenance mode uses reservations made for HA for migrating VMs from the host. Although you are not protected against host failure if you have reservation for HA, hypervisor upgrade occurs without any difficulty because from the perspective of a user it is the same as host failure except that the VMs are migrated (instead of restarted) and

hence no runtime state is lost. The HA status goes through the same states as it goes when the host failure had occurred.

Segment-based reservation is the default method that Acropolis uses to enable VM high availability.

> CAUTION: Do not change segment-based reservation method with the reserve host method.

Fault Detection for High Availability

Acropolis version 6.1 (with minimum supported AHV version 20201105.2229) onwards, the fault detection mechanism for High Availability checks for the heartbeat of the management services on the host. If any one of the services are down, then the host is marked as disconnected.

In addition to this, the fault detection mechanism also performs the following health checks for the host:

- Root file system corruption
- Read-only root file system

If any of the above-mentioned health checks are affirmative, then the host is marked as disconnected.

If the host remains in the disconnected status for 40 seconds, the VMs running on the affected host are automatically restarted (based on the resource availability) .

You can view the alerts raised for any of the above-mentioned checks in the **Activity** > **Alerts** view in Prism UI.

Enabling High Availability for the Cluster

In Acropolis managed clusters, you can enable high availability for the cluster to ensure that VMs can be migrated and restarted on another host in case of failure.

About this task

After you enable high availability for the cluster, if a host failure occurs the cluster goes through following changes.

- OK: This state implies that the cluster is protected against a host failure.
- Healing: Healing period is the time that Acropolis brings the cluster to the protected state. There are two phases to this state. The first phase occurs when the host fails. The VMs are restarted on the available host. After restarting all the VMs if there are enough resources to protect the VM, the HA status of the cluster comes back to OK state. If this does not occur, the cluster goes into critical state. The second phase occurs when the host comes back from the failure. Once the host comes back from failure, no VMs are present on the host and hence during this healing phase restore locality task occurs (VMs are migrated back). Apart from restoring the locality of the VMs, the restore locality task ensures that the cluster is back to the same state before the HA failure. Once it is finished, the HA status is back to OK state.
- Critical: If the host is down, the HA status of the cluster goes into Critical state. This happens because the cluster cannot tolerate any more host failures. You have to ensure that you bring back the host so that your cluster is protected against any further host failures.

> Note: On a less loaded cluster, it is possible for HA to go directly back to OK state if enough resources is reserved to protect another host failure. The start and migrate operations on the VMs are restricted in the Critical state because Acropolis continuously tries to ensure that the HA status is back to the OK state.

Procedure

1. Click the gear icon in the main menu and then select **Manage VM High Availability** in the **Settings** page.

   > Note: This option does not appear in clusters that do not support this feature.

   The Manage VM High Availability dialog box appears.

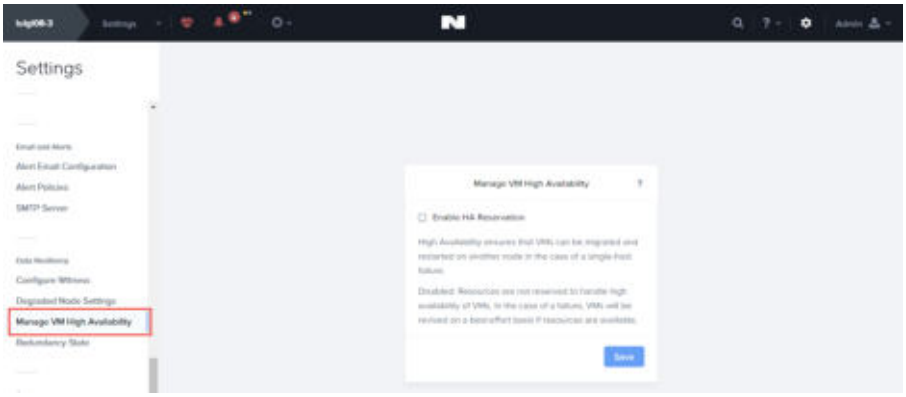


   Figure 60: Manage VM High Availability Window

2. Check the **Enable HA Reservation** box and then click the **Save** button to enable.

Live vDisk Migration Across Storage Containers

vDisk migration allows you to change the container of a vDisk. You can migrate vDisks across storage containers while they are attached to guest VMs without the need to shut down or delete VMs (live migration). You can either migrate all vDisks attached to a VM or migrate specific vDisks to another container.

In a Nutanix solution, you group vDisks into storage containers and attach vDisks to guest VMs. AOS applies storage policies such as replication factor, encryption, compression, deduplication, and erasure coding at the storage container level. If you apply a storage policy to a storage container, AOS enables that policy on all the vDisks of the container. If you want to change the policies of the vDisks (for example, from RF2 to RF3), create another container with a different policy and move the vDisk to that container. With live migration of vDisks across containers, you can migrate vDisk across containers even if those vDisks are attached to a live VM. Thus, live migration of vDisks across storage containers enables you to efficiently manage storage policies for guest VMs.

General Considerations

You cannot migrate images or volume groups.

You cannot perform the following operations during an ongoing vDisk migration:

- Clone the VM
- Resize the VM
- Take a snapshot

> Note: During vDisk migration, the logical usage of a vDisk is more than the total capacity of the vDisk. The issue occurs because the logical usage of the vDisk includes the space occupied in both the source and destination containers. Once the migration is complete, the logical usage of the vDisk returns to its normal value.

Migration of vDisks stalls if sufficient storage space is not available in the target storage container. Ensure that the target container has sufficient storage space before you begin migration.

Disaster Recovery Considerations

Consider the following points if you have a disaster recovery and backup setup:

- You cannot migrate vDisks of a VM that is protected by a protection domain or protection policy. When you start the migration, ensure that the VM is not protected by a protection domain or protection policy. If you want to migrate vDisks of such a VM, do the following:
  - Remove the VM from the protection domain or protection policy.
  - Migrate the vDisks to the target container.
  - Add the VM back to the protection domain or protection policy.
  - Configure the remote site with the details of the new container.

  vDisk migration fails if the VM is protected by a protection domain or protection policy.
- If you are using a third-party backup solution, AOS temporarily blocks snapshot operations for a VM if vDisk migration is in progress for that VM.

Migrating a vDisk to Another Container

You can either migrate all vDisks attached to a VM or migrate specific vDisks to another container.

About this task

Perform the following procedure to migrate vDisks across storage containers:

Procedure

1. Log on to a CVM in the cluster with SSH.

2. Do one of the following:

   » Migrate all vDisks of a VM to the target storage container.

   ```
   nutanix@cvm$ acli vm.update_container vm-name container=target-container wait=false
   ```

   Replace *vm-name* with the name of the VM whose vDisks you want to migrate and *target-container* with the name of the target container.

   » Migrate specific vDisks by using either the UUID of the vDisk or address of the vDisk.

   Migrate specific vDisks by using the UUID of the vDisk.

   ```
   nutanix@cvm$ acli vm.update_container vm-name device_uuid_list=disk-UUID container=target-container wait=false
   ```

   Replace *vm-name* with the name of VM, *disk-UUID* with the UUID of the disk, and *target-container* with the name of the target storage container.

   Run `nutanix@cvm$ acli vm.get <vm-name>` to determine the UUID of the vDisk.

   You can migrate multiple vDisks at a time by specifying a comma-separated list of vDisk UUIDs.

   Alternatively, you can migrate vDisks by using the address of the vDisk.

   ```
   nutanix@cvm$ acli vm.update_container vm-name disk_addr_list=disk-address
    container=target-container wait=false
   ```

   Replace *vm-name* with the name of VM, *disk-address* with the address of the disk, and *target-container* with the name of the target storage container.

   Run `nutanix@cvm$ acli vm.get <vm-name>` to determine the address of the vDisk.

   Following is the format of the vDisk address:

   ```
   bus.index
   ```

   Following is a section of the output of the `acli vm.get vm-name` command:

   ```
   disk_list {
        addr {
          bus: "scsi"
          index: 0
        }
   ```

   Combine the values of bus and index as shown in the following example:

   ```
   nutanix@cvm$ acli vm.update_container TestUVM_1 disk_addr_list=scsi.0 container=test-container-17475
   ```

   You can migrate multiple vDisks at a time by specifying a comma-separated list of vDisk addresses.

3. Check the status of the migration in the **Tasks** menu of the Prism Element web console.

4. (Optional) Cancel the migration if you no longer want to proceed with it.

```
nutanix@cvm$ ecli task.cancel task_list=task-ID
```

Replace *task-ID* with the ID of the migration task.

Determine the task ID as follows:

```
nutanix@cvm$ ecli task.list
```

In the **Type** column of the tasks list, look for **VmChangeDiskContainer**.

VmChangeDiskContainer indicates that it is a vDisk migration task. Note the ID of such a task.

> Note: Note the following points about canceling migration:
>
> - If you cancel an ongoing migration, AOS retains the vDisks that have not yet been migrated in the source container. AOS does not migrate vDisks that have already been migrated to the target container back to the source container.
> - If sufficient storage space is not available in the original storage container, migration of vDisks back to the original container stalls. To resolve the issue, ensure that the source container has sufficient storage space.

Memory Overcommit

Configured and enabled using CLI, the Memory Overcommit feature allows the host to automatically detect the memory under-utilization or over-utilization on a VM. The Memory Overcommit feature allows the host to automatically combine and use excess and unutilized memory of a VM. The host may use this excess memory to provision the memory requirements of another VM, or to create another VM. Thus the Memory Overcommit feature allows you to achieve virtualized optimization of host memory utilization.

Memory overcommit is confined to the VMs in the host and cannot be used for VMs on another host.

> Note: Memory overcommit enables the host to utilize excess memory for any existing or new VM within the host. It does not support utilization of excess memory for VMs across the hosts. For example, excess unutilized memory of VM1 on Host1 cannot be utilized on VMs on Host2. It can only be utilized for another VM on Host1 itself.

With the Memory Overcommit feature enabled on the host, the total of all the memory assigned to VMs by the host can be greater than the total physical memory installed on the host. See Deployment Example.

> Note: The Memory Overcommit feature is only available with AOS version 6.0.2 with minimum AHV version of 20201105.30007. For more information, see the *Acropolis Family Release Notes*.

Deployment Workflow

You can enable or disable the Memory Overcommit feature using CLI on a new or existing VMs.

> Note: Ensure the VMs are shutdown before enabling or disabling Memory Overcommit.

Memory overcommit is useful in test and development environments to verify the performance of the VM. Based on the memory usage of the VM, AHV calculates an appropriate memory size for each VM enabled with memory overcommit.

Essential Concepts

Over Provisioned VM

A VM with a memory allocation that is more than what is required to run the application on it is called an over provisioned VM. For example, a VM with an allocation of 10 GB memory to run an SQL server is only using 7 GB memory. In this case, the VM is over provisioned by 3 GB. This over provisioned memory is unused by the VM. The host can use this unutilized memory to add a new VM or allocate more memory to an existing VM, in spite of non availability of physical memory on the host.

Memory Overcommitted VM

A VM that has the memory overcommit feature enabled and shares its unutilized memory with the host.

You can enable Memory Overcommit on the following types of VM:

- A VM which has unutilized memory due to over provisioning.
- Any VM that is created to utilize the unutilized memory that is available with the host because of the Memory Overcommit feature enabled on the VMs in the host.

Host Swap

After installing or upgrading the current AOS version of the cluster to AOS 6.0.2 with AHV 20201105.30007 or later, Memory Overcommit is available for the cluster. Swap disks are created on the ADSF for every host in the cluster.These host swap disks are only used when the VMs enabled with Memory Overcommit do not have an internal swap disk attached and experience memory pressure.

Memory Overcommit Deployment Example

This is an example to illustrate the working of the memory overcommit feature.

For example, a 128 GB host has three VMs (64 GB + 32 GB + 32 GB). Before you enable memory overcommit, the physical memory of the host is equal to the total memory of all the VMs. Thus, before you enable memory overcommit, the host cannot add a new VM as the host's physical memory is fully committed.

Now, assume that the VMs are not fully utilizing their allocated memory and you enable the memory overcommit feature on the three VMs. Now the host allows you to create a new VM with a memory size that is equal to or less than 36 GB.

Table 10: Memory Overcommit Example

Host 1 (128 GB)			
VMs	VM Memory (GB)	Utilized Memory (GB)	Untilized Memory (GB)
VM1	64 GB	48 GB	16 GB
VM2	32 GB	20 GB	12 GB
VM3	32 GB	24 GB	8 GB
Total	128 GB	92 GB	36 GB

In this example, the host can combine the 36 GB of unutilized memory to create a new VM up to an aggregate size of 36 GB. Even though you can overcommit memory, the aggregate amount of *utilized* memory still cannot exceed the amount of physical memory in the host.

In the table below, VMs 1,2 and 3 contribute 16 GB, 12 GB, and 8 GB of unutilized memory respectively (thus 36 GB cumulatively) to the host. The host creates a new VM - VM4 with an allocated memory of 32 GB. Thus the host is left with 4 GB of unutilized memory. VM4 utilizes 28 GB of the allocated 32 GB, thus contributing 4GB unutilized memory. Thus, the host still has 8 GB of unutilized memory with one new VM (VM4) running.

Table 11: Memory Overcommit Example

Host 1 (128 GB)			
VMs	VM Memory (GB)	Utilized Memory (GB)	Unutilized Memory (GB)
VM1	64 GB	48 GB	4 GB
VM2	32 GB	20 GB	
VM3	32 GB	24 GB	
VM4	32 GB	28 GB	4 GB
Total	160 GB	120 GB	8 GB

Note: Before the memory overcommit feature is enabled, the physical memory of the host is equal to the total memory of all the VMs. After the memory overcommit feature is enabled on the host, the total of all the memory assigned to VMs by the host can *appear* to be greater than the total physical memory installed on the host.

The above example is subject to the considerations mentioned in Requirements and Limitations.

Requirements for Memory Overcommit

Memory Overcommit is only available with AOS 6.0.2 and later, and has the following requirements:

- Ensure that you have an AOS version with AHV 20201105.30007 or later. See *Release Notes*.
- Install the latest VirtIO driver package.

  The latest VirtIO drivers ensure minimal impact on the performance of the VM enabled with Memory Overcommit.

Limitations of Memory Overcommit

Memory overcommit has the following limitations:

- You can enable or disable Memory Overcommit only while the VM is powered off.
- You cannot change memory allocation on a VM if Memory Overcommit is enabled on it.

- Memory overcommit is not supported with the following:
  - VMs that use GPU passthrough and vNUMA.

    For example, you cannot update a VM to a vNUMA VM when it is enabled with memory overcommit. The system displays the following alert: `InvalidArgument: Cannot use memory overcommit feature for a vNUMA VM error`.
  - If memory overcommit is enabled and the VM is powered on then memory hotplug cannot be used.

    For example, you cannot update the memory of a VM that is enabled with memory overcommit when it is still running. The system displays the following alert: `InvalidVmState: Cannot complete request in state on`.
- Memory overcommit feature can slow down the performance and the predictable performance of the VM.

  For example, migrating a VM enabled with Memory Overcommit takes longer than migrating a VM not enabled with Memory Overcommit.
- There may be a temporary spike in the aggregate memory usage in the cluster during the migration of a VM enabled with Memory Overcommit from one node to another.

  For example, when you migrate a VM from Node A to Node B, the total memory used in the cluster during migration is greater than the memory usage before the migration.

  The memory usage of the cluster eventually drops back to pre-migration levels when the cluster reclaims the memory for other VM operations.
- Using Memory Overcommit heavily can cause a spike in the disk space utilization in the cluster. This spike is caused because the Host Swap uses some of the disk space in the cluster.

  If the VMs do not have a swap disk, then in case of memory pressure, AHV uses space from the swap disk created on ADSF to provide memory to the VM. This can lead to an increase in disk space consumption on the cluster.
- All DR operations except Cross Cluster Live Migration (CCLM) are supported.

  On the destination side, if a VM fails when you enable Memory Overcommit, the failed VM fails over (creating the VM on the remote site) as a fixed size VM. You can enable Memory Overcommit on this VM after the failover is complete.

Memory Overcommit Management

You can create a new VM with memory overcommit after installing or upgrading the current AOS version of the cluster to AOS 6.0.2 with AHV 20201105.30007 or later.

The new VM created after installing or upgrading to AOS 6.0.2 supports memory overcommit. You are provided with an option to enable memory overcommit when creating a new VM.

Enabling Memory Overcommit

You can configure memory overcommit on a new VM or on an existing VM after shutting it down and enabling memory overcommit on it.

About this task

Perform the following procedure to enable memory overcommit on an existing VM after shutting it down:

Procedure

1. Log on to the Controller VM using SSH.
2. At the CVM prompt, type `acli` to enter the acropolis CLI mode.
3. Shut down the VM to enforce an update on the VM.

   ```
   acli> vm.shutdown vm-name
   ```

   Replace *vm-name* with the name of the VM.
4. Update the VM to enable memory overcommit.

   ```
   acli> vm.update vm-name memory_overcommit=True
   ```

   Replace *vm-name* with the name of the VM.

   > Note: Use the wildcard * to list out VMs of similar naming scheme and enable memory overcommit on them.
   >
   > `acli> vm.update vm-name* memory_overcommit=True`
   >
   > Use *vm-name** to update multiple VMs that follow identical naming scheme.
   >
   > For example, Host 1 has six VMs (VM1, VM2, VM3, FileServer, SQL server, VM4), then `vm.update VM* memory_overcommit=True` CLI command will only update VM1, VM2, VM3, and VM4. It will not update the FileServer and SQL Server VMs on the host.

   > Note: Use your discretion while using the wildcard * without mentioning the VM naming scheme. Doing so will update all the VMs in the cluster, including the VMs that you may not want to configure with memory overcommit.
5. Verify memory overcommit configuration.

   The memory_overcommit parameter should be set to `True`.

   ```
   acli> vm.get vm-name
   ```

   Replace *vm-name* with the name of the VM.
6. Start the VM.

   ```
   acli> vm.on vm-name
   ```

   Replace *vm-name* with the name of the VM.

Disabling Memory Overcommit

You can disable memory overcommit on an existing VM after shutting it down and changing the memory overcommit configuration of the VM.

About this task

Perform the following procedure to disable memory overcommit on an existing VM:

Procedure

1. Log on to the Controller VM using SSH.
2. At the CVM prompt, type `acli` to enter the acropolis CLI mode.

3. Shut down the VM.

```
acli> vm.shutdown vm-name
```

Replace *vm-name* with the name of the VM.

4. Update the VM to disable memory overcommit.

```
acli> vm.update vm_name memory_overcommit=False
```

Replace *vm-name* with the name of the VM.

> Note: Use the wildcard * to list out VMs of similar naming scheme and disable memory overcommit on them.
>
> `acli> vm.update vm-name* memory_overcommit=False`
>
> Use *vm-name** to update multiple VMs that follow identical naming scheme.
>
> For example, Host 1 has six VMs (VM1, VM2, VM3, FileServer, SQL server, VM4), then `vm.update VM* memory_overcommit=False` CLI command will only update VM1, VM2, VM3, and VM4. It will not update the FileServer and SQL Server VMs on the host.

> Note: Use your discretion while using the wildcard * without mentioning the VM naming scheme. Doing so will update all the VMs in the cluster, including the VMs that you may not want to configure with memory overcommit.

5. Start the VM.

```
acli> vm.on vm-name
```

Replace *vm-name* with the name of the VM.

OVAs

An Open Virtual Appliance (OVA) file is a tar archive file created by converting a virtual machine (VM) into an Open Virtualization Format (OVF) package for easy distribution and deployment. OVA helps you to quickly create, move or deploy VMs on different hypervisors.

Prism Central helps you perform the following operations with OVAs:

- Export an AHV VM as an OVA file.
- Upload OVAs of VMs or virtual appliances (vApps). You can import (upload) an OVA file with the QCOW2 or VMDK disk formats from a URL or the local machine.
- Deploy an OVA file as a VM.
- Download an OVA file to your local machine.
- Rename an OVA file.
- Delete an OVA file.
- Track or monitor the tasks associated with OVA operations in Tasks.

The access to OVA operations is based on your role. See *Role Details View* in the *Prism Central Guide* to check if your role allows you to perform the OVA operations.

For information about:

- Restrictions applicable to OVA operations, see OVA Restrictions on page 174.
- The OVAs dashboard, see *OVAs View* in the *Prism Central Guide*.

- Exporting a VM as an OVA, see *Managing a VM (AHV and Self Service)* in the *Prism Central Guide*.
- Other OVA operations, see *OVA Management* in the *Prism Central Guide*.

OVA Restrictions

You can perform the OVA operations subject to the following restrictions:

- Export to or upload OVAs with one of the following disk formats:
  - QCOW2: Default disk format auto-selected in the Export as OVA dialog box.
  - VMDK: Deselect QCOW2 and select VMDK, if required, before you submit the VM export request when you export a VM.
  - When you export a VM or upload an OVA and the VM or OVA does not have any disks, the disk format is irrelevant.
- Upload an OVA to multiple clusters using a URL as the source for the OVA. You can upload an OVA only to a single cluster when you use the local OVA File source.
- Perform the OVA operations only with appropriate permissions. You can run the OVA operations that you have permissions for, based on your assigned user role.
- The OVA that results from exporting a VM on AHV is compatible with any AHV version 5.18 or later.

5

VM TEMPLATE MANAGEMENT

In Prism Central, you can create VM templates to manage the golden image of a VM. A VM template can be considered as a master copy of a virtual machine. It captures the virtual machine configuration and the contents of the VM including the guest operating system and the applications installed on the VM. You can use this template to deploy multiple VMs across clusters.

> Note: You can create or manage a VM template only as an admin user.

Limitations of VM Template Feature

The current implementation of VM template feature has the following limitations:

- <Version related limitations>
- Templates do not copy the following attributes from the source VMs:
  - Host affinity attributes
  - HA priority attributes
  - NGT installation

  The above-listed attributes must be reconfigured at the deployed VMs.

Creating a VM Template

About this task

To create a VM template, do the following:

Procedure

1. Select the VM in List tab under VM dashboard.

   > Note: Before selecting a VM for creating a template, ensure that the VM is powered off.

2. In the **Actions** list, select **Create VM Template**.

   Alternatively, on the details page of the VM, select **More** and click **Create VM Template**.

3. This displays the Create template from VM dialog box. In the **Create Template from VM** window, do the following in the indicated fields:

   a. In the **Name** field, enter a name for the template.

   b. In the **Description** field, enter a description for the template. Description is an optional field.

   c. In the **Guest Customization** field, select your options for guest operating system (OS) customization.

      You can choose to customize the OS of the VMs that will be deployed using this template. In the **Script Type** field, select Sysprep (Windows) for customizing the Windows OS, and Cloud-init (Linux) for customizing the Linux OS. For configuring each of these script types, you can choose to either upload a custom script or opt for a guided setup in the **Configuration Method** field. The guided setup lets you enter the authentication information such as username, password, locale, hostname, domain join, and license key.

      You can also specify if you want to allow the template users to override the guest customization settings of the template while deploying the VM. If you select a script type and then allow the users to override the settings, the users can change the settings only for the configuration method. For example, the users can change the authentication information at the time of deploying a VM from template, or they can change from a guided setup to a custom script.

      If you select no customization at the time of creating the template but allow the users to override the guest customization settings, it gives the maximum customization control to the users. They can customize the script type and the configuration method.

      Note: If you opt for a guest customization script, ensure that the script is in a valid format. The guest customization scripts are not validated, hence even though the VM deployment

might succeed, the guest customization script may not work as expected and it will be apparent only by observing the VM after it gets deployed.

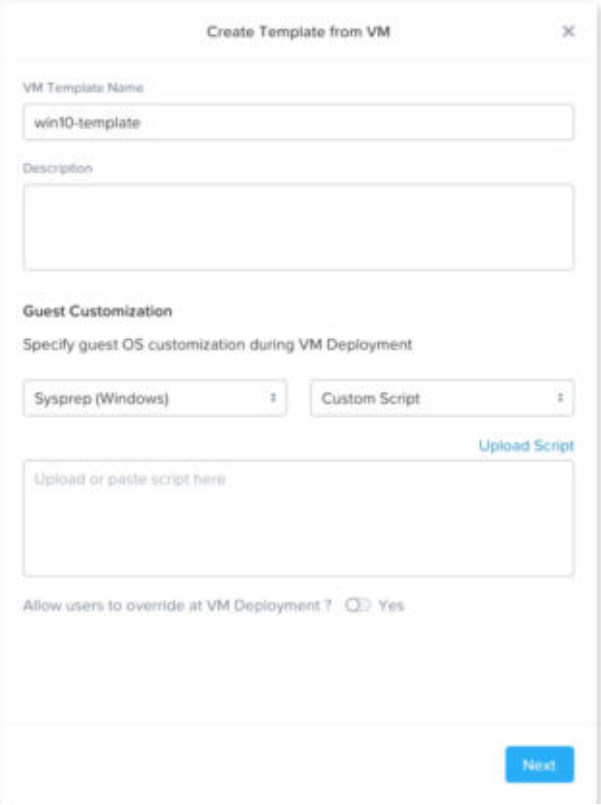


Figure 61: Create VM Template

4. Click **Next**.

   On the next page, you can review the configuration details, resource details, network details, and management details. Except the name, description and guest customization options that you chose on the previous page, you cannot modify any other setting while creating the template.

5. Click **Save** to save the inputs and create a template.

   The new template appears in the templates entity page list.

Deploying VM from a Template

About this task

After creating a VM template, you can use the template to deploy any number of VMs across clusters.

To deploy a VM from a template, do the following:

Procedure

1. You can use any of the following methods to deploy VMs:

   - Select the target template in the templates dashboard (see *VM Template Summary View* topic in the *Prism Central Guide* (see Prism)) and click the **Deploy VMs** button. By default, the active version of the template is used for deployment.
   - Go to the details page of a selected template (see *VM Template Details View* topic in the *Prism Central Guide* (see Prism)) and click the **Deploy VMs** button. By default, the active version of the template is used for deployment.
   - Go to the details page of a selected template (see VM Template Details View) and then go to **Versions** view. Select the desired version that you want to use for VM deployment, and click the **Deploy VMs** button. Here, you can choose any active or non-active version for the deployment.

2. The Deploy VM from Template page appears. By default, you see the **Quick Deploy** page. You can click**Advanced Deploy** button to access the Advanced Deploy page. In **Advanced Deploy** option, you can view and modify some VM properties and network settings.

   - **Quick Deploy**: To deploy VMs using the quick deploy method, provide inputs in the indicated fields:

     - **Name**: Enter a name for the VM.
     - **Cluster**: Select the cluster where you want to deploy the VM.
     - **Number of VMs**: Enter the number of VMs that you want to deploy.
     - **Starting Index Number**: Enter the starting index number for the VMs when you are deploying multiple VMs simultaneously. These index numbers are used in the VM names. For example, if you are deploying two VMs and specify the starting index number as 5, the VMs are named as `vm_name`-5 and `vm_name`-6.
     - **Guest Customization**: The template can have any one of the following options for guest OS customization: **No Customization**, **Sysprep (Windows)**, or **Cloud-init (Linux)**. For **Sysprep (Windows)**, or **Cloud-init (Linux)**, you can choose to either upload a custom script or opt for a guided setup. The fields are enabled for modification only if the template allows you to override its guest customization settings while deploying

the VM. If you are not allowed to override the guest customization settings, the settings that are already provided in the template are used for VM deployment.

- Click **Next** to verify the configuration details of the VMs to be deployed.
- Click **Deploy** to deploy the VMs.

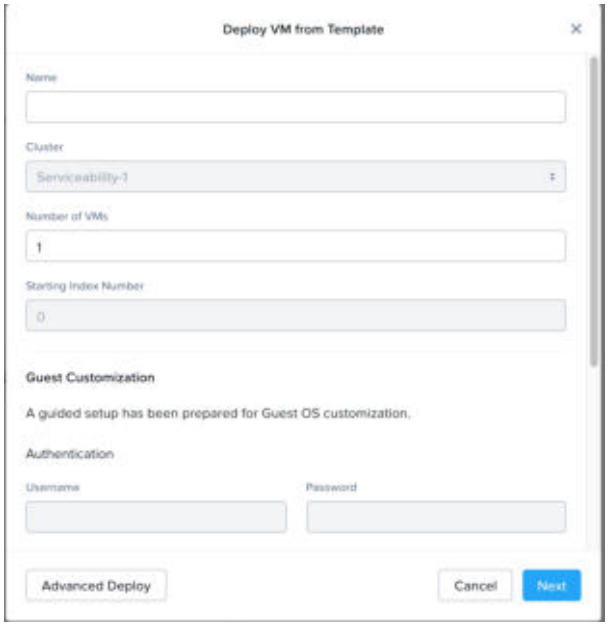


Figure 62: Deploy VM from Template using Quick Deploy Method

- **Advanced Deploy**: To deploy VMs using the advanced deploy method, provide inputs in the indicated fields and review the information displayed under various tabs:
  - **Configuration**: Provide inputs for name and description (optional) of the VM, cluster where you want to deploy the VM, Number of VMs to be deployed, and starting index

number (only if deploying multiple VMs). In this tab, you can also view and modify the VM properties such as CPU, core per CPU, and memory.

- **Resources**: Review the configuration settings for the VM resources such as disks, networks, and boot configuration. Here, you can modify the network settings but cannot modify any other settings.
- **Management**: If the template allows you to modify the settings for guest OS customization, provide inputs for the same.
- **Review**: Review the information displayed on this tab. Click **Deploy** to deploy the VMs.

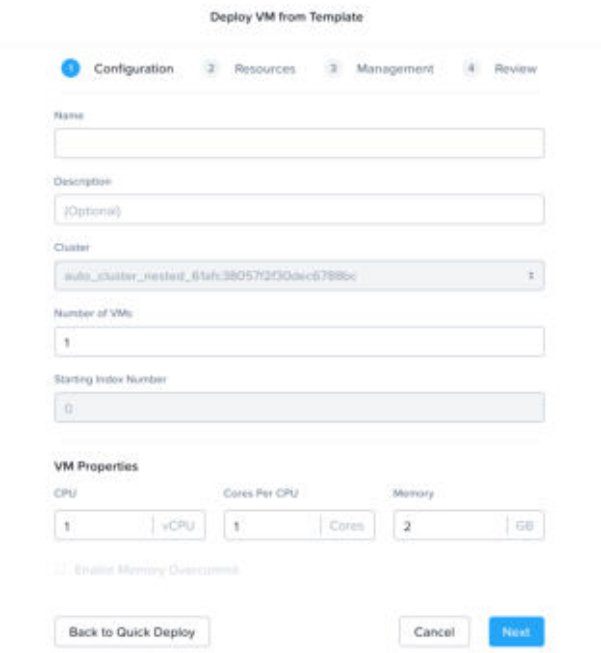


Figure 63: Deploy VM from Template using Advanced Deploy Method

Managing a VM Template

About this task

After creating a template (see Creating a VM Template on page 175), you can update the guest OS of the source VM of the template, complete guest OS update, cancel guest OS update, update the configuration of the template to create a version, or delete the template.

You can perform these tasks by using any of the following methods:

- Select the target template in the templates dashboard (see *VM Template Summary View* topic in the *Prism Central Guide* (see Prism)) and choose the required action from the **Actions** menu.
- Go to the details page of a selected template (see *VM Template Details View* topic in the *Prism Central Guide* (see Prism)) and select the desired action.

> Note: The available actions appear in bold; other actions are not available. The available actions depend on the current state of the template and your permissions.

Procedure

- To update the guest OS of the template, select **Update Guest OS**. Click **Proceed**.

  This step deploys a temporary VM and gives you access to that VM. You can also access the new VM from VM dashboard. You must start the VM, log on to the VM, and update the guest OS of that VM.

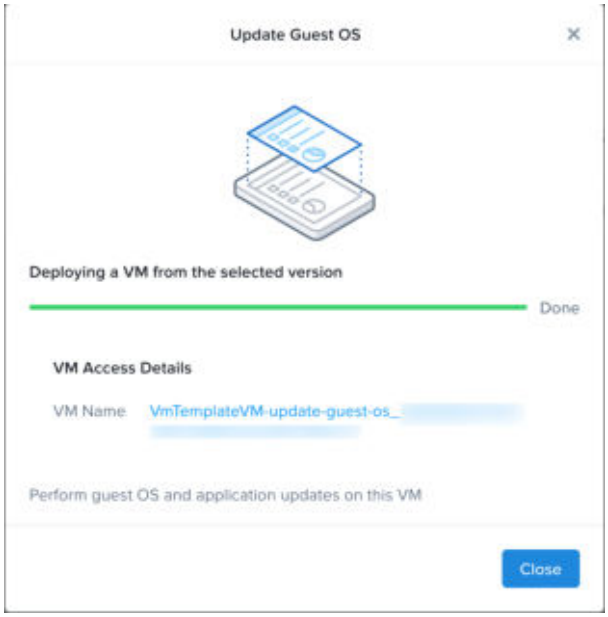


Figure 64: Update Guest OS

Once you have updated the guest OS of the temporary VM, you must complete the guest OS update from Prism Central UI. Alternatively, you can choose to cancel the guest OS update, if you choose to discontinue the guest OS update process.

The temporary VM automatically gets deleted after completion or cancellation of the guest OS upgrade process.

- To complete the process of guest OS update that you had earlier initiated, select **Complete Guest OS Update**. You must select this option only after successful update of the guest OS of the temporary VM.
- To cancel the process of guest OS update that you had earlier initiated, select **Cancel Guest OS Update**.
- To modify the template configuration, select **Update Configuration**. You cannot modify the configuration of the initial version of the template, but you can update the configuration

settings to create a new version of the template. In Select a version to Upgrade window, select the version of the template that you want to modify and create a version on top of it.

The Update Template Configuration window displays the following sections:

- **Configuration**: You can view the name of the base version that you want to update, change notes for that version, cluster name, VM properties (CPU, cores per CPU), and memory). In this section, you can modify only VM properties.
- **Resources**: You can view the information about disks, networks, and boot configuration. In this section, you can modify only network resources.
- **Management**: In this section, you can modify the guest customization settings.
- **Review**: In this section, you can review and modify the configuration settings that you are allowed to modify. You must provide a name and change notes for the new version. You can also choose to set this new version as active version. An active version is the version of the template that by default gets deployed when you click the **Deploy VMs** button.

Click **Save** to save the settings and create a new version of the template.

Figure 65: Update Template Configuration

- To delete a template, select **Delete Template**. A window prompt appears; click the **OK** button to delete the template.

COPYRIGHT

Copyright 2022 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

License

The provision of this software to you does not grant any licenses or other rights under any Microsoft patents with respect to anything other than the file server implementation portion of the binaries for this software, including no licenses or any other rights in any hardware or any devices or software that are used to communicate with or in connection with this software.

Conventions

Convention	Description
`variable_value`	The action depends on a value that is unique to your environment.
`ncli> command`	The commands are executed in the Nutanix nCLI.
`user@host$ command`	The commands are executed as a non-privileged user (such as nutanix) in the system shell.
`root@host# command`	The commands are executed as the root user in the vSphere or Acropolis host shell.
`> command`	The commands are executed in the Hyper-V host shell.
`output`	The information is displayed as output from a command or in a log file.

Version

Last modified: March 22, 2022 (2022-03-22T16:48:32+05:30)

Documentation

Software Documentation
Hardware Replacement Documentation
Hardware Compatibility Lists
Knowledge Base
Solutions Documentation
EOL Information
Field Advisories
Security
Configuration Maximums
Product Notices
Upgrade Paths
Compatibility and Interoperability Matrix
Support Resources
.NEXT Forums
Terms & Conditions
Nutanix Cloud Platform

Compatibility and Interoperability Matrix

Platform | Software Interoperability | AHV Guest OS | Partner Software

Hardware Manufacturer: Nutanix ×
Hardware Model: NX-1065-... × +2
AOS Version: 6.1 ×
Hypervisor: AHV ×
Hypervisor Version: Select
Reset Filters

Download.csv | Supported Products | Intermixing

1 - 3 of 3

Hardware Model	AOS Version	Hypervisor Version
Type to filter...		
NX-1065-G8	6.1	AHV-20201105.30142
NX-8035-G8	6.1	AHV-20201105.30142
NX-8155-G8	6.1	AHV-20201105.30142

1 - 3 of 3

Keys

End of Life

One click download enabled

Hypervisor Support Policy

1. Please see the Nutanix hypervisor support policy statement in our Support FAQ.
2. Please contact your hypervisor vendor for any questions regarding their support policies. Nutanix strongly recommends that customers deploy an 'in-support' hypervisor version on Nutanix systems.
3. Nutanix does not support hypervisor downgrades.
4. This matrix displays the compatible hypervisor version(s) available for each platform at the time of the selected platform release.
5. The message 'None supported' for selected model, AOS version, and hypervisor combinations indicates that the selected model, AOS version, and hypervisor combination are not compatible or QA qualified by Nutanix.
6. This matrix might not show patch release compatibility for all models ("No data found" or a similar message is displayed). However, Nutanix supports the installation of AOS patch releases associated with the qualified AOS major and maintenance releases listed in this matrix. For example, you can install the patch release AOS 5.10.0.1 on the qualified versions of AOS 5.10.
7. This matrix might not show patch release compatibility for all non-AHV hypervisors. However, Nutanix supports the installation of patch releases as stated in the Support FAQ.

Feedback

v5.4 | February 2022 | BP-2005

BEST PRACTICES

Data Protection and Disaster Recovery

NUTANIX
YOUR ENTERPRISE CLOUD

Copyright

Copyright 2022 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

Contents

1. Executive Summary 5

2. Introduction 6
- Audience 6
- Purpose 6

3. Nutanix Enterprise Cloud Overview 8
- Nutanix HCI Architecture 9

4. Web-Scale Data Protection 10

5. Deployment Overview 12
- Native Nutanix Snapshots 12
- Two-Way Mirroring 12
- Many-to-One 13
- To the Cloud 13
- Single-Node Backup 14
- Leap: Disaster Recovery Orchestration 14

6. Local Backup with Snapshots 16
- Native Snapshots 16
- Crash-Consistent vs. Application-Consistent Snapshots 16

7. Protection Domains 20
- Consistency Groups 20

8. Backup and Disaster Recovery on Remote Sites 22
- Remote Site Setup 22
- Scheduling Full Snapshots and Asynchronous Replication 27
- Scheduling LWS and NearSync 29
- Cross-Hypervisor Disaster Recovery 30
- Single-Node Backup Target 30

Cloud Connect.....31

9. Leap: Disaster Recovery Orchestration.....32
Required Infrastructure for Disaster Recovery Orchestration.....32
Availability Zones.....32
Protection Policies.....33
Recovery Plans.....35

10. Sizing Space.....38
Full Local Snapshots.....38
Asynchronous Replication.....39
Lightweight Snapshots (LWS).....40
NearSync.....42

11. Bandwidth.....43
Seeding.....43

12. Failover: Migrate vs. Activate.....45
Protection Domain Cleanup.....46

13. Self-Service File Restore.....47

14. Third-Party Backup Products.....48
Backups with Replication.....48

15. Conclusion.....49

Appendix.....50
Best Practices Checklist.....50
PowerShell Scripts.....56
About the Author.....61
About Nutanix.....61

List of Figures.....62

List of Tables.....63

1. Executive Summary

The Nutanix enterprise cloud software is a hyperconverged infrastructure system delivering storage, compute, and virtualization services for any application. Designed for supporting multiple virtualized environments, including Nutanix AHV, VMware ESXi, and Microsoft Hyper-V, Nutanix invisible infrastructure is exceptionally robust and provides many ways to achieve your required recovery point objectives (RPOs).

Enterprises are increasingly vulnerable to data loss and downtime during disasters as they rely on virtualized applications and infrastructure that their legacy data protection and disaster recovery solutions can no longer adequately support. This best practices guide discusses the optimal configuration for achieving data protection using the native Nutanix disaster recovery capabilities and the Leap disaster recovery orchestration features available both on-premises, in Xi, and in public cloud providers like Amazon Web Services (AWS). Whatever your use case, you can protect your applications with drag-and-drop functionality. The Nutanix Prism UI facilitates seamless management to configure the shortest recovery time objectives (RTOs) possible, so customers can build complex disaster recovery workflows at a moment's notice. With Leap built in, Prism Central allows you to apply protection policies across all your managed clusters. Once the business has decided on the required RPO, you can activate recovery plans to validate, test, migrate, and fail over in a seamless fashion. Recovery plans can protect availability zones both on-premises and hosted in Xi.

As application requirements change and grow, Nutanix can easily adapt to business needs. Nutanix is uniquely positioned to protect and operate in environments with minimal administrative effort because of its web-scale architecture and commitment to enterprise cloud operations.

© 2022 Nutanix, Inc. All rights reserved

2. Introduction

Audience

This best practices guide is part of the Nutanix Solutions Library. We wrote it for IT administrators and architects who want more information about the data protection and disaster recovery features built into the Nutanix enterprise cloud software. Readers should have basic familiarity with Nutanix and AOS.

Purpose

This document provides best practice guidance for data protection solutions implementation on Nutanix servers running AOS 5.19. We present the following concepts:

- Scalable metadata.
- Backup.
- Crash-consistent versus application-consistent snapshots.
- Protection domains.
- Protection policies.
- Recovery plans.
- Scheduling snapshots and asynchronous replication.
- Sizing disk space for local snapshots and replication.
- Scheduling lightweight snapshots (LWS) and NearSync replication.
- Sizing disk space for LWS and NearSync replication.
- Determining bandwidth requirements.
- File-level restore.

© 2022 Nutanix, Inc. All rights reserved

Table 1: Document Version History

Version Number	Published	Notes
1.0	December 2014	Original publication.
2.0	March 2016	Updated recommendations for current best practices throughout.
2.1	June 2016	Updated Backup and Disaster Recovery on Remote Sites section.
2.2	July 2016	Updated bandwidth sizing information.
2.3	December 2016	Updated for AOS 5.0.
2.4	May 2017	Updated information on sizing SSD space on a remote cluster.
3.0	December 2017	Updated for AOS 5.5.
3.1	September 2018	Updated overview and Remote Site Setup section.
4.0	December 2018	Updated for AOS 5.10 and Xi Leap.
4.1	February 2019	Updated Sizing Space section and Leap product details.
4.2	August 2019	Updated for AOS 5.11.
4.3	October 2019	Updated for AOS 5.11.1 and Nutanix Files support for NearSync replication.
5.0	May 2020	Updated for AOS 5.17.
5.1	September 2020	Updated for AOS 5.18.
5.2	December 2020	Updated for AOS 5.19 and Leap multisite replication.
5.3	March 2021	Refreshed content.
5.4	February 2022	Updated Third-Party Backup Products section.

© 2022 Nutanix, Inc. All rights reserved

3. Nutanix Enterprise Cloud Overview

Nutanix delivers a web-scale, hyperconverged infrastructure solution purpose-built for virtualization and both containerized and private cloud environments. This solution brings the scale, resilience, and economic benefits of web-scale architecture to the enterprise through the Nutanix enterprise cloud platform, which combines the core HCI product families—Nutanix AOS and Nutanix Prism management—along with other software products that automate, secure, and back up cost-optimized infrastructure.

Available attributes of the Nutanix enterprise cloud OS stack include:

- Optimized for storage and compute resources.
- Machine learning to plan for and adapt to changing conditions automatically.
- Intrinsic security features and functions for data protection and cyberthreat defense.
- Self-healing to tolerate and adjust to component failures.
- API-based automation and rich analytics.
- Simplified one-click upgrades and software life cycle management.
- Native file services for user and application data.
- Native backup and disaster recovery solutions.
- Powerful and feature-rich virtualization.
- Flexible virtual networking for visualization, automation, and security.
- Cloud automation and life cycle management.

Nutanix provides services and can be broken down into three main components: an HCI-based distributed storage fabric, management and operational intelligence from Prism, and AHV virtualization. Nutanix Prism furnishes one-click infrastructure management for virtual environments running on AOS. AOS is hypervisor agnostic, supporting two third-party hypervisors—VMware ESXi and Microsoft Hyper-V—in addition to the native Nutanix hypervisor, AHV.

© 2022 Nutanix, Inc. All rights reserved

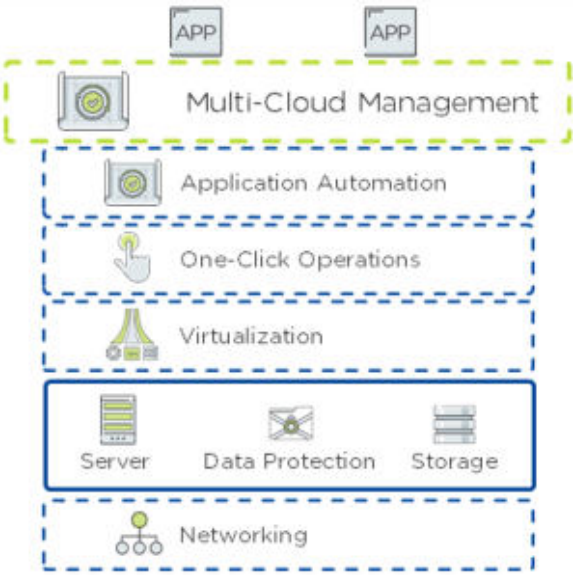


Figure 1: Nutanix Enterprise Cloud OS Stack

Nutanix HCI Architecture

Nutanix does not rely on traditional SAN or network-attached storage (NAS) or expensive storage network interconnects. It combines highly dense storage and server compute (CPU and RAM) into a single platform building block. Each building block delivers a unified, scale-out, shared-nothing architecture with no single points of failure.

The Nutanix solution requires no SAN constructs, such as LUNs, RAID groups, or expensive storage switches. All storage management is VM-centric, and I/O is optimized at the VM virtual disk level. The software solution runs on nodes from a variety of manufacturers that are either entirely solid-state storage with NVMe for optimal performance or a hybrid combination of SSD and HDD storage that provides a combination of performance and additional capacity. The storage fabric automatically tiers data across the cluster to different classes of storage devices using intelligent data placement algorithms. For best performance, algorithms make sure the most frequently used data is available in memory or in flash on the node local to the VM.

To learn more about Nutanix enterprise cloud software, visit the Nutanix Bible and Nutanix.com.

© 2022 Nutanix, Inc. All rights reserved

4. Web-Scale Data Protection

One of the key architectural differentiators for Nutanix is the ability to scale. Nutanix isn't bound by the same limitations as dual-controller architectures or federations relying on special hardware like NVRAM or custom ASICs for performance. When it comes to snapshots and disaster recovery, scaling metadata is a key part of delivering performance while ensuring availability and reliability. Each Nutanix node is responsible for a subset of the overall platform's metadata. All nodes in the cluster serve and manipulate metadata entirely through software, eliminating traditional bottlenecks.

Because each node has its own virtual storage controller and access to local metadata, replication scales with the system. Every node participates in replication to reduce hotspots throughout the cluster.

Nutanix uses two different forms of snapshots: full snapshots for asynchronous replication (when the RPO is 60 minutes or greater) and lightweight snapshots (LWS) for NearSync replication (when the RPO is between 15 minutes and 1 minute). Full snapshots keep system resource usage low when you use many snapshots over an extended period. The LWS feature reduces metadata management overhead and increases storage performance by decreasing the high number of storage I/O operations that long snapshot chains can cause.

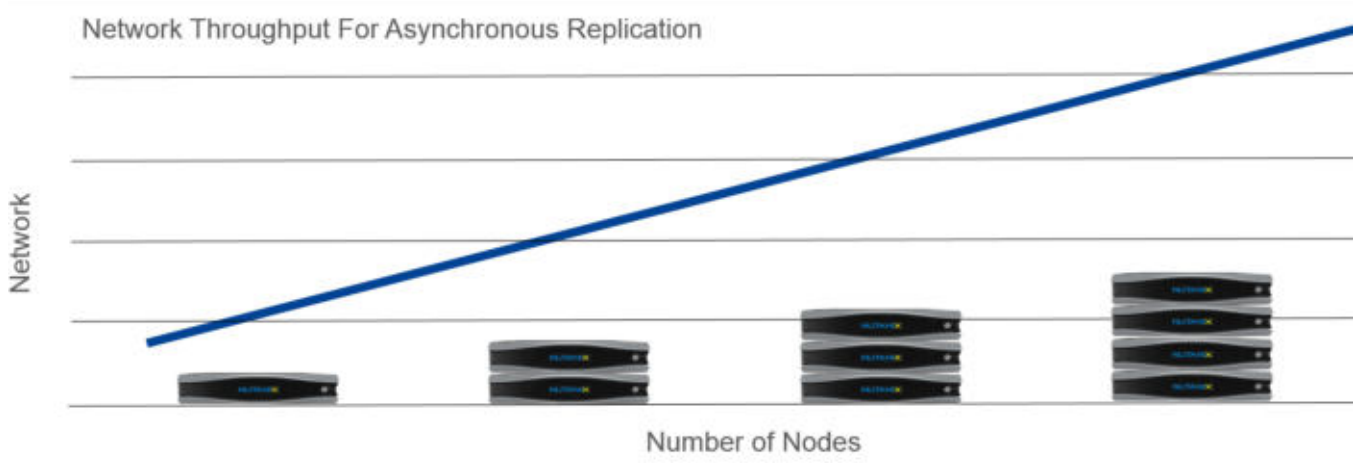


Figure 2: Scalable Replication

© 2022 Nutanix, Inc. All rights reserved

In asynchronous replication, every node can replicate four files, up to an aggregate of 100 MBps at one time. Thus, in a four-node configuration, the cluster can replicate 400 MBps or 3.2 Gbps. As you grow the cluster, the virtual storage controllers keep replication traffic distributed. In many-to-one deployments, as when remote branch offices communicate with a main datacenter, the main datacenter can use all its available resources to handle the increased replication load from the branch offices. When the main site is scalable and reliable, administrators don't have multiple replication targets to maintain, monitor, and manage. You can protect both VMs and volume groups with asynchronous replication.

NearSync offers unbound throughput. Because all writes go to SSD, we want to make sure that the performance tier doesn't fill up. To this end, AOS automatically allocates 7 percent of the performance tier to be used by NearSync. NearSync, which covers both VMs and volume groups, is supported for bidirectional replication between two clusters.

Nutanix also provides cross-hypervisor disaster recovery natively with asynchronous replication. Existing vSphere clusters can target AHV-based clusters as their disaster recovery and backup targets. Thanks to true VM mobility, Nutanix customers can place their workloads on the platform that best meets their needs.

© 2022 Nutanix, Inc. All rights reserved

5. Deployment Overview

Nutanix meets real-world requirements with native backup and replication infrastructure and management features that support a wide variety of enterprise topologies.

Native Nutanix Snapshots

Per-VM or per–volume group snapshots enable instant recovery. Depending on the workload and associated SLAs, customers can tune the snapshot schedule and retention periods to meet the appropriate RPOs. With the intuitive UI snapshot browser, you can perform restore and cloning operations instantly on the local cluster.

Two-Way Mirroring

The ability to mirror VM and volume group replication between multiple sites is necessary in environments where all sites must support active traffic. Consider a two-site example. Site B is the data protection target for selected workloads running on site A. At the same time, site A serves as the target for designated workloads running on site B. While asynchronous replication is supported for all workflows listed in this section, two-way mirroring is the only supported topology for NearSync.

Figure 3: Two-Way Mirroring

© 2022 Nutanix, Inc. All rights reserved

Many-to-One

In a many-to-one or hub-and-spoke architecture, you can replicate workloads running on sites A and B to a central site C. Centralizing replication to a single site may improve operational efficiency for geographically dispersed environments. Remote and branch offices (ROBO) are a classic many-to-one topology use case.

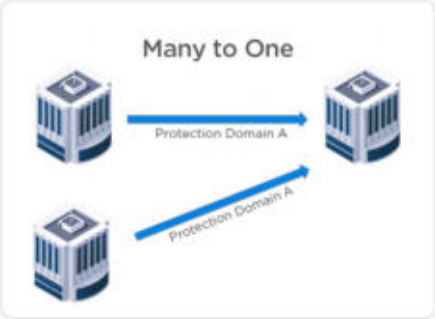


Figure 4: Many-to-One Architecture

To the Cloud

With Cloud Connect, customers can now use the public cloud as a destination for backing up their on-cluster VMs and volume groups. At this time, Nutanix supports Amazon Web Services (AWS) as the cloud destination. This option is particularly suitable for customers who don't have an offsite location for their backups or who are currently relying on tapes for storing their backups offsite. Cloud Connect provides customers with backup options for both Hyper-V and ESXi using the same Prism UI. Customers can also deploy a Nutanix cluster in a public cloud provider like AWS, setting up protection domains and employing the same workflows they use with on-prem clusters.

© 2022 Nutanix, Inc. All rights reserved

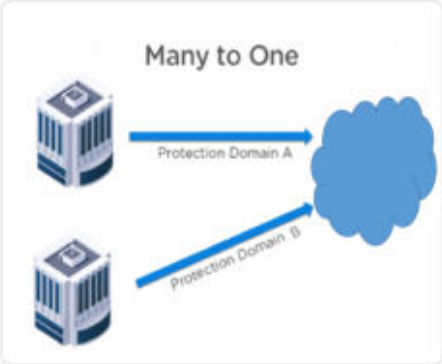


Figure 5: Public Cloud as a Backup Destination

Single-Node Backup

Nutanix has added support for single-node backup as a cost-efficient solution for providing full native backups to branch offices and SMBs. Using the same underlying disaster recovery technology, the single node can be either on-site or remote. Nutanix protects data on the node from single drive failure and provides native backup end to end. Single-node systems can also run VMs for testing or remote branch offices.

Leap: Disaster Recovery Orchestration

Prism Central provides a single web console for monitoring and managing multiple clusters. AOS versions 5.11 and later provide protection policies and recovery plans in Prism Central, offering an easy way to orchestrate operations around migrations and unplanned failures. Now you can apply orchestration policies for ESXi and AHV from a central location, ensuring consistency across all your sites and clusters.

To help manage these protection policies and recovery plans, Nutanix uses a construct called availability zones. On-premises, an availability zone includes all the Nutanix clusters managed by one Prism Central. An availability zone can also represent a region in Nutanix Xi Cloud Services. For disaster recovery, availability zones exist in pairs —either on-prem to on-prem or on-prem to Xi. Once you have paired your on-prem environment to a Xi-based availability zone, you can take advantage of Xi Leap, which is Nutanix disaster recovery as a service. There are multiple Xi Leap subscription plans available so you don’t have to pay the full cost associated with buying a secondary cluster up front and you save the time it takes to manage and operate the infrastructure.

© 2022 Nutanix, Inc. All rights reserved

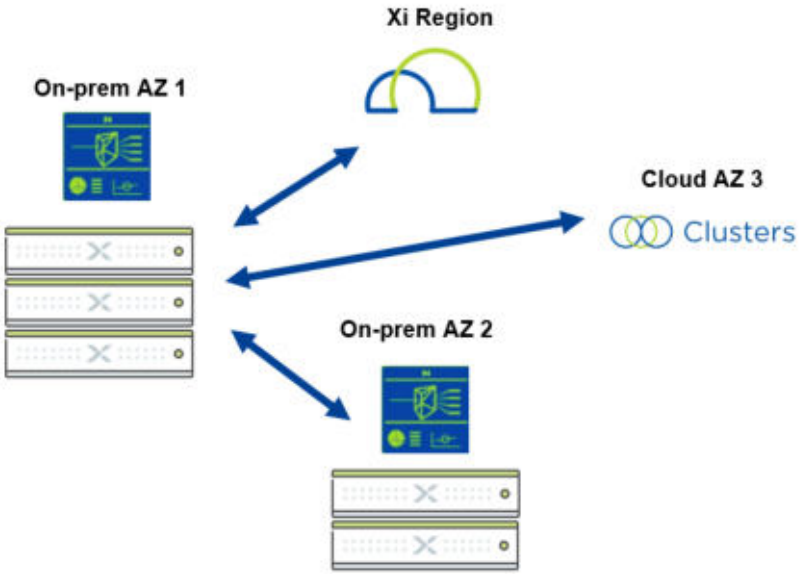


Figure 6: On-Prem, Cloud, and Xi-Based Availability Zones

6. Local Backup with Snapshots

Native Snapshots

Nutanix native snapshots provide production-level data protection without sacrificing performance. Nutanix uses a redirect-on-write algorithm to dramatically improve system efficiency for snapshots. Native snapshots operate at the VM level, and our crash-consistent snapshot implementation is the same across hypervisors. Implementation varies for application-consistent snapshots because of differences in the hypervisor layer. Nutanix can create local backups and recover data instantly to meet a wide range of data protection requirements.

Best practices:

- All VM files should sit on Nutanix storage. If you make non-Nutanix storage available to store files (VMDKs), the storage should have the same file path on both the source and destination clusters.
- Remove all external devices, including ISOs and floppy devices.

Crash-Consistent vs. Application-Consistent Snapshots

VM snapshots are by default crash-consistent, which means that the vDisks captured are consistent with a single point in time. The snapshot represents the on-disk data as if the VM crashed or the power cord was pulled from the server—it doesn't include anything that was in memory when the snapshot was taken. Today, most applications can recover well using crash-consistent snapshots.

Application-consistent snapshots capture the same data as crash-consistent snapshots, with the addition of all data in memory and all transactions in process. Because of their extra content, application-consistent snapshots are the most involved and take the longest to perform.

While most organizations find crash-consistent snapshots to be sufficient, Nutanix also supports application-consistent snapshots. The Nutanix application-consistent snapshot uses the Nutanix Volume Shadow Copy Service (VSS) to quiesce the file system for

© 2022 Nutanix, Inc. All rights reserved

ESXi and AHV prior to taking the snapshot. You can configure which type of snapshot each protection domain should maintain.

VSS Support with Nutanix Guest Tools

The Nutanix Guest Tools (NGT) software package for VMs plays an important role in application-consistent snapshots. ESXi and AHV-based snapshots call the Nutanix VSS provider from the Nutanix Guest Agent, which is one component of NGT. VMware Tools talk to the guest VM's VSS writers. Application-consistent snapshots quiesce all I/O, complete all open transactions, and flush the caches so everything is at the same point. VSS freezes write I/O while the native Nutanix snapshot takes place, so all data and metadata is written in a consistent manner. Once the Nutanix snapshot takes place, VSS thaws the system and allows queued writes to occur. Application-consistent snapshots don't snapshot the OS memory during this process.

Requirements for Nutanix VSS snapshots:

- Configure an external cluster IP address.
- Guest VMs should be able to reach the external cluster IP on port 2074.
- Guest VMs should have an empty IDE CD-ROM for attaching NGT.
- Only available for ESXi and AHV.
- Virtual disks must use the SCSI bus type.
- For NearSync support, you must have NGT version 1.3 installed.
- VSS isn't supported with NearSync or if the VM has any delta disks (hypervisor snapshots).
- Nutanix VSS snapshots are only available for these supported versions:
  - Windows 7
  - Windows Server 2008 R2 and later
  - CentOS 6.5 and 7.0
  - Red Had Enterprise Linux (RHEL) 6.5 and 7.0
  - Oracle Linux 6.5 and 7.0
  - SUSE Linux Enterprise Server (SLES) 11 SP4 and 12

© 2022 Nutanix, Inc. All rights reserved

- VSS must be running on the guest VM. Check the PowerShell Scripts section of the appendix for a script that verifies whether the service is running.
- The guest VM must support the use of VSS writers. Check the PowerShell Scripts section of the appendix for a script that ensures VSS writer stability (Windows only).
- VSS isn't supported for volume groups.
- You can't include volume groups in a protection domain configured for Metro Availability.
- You can't include volume groups in a protected VStore.
- You can't use Nutanix native snapshots to protect VMs that have VMware fault tolerance enabled.

For ESXi, if you haven't installed NGT, the process fails back using VMware Tools. Because the VMware Tools method creates and deletes an ESXi-based snapshot whenever it creates a native Nutanix snapshot, it generates more I/O stress. To eliminate this stress, we strongly recommend installing NGT.

Best practices:

- Schedule application-consistent snapshots during off-peak hours. NGT takes less time to quiesce applications than VMware Tools, but application-consistent snapshots still take longer than crash-consistent snapshots.
- Increase cluster heartbeat settings when using Windows Server Failover Cluster (WSFC).
- To avoid accidental cluster failover when you perform a vMotion, follow VMware best practices to increase heartbeat probes:
  - Change the tolerance of missed heartbeats from the default of 5 to 10.
  - Increase the number to 20 if your servers are on different subnets.
  - If you run Windows Server 2012, adjust the RouteHistoryLength to double the CrossSubnetThreshold value.

Table 2: MS Failover Settings Adjusted for Using VSS

(get-cluster).SameSubnetThreshold = 10

(get-cluster).CrossSubnetThreshold = 20
(get-cluster).RouteHistoryLength = 40

VSS on Hyper-V

Hyper-V on Nutanix supports VSS only through third-party backup applications, not snapshots. Because the Microsoft VSS framework requires a full share backup for every virtual disk in the share, Nutanix recommends limiting the number of VMs on any container using VSS backup.

Best practices:

- Create different containers for VMs that need VSS backup support. Don’t exceed 50 VMs on each container.
- Create a separate large container for crash-consistent VMs.

© 2022 Nutanix, Inc. All rights reserved

7. Protection Domains

A protection domain is a group of VMs or volume groups that you can either snapshot locally or replicate to one or more clusters when you have a remote site configured. Prism Element uses protection domains when replicating between remote sites.

Best practices:

- Protection domain names must be unique across sites.
- No more than 200 VMs per protection domain.
  - VMware Site Recovery Manager and Metro Availability protection domains are limited to 3,200 files.
  - No more than 10 VMs per protection domain with LWS.
- Group VMs with similar RPO requirements.
  - NearSync can only have one schedule, so place NearSync VMs in their own protection domain.

Consistency Groups

Administrators can create a consistency group for VMs and volume groups that are part of a protection domain and need to be snapshotted in a crash-consistent manner.

Best practices:

- Keep consistency groups as small as possible. Collect dependent applications or service VMs into a consistency group to ensure that they're recovered in a consistent state (for example, put a web server and database in the same consistency group).
- For all hypervisors, try to limit consistency groups to fewer than 10 VMs following the above best practices. Although we have tested consistency groups with up to 50 VMs, it's more efficient to have smaller consistency groups.
- Each consistency group using application-consistent snapshots can contain only one VM.

© 2022 Nutanix, Inc. All rights reserved

- When you provide disaster recovery for VDI using VMware View Composer or Machine Creation Services (MCS), place each protected VM in its own consistency group (including the gold image) inside a single protection domain.

© 2022 Nutanix, Inc. All rights reserved

8. Backup and Disaster Recovery on Remote Sites

Nutanix allows administrators to set up remote sites and select whether they use those remote sites for simple backup or for both backup and disaster recovery.

Remote sites are a logical construct. You must configure any AOS cluster—either physical or based in the cloud—as a remote site from the perspective of the source cluster before you use it as the destination for storing snapshots. Similarly, on this secondary cluster, you must configure the primary cluster as a remote site before snapshots from the secondary cluster start replicating to it.

Configuring the backup option on Nutanix allows an organization to use its remote site as a replication target. This option means you can back up data to this site and retrieve snapshots from it to restore locally, but failover protection (that is, running failover VMs directly from the remote site) isn't enabled. Backup supports using multiple hypervisors; as an example, an enterprise might have ESXi in the main datacenter but use Hyper-V at a remote location. With the backup option configured, the Hyper-V cluster could use storage on the ESXi cluster for backup. Using this method, Nutanix can also back up to AWS from Hyper-V or ESXi.

Configuring the disaster recovery option allows you to use the remote site both as a backup target and as a source for dynamic recovery. In this arrangement, failover VMs can run directly from the remote site. Nutanix provides cross-hypervisor disaster recovery between ESXi and AHV clusters. Currently, Hyper-V clusters can only provide disaster recovery to other Hyper-V-based clusters.

For data replication to succeed, configure forward (DNS A) and reverse (DNS PTR) DNS entries for each ESXi management host on the DNS servers used by the Nutanix cluster.

Remote Site Setup

You can customize several options when you set up a remote site. Protection domains inherit all remote site properties during replication.

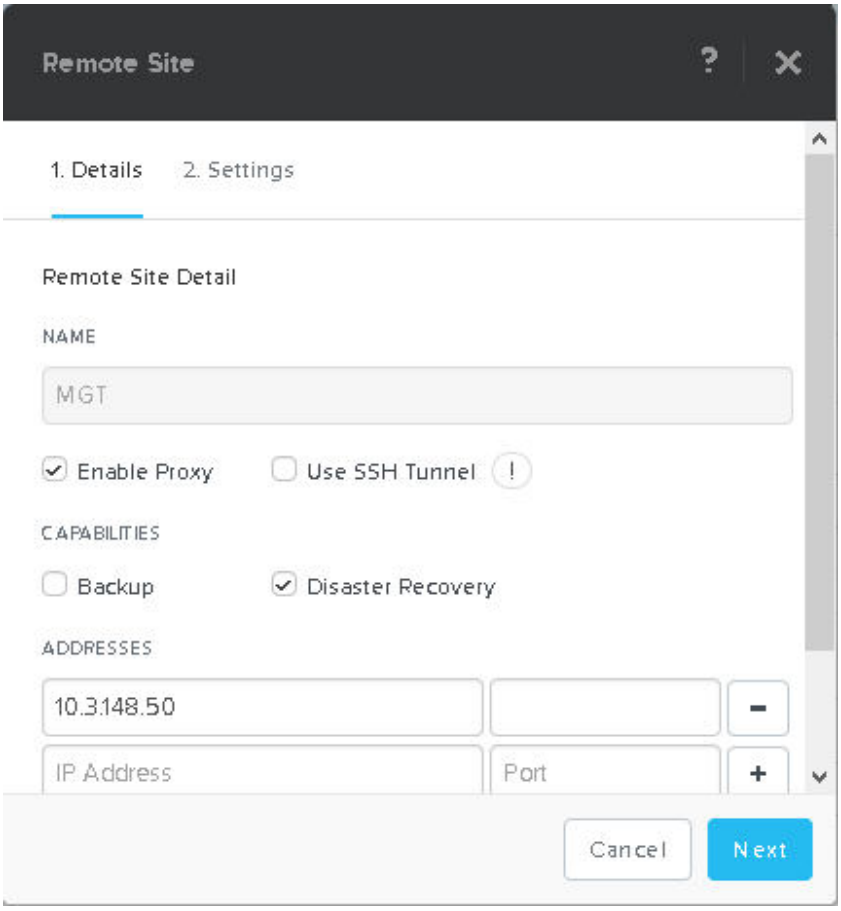


Figure 7: Setup Options for a Remote Site

Address

Use the external cluster IP as the address for the remote site. The external cluster IP is highly available, as it creates a virtual IP address for all the virtual storage controllers. You can configure the external cluster IP in the Prism UI under cluster details.

Other recommendations include:

- Try to keep both sites at the same AOS version. If both sites require compression, both must have the compression feature licensed and enabled.
- Open the following ports between both sides: 2009 TCP, 2020 TCP, 9440 TCP, and 53 UDP. If you use the SSH tunnel, also open 22. Use the external cluster IP address for source and destination. Cloud Connect uses a port between 3000–3099 but that

setup occurs automatically. You must allow all CVM IPs to pass replication traffic between sites with the ports detailed above. To simplify firewall rules, you can use the proxy described in the following section.

Enable Proxy

The enable proxy option redirects all egress remote replication traffic through one node. This remote site proxy is different from the Prism proxy. When you enable the proxy, replication traffic goes to the remote site proxy, which then forwards it to other nodes in the cluster. This arrangement significantly reduces the number of firewall rules you need to set up and maintain.

Best practice:

- Use the proxy in conjunction with the external address.

SSH Tunnel

An SSH tunnel is a point-to-point connection—one node in the primary cluster connects to a node in the remote cluster. By enabling proxy, we force replication traffic to go over this node pair. You can use the SSH tunnel between Cloud Connect and physical Nutanix clusters when you can't set up a virtual private network (VPN) between the two clusters. We recommend using an SSH tunnel as a fail-back option in lieu of a VPN.

Best practices:

- To use an SSH tunnel, enable the proxy.
- Open port 22 between external cluster IPs.
- Only use SSH tunnel for testing—not production. Use a VPN between remote sites or a Virtual Private Cloud (VPC) with AWS.

Capabilities

The disaster recovery option requires that both sites either support cross-hypervisor disaster recovery or have the same hypervisor. Today, Nutanix supports only ESXi and AHV for cross-hypervisor disaster recovery with full snapshots. When using the backup option, the sites can use different hypervisors, but you can't restore VMs on the remote side. You also use the backup option when you back up to AWS and Azure.

© 2022 Nutanix, Inc. All rights reserved

Bandwidth Throttling

Max bandwidth is set to throttle traffic between sites when no network device can limit replication traffic. The max bandwidth option allows for different settings throughout the day, so you can assign a max bandwidth policy when your sites are busy with production data and disable the policy when they aren't as busy. Max bandwidth doesn't imply a maximum observed throughput.

When you talk with your networking teams, note that this setting is in megabytes per second, not megabits per second. NearSync doesn't currently honor maximum bandwidth thresholds.

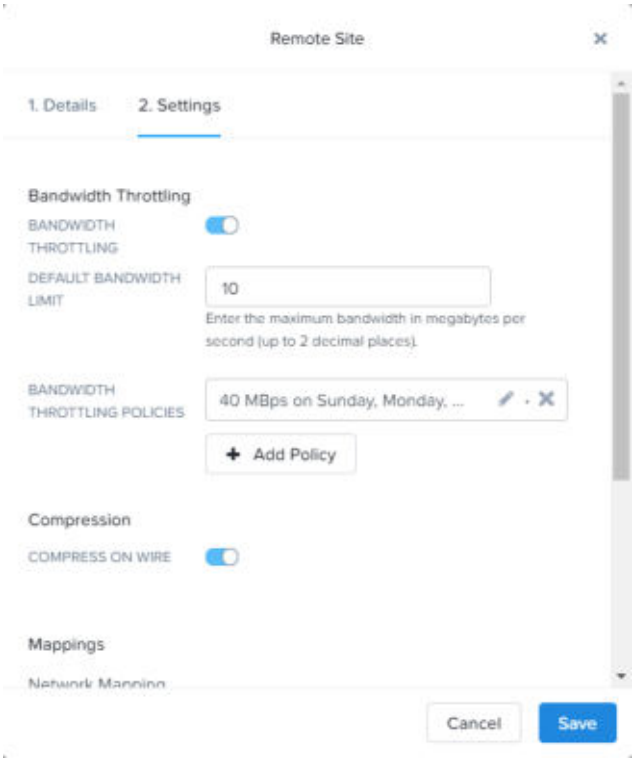


Figure 8: Max Bandwidth Settings

Remote Container

VStore name mapping identifies the container on the remote cluster used as the replication target. When you establish the VStore mapping, we recommend that you create a new, separate remote container with no VMs running on it on the remote side.

This configuration allows the hypervisor administrator to distinguish failed-over VMs quickly and apply polices on the remote side easily in case of a failover.

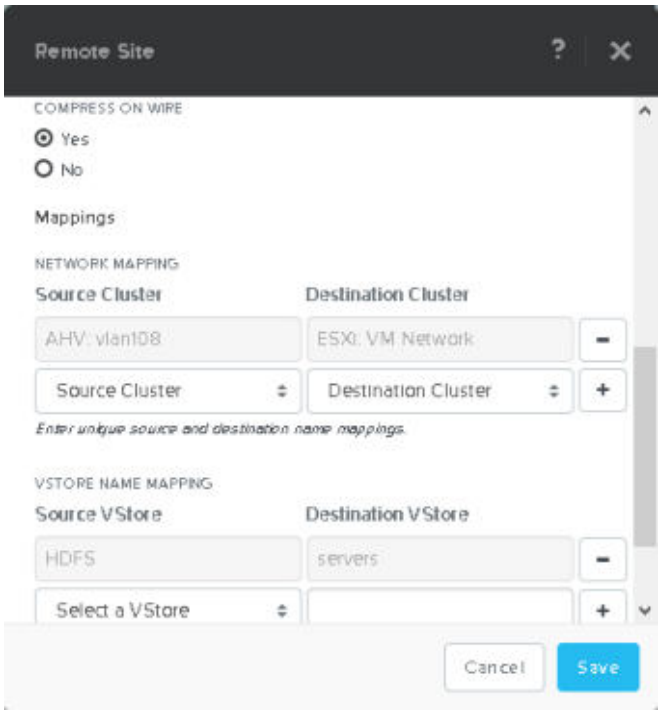


Figure 9: VStore Container Mappings for Replication

Best practices:

- Create a new remote container as the target for the VStore mapping.
- If you're backing up many clusters to one destination cluster, use only one destination container if the source containers have similar advanced settings.
- Enable MapReduce compression if licensing permits.
- If you use vCenter Server to manage both the primary and remote sites, don't have storage containers with the same name on both sites.

If the aggregate incoming bandwidth required to maintain the current change rate is less than 500 Mbps, we recommend skipping the performance tier. This setting saves your flash for other workloads while also saving on SSD write endurance. To skip the performance tier, use the following command from the nCLI:

```
ncli ctr edit sequential-io-priority-order=DAS-SATA,SSD-SATA,SSD-PCIe
 name=<container-name>
```

You can reverse this command at any time.

Network Mapping

AOS supports network mapping for disaster recovery migrations moving to and from AHV. Whenever you delete or change the network attached to a VM specified in the network map, modify the network map accordingly.

Scheduling Full Snapshots and Asynchronous Replication

The snapshot schedule should be equal to your desired RPO. In practical terms, the RPO determines how much data you can afford to lose in the event of a failure. The failure could be due to a hardware, human, or environmental issue. Taking a snapshot every 60 minutes for a server that changes infrequently or when you don't need a low RPO uses resources that could benefit more critical services.

The RPO is set from the local site. If you set a schedule to take a snapshot every hour, bandwidth and available space at the remote site determine if you can achieve the RPO. In constrained environments, limited bandwidth may cause the replication to take longer than the one-hour RPO, increasing the RPO. We list guidelines for sizing bandwidth and capacity to avoid this scenario later in this document.

© 2022 Nutanix, Inc. All rights reserved

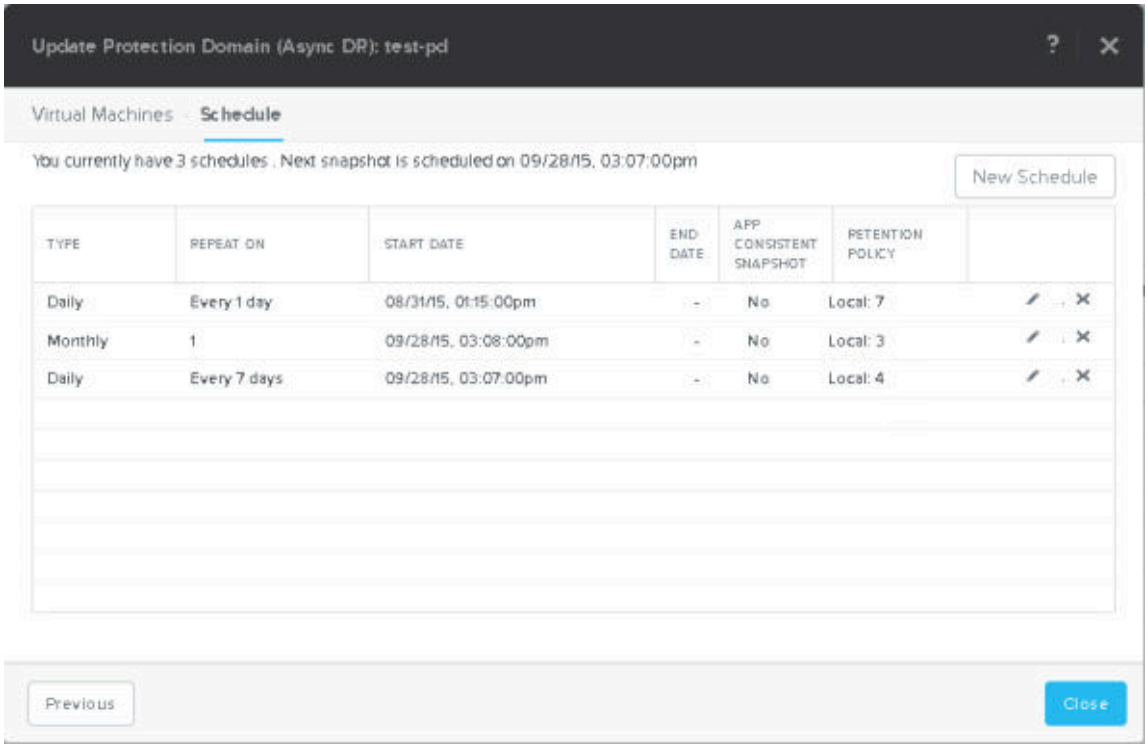


Figure 10: Multiple Schedules for a Production Domain

You can create multiple schedules for a protection domain (PD) using full snapshots, and you can have multiple PDs. The previous figure shows seven daily snapshots, four weekly snapshots, and three monthly snapshots to cover a three-month retention policy. This policy manages metadata on the cluster more efficiently than a daily snapshot with a 180-day retention policy.

Best practices:

- Stagger replication schedules across PDs. If you have a PD starting at the top of the hour, stagger the PDs by half of the most common RPO. The goal is to spread out replication impact on performance and bandwidth.
- Configure snapshot schedules to retain the lowest number of snapshots while still meeting the retention policy, as shown in the previous figure.

Remote snapshots implicitly expire based on how many snapshots there are and how frequently they are taken. For example, if you take daily snapshots and keep a maximum

of five, on the sixth day the first snapshot expires. At that point, you can't recover from the first snapshot because the system deletes it automatically.

In case of a prolonged network outage, Nutanix always retains the last snapshot to ensure that you don't ever lose all the snapshots. You can modify the retention schedule from the nCLI by changing the min-snap-retention-count. This value ensures that you retain at least the specified number of snapshots, even if all the snapshots have reached the expiry time. This setting works at the PD level.

Scheduling LWS and NearSync

Nutanix offers NearSync replication with a telescopic schedule (time-based retention). When you set the RPO to at most 15 minutes and at least 1 minute, you can save your snapshots for X number of weeks or months. Once you select NearSync, you can't add any more schedules.

The following table presents the default telescopic schedule to save recovery points for one month.

Table 3: Default Telescopic Schedule for One Month

Type	Frequency	Retention
Minute increments	Every minute	15 minutes
Hourly	Every hour	6 hours
Daily	Every 24 hours	7 days
Weekly	Every week	4 weeks
Monthly	Every month	1 month

As NearSync continues to improve, please refer to the latest set of requirements and limitations in the Prism Web Console Guide.

Limit the number of VMs to 10 or fewer per protection domain. If you can, maintain one VM per protection domain to help you transition back to NearSync if you run out of LWS reserve storage.

© 2022 Nutanix, Inc. All rights reserved

Cross-Hypervisor Disaster Recovery

Nutanix provides cross-hypervisor disaster recovery for migrating between ESXi and AHV clusters. The migration works with one click and uses the Prism data protection workflow. Once you have installed the mobility drivers through NGT, VMs can move freely between the hypervisors.

Best practices:

- Configure the CVM external IP address.
- Obtain the mobility driver from NGT.
- Don’t migrate VMs:
  - With delta disks (hypervisor-based snapshots).
  - Using SATA disks.
- Ensure that protected VMs have an empty IDE CD-ROM attached.
- Ensure that network mapping is complete.

Single-Node Backup Target

Nutanix offers the ability to use an NX-1155 or NX-1175 appliance as a single-node backup target for an existing Nutanix cluster. Because this target has different resources than the original cluster, its primary use case is to provide backup for a small set of VMs. This utility gives SMB and ROBO customers a fully integrated backup option.

Best practices:

- Combined, all protection domains should have fewer than 30 VMs.
- Limit backup retention to a three-month policy.
  - We recommend a policy that includes seven daily, four weekly, and three monthly backups.
- Only map an NX-1155 or NX-1175 to one physical cluster.
- Snapshot schedule should be at least six hours.
- Turn off deduplication.

© 2022 Nutanix, Inc. All rights reserved

Cloud Connect

The CVM running in AWS and Azure has limited SSD space, so we recommend following these best practices when you size:

- Try to limit each protection domain to one VM to speed up restores. This approach also saves money, as it limits the amount of data going across the WAN.
- The RPO shouldn't be lower than four hours.
- Turn off deduplication.
- Try to use Cloud Connect to protect workloads that have an average change rate of less than 0.5 percent.

© 2022 Nutanix, Inc. All rights reserved

9. Leap: Disaster Recovery Orchestration

The following best practices for disaster recovery orchestration cover both on-premises environments and Xi Leap. We have noted any differences between the two.

Required Infrastructure for Disaster Recovery Orchestration

- Deploy Prism Central.
  - For on-prem environments, deploy two Prism Central instances or one Prism Central instance at a separate site for high availability.
  - Leap for ROBO sites only requires one Prism Central, at the main site. When you replicate between two clusters managed from the same Prism Central instance, the minimum RPO is one hour or more and cross-hypervisor disaster recovery isn't supported.
- Deploy Prism Central on a subnet that doesn't fail over.
- Place the CVM and hypervisor IPs on a subnet different from the subnets used by VMs.
- The test network for on-prem disaster recovery orchestration requires a nonroutable VLAN.

Availability Zones

Paired availability zones synchronize the following disaster recovery configuration entities:

- Protection policies.
- Recovery plans.
- Categories used in protection polices and recovery plans.

Issues such as network connectivity loss between paired availability zones or user actions such as unpairing availability zones followed by pairing those availability zones

© 2022 Nutanix, Inc. All rights reserved

again can affect entity synchronization. Pairing previously unpaired availability zones triggers an automatic synchronization event. If you don't update entities before you resolve a connectivity issue or pair the availability zones again, the synchronization behavior resumes. If you update entities in either or both availability zones before you resolve such issues or pair unpaired availability zones again, you can't synchronize the entities. In such a scenario, you can force the entities in one availability zone to synchronize with the paired availability zone. This forced synchronization overwrites entities at the paired availability zone.

Observe the following recommendations to avoid inconsistencies and the resulting synchronization issues:

- During network connectivity issues, don't update an entity at both availability zones in a pair. You can safely make updates at any one location. After the connectivity issue is resolved, force synchronization from the availability zone you made the updates in. Failure to adhere to this recommendation results in synchronization failures.
- You can safely create new entities in either or both availability zones as long as you don't assign the same name to entities in both availability zones. After you resolve the connectivity issue, force synchronization from the availability zone where you created the entities.
- If one of the availability zones becomes unavailable or if a service in the paired availability zone is down, perform a forced sync from the paired availability zone after you resolve the issue.

Protection Policies

A protection policy automates the creation and replication of snapshots. When you configure a protection policy for creating local snapshots, simply specify the RPO, retention policy, and the entities you want to protect. If you want to automate snapshot replication to a remote location, you can also specify the remote location. Leap uses entity-centric disaster recovery. There are a few requirements for using protection policies:

- A VM can only belong to a PD or a protection policy, not both.
- If you don't use Nutanix AHV IPAM and need to retain your IP addresses, you must install NGT on the VMs you want protected.

© 2022 Nutanix, Inc. All rights reserved

- You can associate a VM with a maximum of two protection policies.
- When you protect a VM with more than one protection policy, only one of those policies can use NearSync or have Xi Leap as the target.

Best practices for using protection policies:

- Apply protection polices using categories.
- For VMs that haven't been replicated before, create a container with the same name on both sides. If the VM was in a PD before, it continues to use the same container.
- All protection policies that use NearSync must have fewer than 200 VMs.

Cross-Hypervisor Disaster Recovery

Beginning with AOS 5.11, cross-hypervisor disaster recovery is supported for Xi Leap for ESXi on-premises. (Because Xi Leap already runs AHV, AHV on-premises just works.) Cross-hypervisor disaster recovery preserves these elements of the ESXi configuration on failback when your on-prem environment is once again healthy:

- Port group (based on network mapping).
- Virtual hardware version.
- Network adapter types.
- Disk adapter type.
- VMware Tools state.
- Number of vCPUs.
- Number of cores per vCPU.
- Static MAC address.
- Disk provisioning (thick or thin).
- OS type.

Cross-hypervisor disaster recovery has the following limitations:

- Doesn't preserve SCSI controllers. Nutanix uses LSI logic on failback.
- No support for UEFI boot.
- Doesn't preserve HA and Distributed Resource Scheduler (DRS) settings.

© 2022 Nutanix, Inc. All rights reserved

- No support for hypervisor-based snapshots, VMware VSS, or linked clones.

Recovery Plans

A recovery plan orchestrates restoring protected VMs at a backup location. Recovery plans can either recover all specified VMs at once or, using what is essentially a runbook functionality, use power-on sequences with optionally configurable interstage delays to recover applications gracefully and in the required order. Recovery plans that restore applications in Xi Leap can also create the required networks during failover and can assign public-facing IP addresses to VMs.

Requirements for recovery plans:

- To enable disaster recovery orchestration, you must have set up a Prism Element external data services IP.
- Prism Central must run on a Nutanix cluster with the external data services IP.

Xi Leap doesn't allow you to create a recovery plan if:

- The recovery network that you specify exists with the same name on multiple clusters.
- The networks have the same name but different IP address spaces.

> **Note:** If you add a VM to multiple recovery plans and perform failover simultaneously on those recovery plans, each recovery plan creates an instance of the VM at the recovery location. You must manually clean up the additional instances.

Best practices:

- For on-premises availability zones, create a nonroutable network for testing failovers.
- Run the Validate workflow after you make changes to recovery plans.
- After you run the Test workflow, run the Clean-Up workflow instead of manually deleting VMs.
- A recovery plan should cover a maximum of 200 VMs.
- Maximum of 50 categories in a recovery plan.
- Maximum of 20 stages in a recovery plan.
- Maximum of 15 categories per stage in a recovery plan.
- Only run a maximum of 5 recovery plans in parallel.

© 2022 Nutanix, Inc. All rights reserved

- Keep asynchronous or NearSync replication VMs in separate recovery plans from synchronous replication VMs because synchronous recovery plans don't currently support planned failover.

Network Mapping

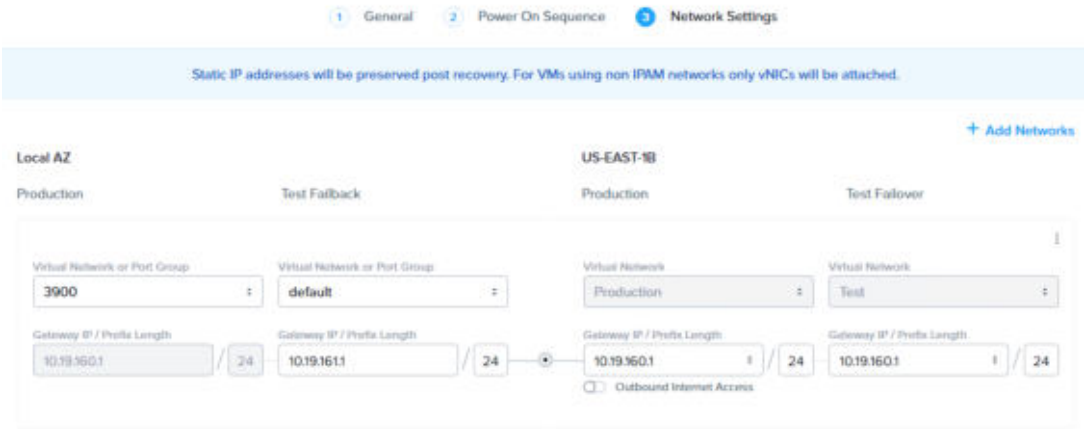


Figure 11: Network Mapping for Xi Leap

Virtual Networks in On-Premises Clusters

Virtual networks in on-premises Nutanix clusters are virtual subnets bound to a single VLAN. At physical locations, including the recovery location, administrators must create these virtual subnets manually, with separate virtual subnets created for production and test purposes. You must create these virtual subnets before you configure recovery plans.

When you configure a recovery plan, map the virtual subnets at the source location to the virtual subnets at the recovery location.

Virtual Networks in Xi Cloud

Xi Cloud features two built-in virtual networks: Production and Test. These virtual networks aren't analogous to the virtual subnets used in on-premises Nutanix clusters; they only provide two separate IP address spaces for production and testing so that activities performed in one do not affect the other. These separate Production and Test IP address spaces contain virtual subnets that are analogous to the virtual subnets in on-premises Nutanix clusters.

The Production virtual network contains subnets used for production workloads. These production workloads can be either workloads created and maintained in Xi Cloud or workloads that have failed over from a paired physical location.

The Test virtual network contains the subnets you want to recover VMs to when you test failover from the virtual subnets at a paired physical location. When you configure a recovery plan, map the virtual subnets on your on-premises clusters to virtual subnets in the Production and Test networks in Xi Cloud.

Best practices:

- Set up administrative distances on VLANs for subnets that completely fail over. If you don't set up administrative distances, shut down the VLAN on the source side after failover if the VPN connection is maintained between the two sites. If you are failing over to a new subnet, set up the subnet beforehand so you can test the routing.
- The prefix length for network mappings at the source and the destination must be the same.
- If you don't use Nutanix IPAM, you must install NGT to maintain a static address.
- To maintain a static address for Linux VMs that don't use Nutanix IPAM, the VMs must have the NetworkManager command-line tool (nmcli) version 0.9.10.0 or later installed. Additionally, you must use NetworkManager to manage the network for the Linux VMs. To enable NetworkManager on a Linux VM, set the value of the **NM_CONTROLLED** field to yes in the interface configuration file (for example, in CentOS, the file is /etc/sysconfig/network-scripts/ifcfg-eth0). After you set this field, restart the network service on the VM.

For networking best practices specific to Xi Leap, refer to the Xi Connectivity Tech Note.

Xi Leap Hypervisor Support

- Xi Leap supports clusters running AHV and ESXi as the source.

Xi Leap VM Configuration Limitations

VMs with the following configurations can't start:

- VMs configured with a GPU resource.
- VMs configured with four vNUMA sockets.

© 2022 Nutanix, Inc. All rights reserved

10. Sizing Space

Full Local Snapshots

To size space for local snapshots, you need to account for the rate of change in your environment and how long you plan to keep your snapshots on the cluster. Reduced snapshot frequency may increase the rate of change because there's a greater chance of common blocks changing before the next snapshot.

To find the space needed to meet your RPO, you can use the following formula. As you lower the RPO for asynchronous replication, you may need to account for an increased rate of transformed garbage. Transformed garbage is space that was allocated for I/O optimization or space that was assigned but to which the metadata no longer refers.

Table 4: Local Full Snapshot Reserve Formula

snapshot reserve = (frequency of snapshots × change rate per frequency) + (change rate per frequency × # of snapshots in a full curator scan × 0.1)

Note: A full curator scan runs every six hours.

You can look at your backups and compare the incremental differences between them to find the change rate. You could also take a conservative approach and start with a low snapshot frequency and a short expiry policy, then gauge the size difference between backups before consuming too much space.

© 2022 Nutanix, Inc. All rights reserved

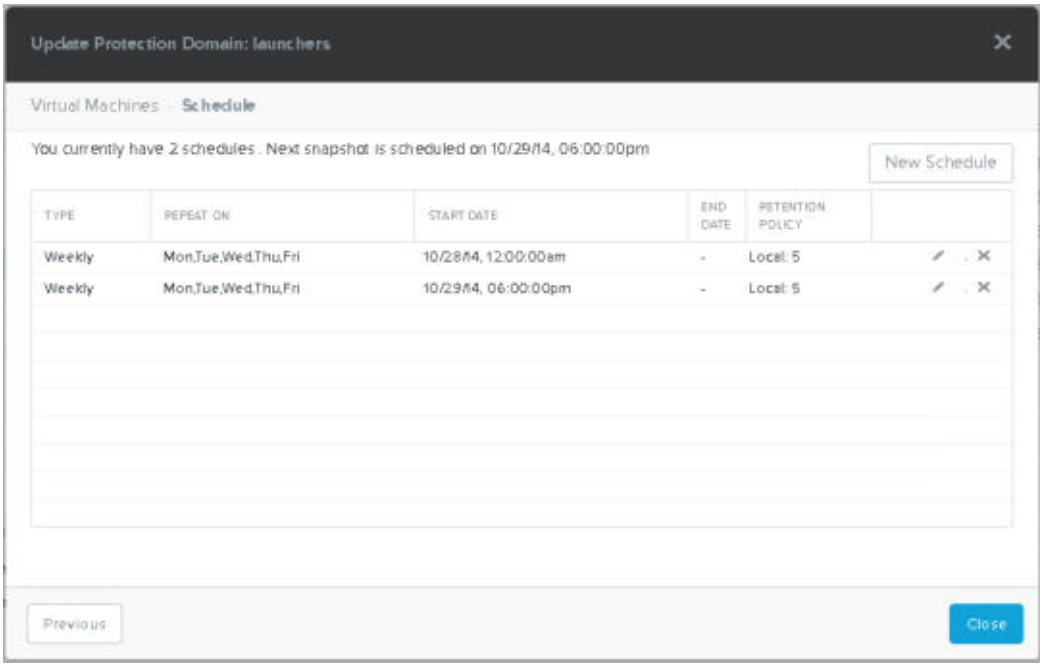


Figure 12: Example Snapshot Schedule: Snapshot at Noon and 6 PM

Using the local snapshot reserve formula presented above and assuming for demonstration purposes that the change rate is 35 GB of data every six hours and that we keep ten snapshots:

```
snapshot reserve = (frequency of snapshots × change rate per frequency) +
(change rate per frequency × # of snapshots in a full curator scan × 0.1)
= (10 × 35,980 MiB) + (35,980 MiB × 1 × 0.1)
= 359,800 + (35,980 × 1 × 0.1)
= 359,800 + 3,598
= 363,398 MiB
= 363 GiB
```

Asynchronous Replication

Asynchronous replication uses the same process, but you must include the first full copy of the PD plus delta changes based on the set schedule.

Table 5: Remote Snapshot Reserve Formula

snapshot reserve = (frequency of snapshots × change rate per frequency) + (change rate per frequency × # of snapshots in a full curator scan × 0.2) + total size of the source PD

For the minimum amount of space needed at the remote side, 130 percent of the PD is a good average to work from.

If the remote target also runs a workload, incoming replication uses the performance tier. If you use a hybrid cluster, size for the additional hot data. You can also skip the performance tier by creating a separate container for incoming replication and following the steps provided in the Remote Container section.

Lightweight Snapshots (LWS)

Sizing for LWS is very similar to sizing for full snapshots in that you must account for change rate and retention time. During the LWS retention time, you don't have to size for transformed garbage space, but LWS does use additional SSD space that you have to size for. By default, 7 percent of each node's SSD space forms the LWS reserve, which is an additional factor.

Duplicating the workload from the full snapshot example, let's change the RPO to one minute. LWS data exists for 75 minutes plus one extra snapshot based on the RPO value. Because the frequency rate in this scenario is high, it's very important to account for overwrites in the change rate. Because all data goes through the oplog, it's compressed; the type and amount of compression varies by workload. Using inline compression, we typically see a 2:1 compression rate.

Example Using 35 GB Change Rate Over 6 Hours

If your change rate is 35 GB of data every six hours, add 5 GB to account for overwrites. 40 GB every six hours has a change rate of approximately 114 MB per minute. If the cluster is running with replication factor 2, you must account for the total physical space.

© 2022 Nutanix, Inc. All rights reserved

Table 6: LWS Reserve

LWS reserve	= (frequency of snapshots × change rate per frequency) × oplog compression × replication factor
	= (75 × 114 MB per minute) × 0.50 × 2
	= 8,550 MB
	= 8.6 GB

If you run a 3460 hybrid system with two 1.9 TB SSDs, your LWS reserve is shown with the following calculation.

Table 7: Cluster LWS Reserve

LWS cluster reserve	= ((total SSD capacity per node - (CVM + metadata + oplog + cache overhead)) × number of nodes) × LWS reserve percentage
	= (3,539 GiB - (120 GiB + 30 GiB + 200 GiB + 40 GiB)) × 4 × 0.07
	= 3,149 GB × 4 × 0.07
	= 881.72 GiB
	= ~945 GB

Apply the 7 percent of the SSD space used by the extent store after you account for the rest of the system. Because the LWS cluster reserve is 945 GB, this system has lots of room for the workload as well as for additional business-critical applications.

As we discussed in the Scheduling LWS and NearSync section, a telescopic schedule has a total of six hourly, seven daily, four weekly, and one monthly snapshots. You only need to account for additional garbage space for the dailies because of the higher frequency. Using the change rate above, you can calculate each separate schedule and add them all together. Because full snapshots occur less often than LWS, you don't have to be concerned with overwrites and can use the original 35 GB change rate for every six hours.

Example of a Telescopic NearSync Snapshot Schedule

```
Hourly change rate:      5,833 MB
Daily change rate:     140,000 MB
```

```
Weekly change rate:    980,000 MB

Monthly change rate: 3,920,000 MB

snapshot overall capacity reserve = (hourly schedule + daily schedule + weekly
 schedule + monthly schedule)

For the hourly schedule:

(frequency of snapshots × change rate per frequency) + (change rate per frequency
 × # of snapshots in a full curator scan × 0.1)

The remaining schedules:

= (frequency of snapshots × change rate per frequency)

snapshot overall capacity reserve = ((6 × 5,833)+(5,833 × 6 × 0.1)) + (7 ×
 140,000) + (4 × 980,000) + (2 × 3,920,000)

= 34,998 + 3,500 + 980,000 + 3,920,000 + 8,257,538

= 13,196,036 MB

= ~13 TB
```

For most small and medium businesses, a daily change of 140 GB is considered high. This NearSync example highlights the difference between keeping a lot of snapshots around and keeping only the 10 snapshots in the full snapshot example.

NearSync

NearSync uses the same process as the snapshot process outlined above, but you must include the first full copy of the protection domain as well as delta changes based on the set schedule.

For the minimum amount of space needed at the remote site, start with 130 percent of the PD plus the required LWS reserve space.

Best practices:

- Use drives with the same or greater capacity at the remote site compared to those in your primary cluster to ensure that there's enough LWS reserve space.
- Don't enable deduplication on the source container for NearSync.

11. Bandwidth

You must have enough available bandwidth to keep up with the replication schedule. If you're still replicating when the next snapshot is scheduled, the current replication job finishes first. The newest outstanding snapshot then starts to get the newest data to the remote side first. To help replication run faster when you have limited bandwidth, you can seed data on a secondary cluster at the primary site before shipping that cluster to the remote site.

Seeding

To seed data for a new site:

- Set up a secondary cluster with local IPs at the primary site.
- Enable compression on the remote site in the production domain.
- Set the initial retention time to 3 months.
- Once setup completes, reconfigure the secondary cluster with remote IPs.
- Shut down the secondary cluster and ship it to the remote site.
- Power on the remote cluster and update the remote site on the primary cluster to the new IP.

If you can't seed the protection domain at the local site, you can create the remote cluster as a normal install and turn on compression over the wire. Manually create a one-time replication with retention time set to 3 months. We recommend this retention time setting because of the extra time it takes to replicate the first data set across the wire.

To figure out the needed throughput, you must know your RPO. If you set the RPO to one hour, you must be able to replicate the changes within that time.

Assuming you know your change rate based on incremental backups or local snapshots, you can calculate the bandwidth needed. The next example uses a change rate of 15 GB and an RPO of one hour. We don't use deduplication in the calculation, partly so the dedupe savings can serve as a buffer in the overall calculation and partly because

© 2022 Nutanix, Inc. All rights reserved

the one-time cost for deduped data going over the wire has less impact once the data is present at the remote site. We assume an average of 30 percent bandwidth savings for compression on the wire.

Table 8: Bandwidth Sizing

Bandwidth needed = (RPO change rate × (1 - compression on wire savings %)) / RPO
Example: (15 GB × (1 - 0.3)) / 3,600s (15 GB × 0.7) / 3,600s 10.5 GB / 3,600s (10.5 × 1,000 MB) / 3,600s (changing to MBps) (10,500 MB) / 3,600s 10,500 MB / 3,600 = 2.92 MBps Bandwidth needed = 23.33 Mbps

You can easily perform the calculation online using http://www.wolframalpha.com/input/?i=(15+GB+*+(1-30%25))%2F(1+hour).

If you keep missing your replication schedule, either increase your bandwidth or your RPO. To allow more RPO flexibility, you can run different schedules on the same production domain; for example, have one daily replication schedule and create a separate schedule to take local snapshots every two hours.

© 2022 Nutanix, Inc. All rights reserved

12. Failover: Migrate vs. Activate

Nutanix offers two options for handling VMs and volume groups during failover: migrate and activate.

Use migrate when you know that both the production site and the remote site are still healthy. This option is only available on the active production site. Migrate shuts down the VM or volume group, takes another snapshot, and replicates the VM or volume group to the selected remote site.

The activate option for restoring a VM is only available on the remote site. When you select activate, the system uses the last snapshot on the remote side to bring up the VM regardless of whether the active production site is healthy. You can't sync any outstanding changes to the VM from the production site.

If you activate a PD because the primary site is down but the primary site comes back up after the failover, you can have a split-brain scenario. To resolve this situation, deactivate the PD on the former primary site. The following command is hidden from the nCLI because it deletes the VMs, but it resolves the split while keeping the existing snapshots:

```
ncli> pd deactivate_and_destroy_vms name=<protection_Domain_Name>
```

You can test a VM at the remote site without breaking replication using the restore or clone functionality.

Figure 13: Cloning a VM for Testing Using the Remote Local Snapshot Browser

Using the local snapshot browser on the inactive production domain at the remote site, choose the restore option to clone a VM to the datastore. Add a prefix to the VM's path.

Best practices:

- When you activate PDs on the remote site, use intelligent placement for Hyper-V and DRS for ESXi clusters. Intelligent placement evenly spreads out the VMs on boot during a failover. AOS powers on VMs uniformly at boot time.
- Install NGT on machines using volume groups.
- Configure the data services cluster IP on the remote cluster.

Protection Domain Cleanup

Because inactive protection domains still consume space in existing snapshots, you should remove any unused PDs to reclaim that space.

To remove a PD from a cluster, follow these steps:

1. Remove existing schedules.
2. Remove both local and remote snapshots.
3. Remove the VMs from the active PD.
4. Delete the PD.

© 2022 Nutanix, Inc. All rights reserved

13. Self-Service File Restore

Self-service file restore is available for ESXi and AHV. The goal is to offer self-service file-level restore with minimal support from infrastructure administrators. Once you've installed NGT, you can open a web browser and go to http://localhost:5000 to browse snapshots sorted by today, last week, last month, and user-defined criteria. Self-service works with both protection domains and Leap.

- Guest VMs should be:
  - Windows 2008, Windows 7, or a later version
  - Centos 6.5+ and 7.0+
  - Red Hat 6.5+ and 7.0+
  - OEL 6.5+
  - SLES 11+
- Install VMware Tools.
- Remove JRE 1.8 or later if installed with a previous release.
- Configure a cluster external IP address.
- Add the VM to a protection domain.
- Use the default **disk.EnableUUID = true** for the VM in advanced settings for ESXi.
- The VM must have an IDE-based CD-ROM configured for installation. On newer versions of ESXi, a SATA-based CD-ROM is added by default.
- Detach the mounted disk after you restore your files.

Limitations:

- For Linux VMs, logical volumes spanning multiple disks aren't supported.

© 2022 Nutanix, Inc. All rights reserved

14. Third-Party Backup Products

Nutanix provides its own hypervisor-agnostic changed-region tracking API that vendors can access using a REST API. Currently, Cohesity, Commvault, HYCU, and Rubrik provide backup support specifically for AHV. Nutanix systems can integrate with any backup vendor that supports vStorage APIs for Data Protection for ESXi. Nutanix also offers a VSS provider for Hyper-V backup software vendors. Visit the Nutanix Elevate Technology Alliance Partner Program page for additional details on certified third-party solutions.

Backups with Replication

Note: Hypervisor-based snapshots left on the VM when you take a hardware-based snapshot may cause the recovery to fail. You must manually turn on the VM. Refer to VMware KB 1025279 for more information.

Backup software interacting with AHV uses scoped production domains, also known as backup snapshots. When the backup vendor requests a snapshot of a VM that isn't protected under any user-created production domain, temporarily protect the VM by placing it in a scoped production domain. Issue a snapshot of this production domain, then remove the VM from the production domain. Backup snapshots also allow you to back up VMs and use Leap disaster recovery orchestration. The only possible issue is that while the snapshot operation on the scoped production domain is in progress, the user-initiated VM protection may fail in the regular production domain.

Best practices:

- Avoid using hypervisor-based snapshots.
- Try to schedule hypervisor snapshots or user-created hardware-based snapshots at different times than the backup window.

Nutanix has published recommendations for optimizing and scaling third-party backup solutions in best practices guides focused on an enhanced disk-to-disk backup architecture. Find these guides on the Nutanix Support Portal.

© 2022 Nutanix, Inc. All rights reserved

15. Conclusion

Nutanix offers granular data protection based on the required recovery point objectives across many different deployment models. As your application requirements change and your cluster grows, you have the flexibility to add and remove VMs from protection domains. Sizing capacity and bandwidth are key to achieving optimal data protection and maintaining cluster health. Snapshot frequency and the daily change rate affect the capacity and bandwidth needed between sites to meet the needs of the business.

With data protection features that are purpose-built, fully integrated, and 100 percent software defined, Nutanix provides the ultimate in adaptability and flexibility for meeting your enterprise's backup and recovery needs.

© 2022 Nutanix, Inc. All rights reserved

Appendix

Best Practices Checklist

Following are high-level best practices for backup and disaster recovery on Nutanix.

General

- All VM files should sit on Nutanix storage. If non-Nutanix storage stores files externally, it should have the same file path on both sides.
- Remove all external devices, including ISOs or floppy devices.

Nutanix Native VSS Snapshots

- Configure an external cluster IP address.
- Guest VMs should be able to reach the external cluster IP on port 2074.
- Guest VMs should have an empty IDE CD-ROM for attaching NGT.
- Only available for ESXi and AHV.
- Virtual disks must use the SCSI bus type.
- VSS must be running in the guest VM. The appendix has a Check for VSS Service PowerShell script that verifies whether the service is running.
- The guest VM needs to support the use of VSS writers. The appendix has a Check VSS Writers PowerShell script that makes sure the VSS writers are stable.
- Schedule application-consistent snapshots during off-peak hours or ensure that additional I/O performance is available. If you take a VSS snapshot during peak usage, the delta disk from the hypervisor-based snapshot could become large. When you delete the hypervisor-based snapshot, collapsing it takes additional I/O; account for this additional I/O to avoid affecting performance.

Hyper-V VSS Provider

- VSS support is only for backup.

© 2022 Nutanix, Inc. All rights reserved

- Create different containers for VMs that need VSS backup support. Limit the number of VMs on each container to 50.
- Create a separate large container for crash-consistent VMs.

Protection Domains

- Protection domain names must be unique across sites.
- Group VMs with similar RPO requirements.
- Maximum of 200 VMs per protection domain for full snapshots.
- Maximum of 10 VMs per protection domain for LWS.
- VMware Site Recovery Manager and Metro Availability protection domains are limited to 50 VMs.
- Linked clone VMs (typically nonpersistent View desktops) aren't supported with NearSync.
- Remove unused protection domains to reclaim space.
- If you must activate a protection domain rather than migrate it, deactivate the old primary protection domain when the site comes back up.

Consistency Groups

- Keep consistency groups as small as possible. Keep dependent applications or service VMs in one consistency group to ensure that they are recovered in a consistent state (for example, App and DB).
- Each consistency group using application-consistent snapshots can contain only one VM.

Disaster Recovery and Backup

- Ensure that you configure forward (DNS A) and reverse (DNS PTR) DNS entries for each ESXi management host on the DNS servers used by the Nutanix cluster.

Remote Sites

- Use the external cluster IP as the address for the remote site.
- Use disaster recovery proxy to limit firewall rules.

© 2022 Nutanix, Inc. All rights reserved

- Use max bandwidth to limit replication traffic.
- When activating protection domains, use intelligent placement for Hyper-V and DRS for ESXi clusters on the remote site. Intelligent placement evenly spreads out the VMs on boot during a failover. AOS powers on VMs uniformly at boot time.
- If you use vCenter Server to manage both the primary and remote sites, don't use storage containers with the same name on both sites.

Remote Containers

- Create a new remote container as the target for the VStore mapping.
- When you back up many clusters to one destination cluster, use only one destination container if the source containers have similar advanced settings.
- Enable MapReduce compression if licensing permits.
- If the aggregate incoming bandwidth required to maintain the current change rate is less than 500 Mbps, we recommend skipping the performance tier to save flash capacity and increase device longevity.

Network Mapping

- Whenever you delete or change the network attached to a VM specified in the network map, modify the network map accordingly.

Scheduling

- To spread out replication impact on performance and bandwidth, stagger replication schedules across protection domains. If you have a PD starting at the top of the hour, stagger the PDs by half of the most common RPO.
- Configure snapshot schedules to retain the smallest number of snapshots while still meeting the retention policy.
- Metro and SRM-protected containers aren't supported.
- Deduplication on the source container isn't currently supported for NearSync.

Cross-Hypervisor Disaster Recovery

- Configure CVM external IP address.
- Obtain the mobility driver from NGT.

© 2022 Nutanix, Inc. All rights reserved

- Don’t migrate VMs with delta disks (hypervisor-based snapshots), using SATA disks, or using volume groups.
- Ensure that protected VMs have an empty IDE CD-ROM attached.
- Ensure that network mapping is complete.

Disaster Recovery Orchestration

- Deploy Prism Central to each on-prem site.
- Deploy Prism Central on a subnet that doesn’t fail over.
- Place CVM and hypervisor IPs on a different subnet than the subnets used by VMs.
- On-prem disaster recovery orchestration requires a nonroutable VLAN for the test network.

Availability Zones

- If one of the availability zones becomes unavailable or if a service in the paired availability zone is down, perform a forced sync from the paired availability zone after the issue is resolved.

Protection Polices

- A VM can only belong to either a protection domain or a protection policy.
- If you don’t use Nutanix AHV IPAM and need to retain your IP addresses, install NGT on the VMs to be protected.
- Use categories to apply protection policies.
- For on-premises Leap, create the same container name on both sides. If the container name doesn’t match on both sides, data replicates by default to the SelfServiceContainer.

Cross-Hypervisor Support with ESXi for Protection Polices

- No support for UEFI boot.
- Doesn’t preserve HA and DRS settings.
- No support for hypervisor-based snapshots, VMware VSS, or linked clones.

© 2022 Nutanix, Inc. All rights reserved

Recovery Plans

- For on-premises availability zones, create a nonroutable network for testing failovers.
- Run the Validate workflow after making changes to recovery plans.
- After you run the Test workflow, run the Clean-Up workflow instead of manually deleting VMs.
- A recovery plan should cover a maximum of 200 VMs.
- Maximum of 50 categories in a recovery plan.
- Maximum of 20 stages in a recovery plan.
- Maximum of 15 categories per stage in a recovery plan.
- You can run a maximum of 5 recovery plans in parallel.

Network Mapping

- Set up administrative distances on VLANs for subnets that completely fail over. If you don't set up administrative distances, shut down the VLAN on the source side after failover if the VPN connection is maintained between the two sites. If you're failing over to a new subnet, set up the subnet beforehand so you can test the routing.
- The prefix length for network mappings at the source and the destination must be the same.
- If you don't use Nutanix IPAM, you must install the NGT software package to maintain a static address.
- To maintain a static address for Linux VMs that don't use Nutanix IPAM, the VMs must have the NetworkManager command-line tool (nmcli) version 0.9.10.0 or later installed. Additionally, you must use NetworkManager to manage the network for the Linux VMs. To enable NetworkManager on a Linux VM, set the value of the **NM_CONTROLLED** field to yes in the interface configuration file (for example, in CentOS, the file is /etc/sysconfig/network-scripts/ifcfg-eth0). After you set the field, restart the network service on the VM.

Xi Leap Hypervisor Support

- Xi Leap only supports clusters running AHV.

© 2022 Nutanix, Inc. All rights reserved

Xi Leap Virtual Machine Configuration Restrictions

- Can’t power on VMs configured with a GPU resource.
- Can’t power on VMs configured with four vNUMA sockets.

Single-Node Backup

- Combined, all protections domains should be under 30 VMs.
- Limit backup retention to a three-month policy. We recommended seven daily, four weekly, and three monthly backups.
- Only map an NX-1155 to one physical cluster.
- Snapshot schedule should be at least six hours.
- Turn off deduplication.

Cloud Connect

- Try to limit each protection domain to one VM to speed up restores. This approach also saves money, as it limits the amount of data going across the WAN.
- The RPO shouldn’t be lower than four hours.
- Turn off deduplication.
- Try to use Cloud Connect to protect workloads that have an average change rate of less than 0.5 percent.

Sizing

- Size local and remote snapshot usage using the application’s change rate.
- Remember to either size the performance tier for hybrid clusters to accommodate incoming data or bypass the performance tier and write directly to disk.

Bandwidth

- Seed locally for replication if WAN bandwidth is limited.
- Set a high initial retention time for the first replication when you seed.

File-Level Restore

- Guest VMs should be Windows 2008, Windows 7, or a later version.

© 2022 Nutanix, Inc. All rights reserved

- Install VMware tools.
- Install JRE 1.8 or later.
- Configure a cluster external IP address.
- Add the VM to a protection domain.
- Use the default **disk.EnableUUID = true** for the VM in advanced settings.
- The VM must have a CD-ROM configured.
- Detach the mounted disk after restoring your files.

PowerShell Scripts

Check for VSS Service

This PowerShell script checks whether the VSS service is running as required for application-consistent snapshots.

```
#Connect to the Nutanix cluster of your choice, try to use the external address.

Connect-NutanixCluster -AcceptInvalidSSLCerts -server External_cluster_IP -
UserName admin

#load Nutanix CMDlets, make sure your local version matches the cluster version

Add-PSSnapin NutanixCmdletsPSSnapin

#Get a list of all Consistency Groups

$pdvss = Get-NTNXProtectionDomainConsistencyGroup

#array of all the appConsistentVMs

$appConsistentVM = @()

Foreach ($vssVM in $pdvss)

  {

    if ($vssVM.appConsistentSnapshots)

  {

     $appConsistentVM += $vssVM.consistencyGroupName

    }

  }

 get-service -name VSS -computername $appConsistentVM | format-table -property
 MachineName, Status, Name, DisplayName -auto
```

Check VSS Writers

This PowerShell script checks whether VSS writers are stable for VMs that run application-consistent snapshots.

```
function Get-VssWriters {
<#
 .Synopsis
 Function to get information about VSS Writers on one or more computers
 .Description
 Function will parse information from VSSAdmin tool and return object containing
 WriterName, StateID, StateDesc, and LastError
 Function will display a progress bar while it retrives information from
different
 computers.
 .Parameter ComputerName
 This is the name (not IP address) of the computer.
 If absent, localhost is assumed.
 .Example
 Get-VssWriters
 This example will return a list of VSS Writers on localhost
 .Example
 # Get VSS Writers on localhost, sort list by WriterName
 $VssWriters = Get-VssWriters | Sort "WriterName"
 $VssWriters | FT -AutoSize # Displays it on screen
 $VssWriters | Out-GridView # Displays it in GridView
 $VssWriters | Export-CSV ".\myReport.csv" -NoTypeInformation # Exports it to CSV
 .Example
 # Get VSS Writers on the list of $Computers, sort list by ComputerName
 $Computers = "xHost11","notThere","xHost12"
 $VssWriters = Get-VssWriters -ComputerName $Computers -Verbose | Sort
"ComputerName"
 $VssWriters | Out-GridView # Displays it in GridView
 $VssWriters | Export-CSV ".\myReport.csv" -NoTypeInformation # Exports it to CSV
 .Example
```

```
# Reports any errors on VSS Writers on the computers listed in
MyComputerList.txt, sorts list by ComputerName

$Computers = Get-Content ".\MyComputerList.txt"

$VssWriters = Get-VssWriters $Computers -Verbose |

 Where { $_.StateDesc -ne 'Stable' } | Sort "ComputerName"

$VssWriters | Out-GridView # Displays it in GridView

$VssWriters | Export-CSV ".\myReport.csv" -NoTypeInformation # Exports it to
CSV

.Example

# Get VSS Writers on all computers in current AD domain, sort list by
ComputerName

$Computers = (Get-ADComputer -Filter *).Name

$VssWriters = Get-VssWriters $Computers -Verbose | Sort "ComputerName"

$VssWriters | Out-GridView # Displays it in GridView

$VssWriters | Export-CSV ".\myReport.csv" -NoTypeInformation # Exports it to CSV

.Example

# Get VSS Writers on all Hyper-V hosts in current AD domain, sort list by
ComputerName

$FilteredComputerList = $null

$Computers = (Get-ADComputer -Filter *).Name

Foreach ($Computer in $Computers) {

  if (Get-WindowsFeature -ComputerName $Computer -ErrorAction SilentlyContinue
|

   where { $_.Name -eq "Hyper-V" -and $_.InstallState -eq "Installed"}) {

    $FilteredComputerList += $Computer

  }

}

$VssWriters = Get-VssWriters $FilteredComputerList -Verbose | Sort
"ComputerName"

$VssWriters | Out-GridView # Displays it in GridView

$VssWriters | Export-CSV ".\myReport.csv" -NoTypeInformation # Exports it to CSV

.OUTPUT

Scripts returns a PS Object with the following properties:

 ComputerName


 WriterName
```

```
  StateID

  StateDesc

  LastError

 .Link
 https://superwidgets.wordpress.com/category/powershell/
 .Notes
 Function by Sam Boutros
 v1.0 - 09/17/2014
#>
[CmdletBinding(SupportsShouldProcess=$true,ConfirmImpact='Low')]
  Param(
    [Parameter(Mandatory=$false,
         ValueFromPipeLine=$true,
         ValueFromPipeLineByPropertyName=$true,
         Position=0)]
      [ValidateNotNullorEmpty()]
      [String[]]$ComputerName = $env:COMPUTERNAME
  )
  $Writers = @()
  $k = 0
  foreach ($Computer in $ComputerName) {
    try {
      Write-Verbose "Getting VssWriter information from computer $Computer"
      $k++
      $Progress = "{0:N0}" -f ($k*100/$ComputerName.count)
      Write-Progress -Activity "Processing computer $Computer ... $k out of
 $($ComputerName.count) computers" `
        -PercentComplete $Progress -Status "Please wait" -CurrentOperation
 "$Progress% complete"
      $RawWriters = Invoke-Command -ComputerName $Computer -ErrorAction Stop -
ScriptBlock {
        return (VssAdmin List Writers)
      }
```

```
      for ($i=0; $i -lt ($RawWriters.Count-3)/6; $i++) {
        $Writer = New-Object -TypeName psobject
        $Writer| Add-Member "ComputerName" $Computer
        $Writer| Add-Member "WriterName" $RawWriters[($i*6)+3].Split("'")[1]
        $Writer| Add-Member "StateID" $RawWriters[($i*6)+6].SubString(11,1)
        $Writer| Add-Member "StateDesc" $RawWriters[($i*6)+6].SubString(14,
$RawWriters[($i*6)+6].Length - 14)
        $Writer| Add-Member "LastError" $RawWriters[($i*6)+7].SubString(15,
$RawWriters[($i*6)+7].Length - 15)
        $Writers += $Writer
      }
      Write-Debug "Done"
    } catch {
      Write-Warning "Computer $Computer is offline, does not exist, or cannot be
 contacted"
    }
  }
  return $Writers
}
#Connect to the Nutanix cluster of your choice, try to use the external address.
Connect-NutanixCluster -AcceptInvalidSSLCerts -server External_cluster_IP -
UserName admin
#load Nutanix CMDlets, make sure your local version matches the cluster version
Add-PSSnapin NutanixCmdletsPSSnapin
#Get a list of all Consistency Groups
$pdvss = Get-NTNXProtectionDomainConsistencyGroup
#array of all the appConsistentVMs
$appConsistentVM = @()
Foreach ($vssVM in $pdvss)
  {
    if ($vssVM.appConsistentSnapshots)
  {
     $appConsistentVM += $vssVM.consistencyGroupName
    }
  }
```

```
$VssWriters = Get-VssWriters $appConsistentVM -Verbose |
  Where { $_.StateDesc -ne 'Stable' } | Sort "ComputerName"
$VssWriters | Out-GridView # Displays it in GridView
$VssWriters | Export-CSV ".\vssWriterReport.csv" -NoTypeInformation # Exports it
 to CSV
```

About the Author

Dwayne Lessner is a Principal Technical Marketing Engineer on the Product and Technical Marketing team at Nutanix. Follow Dwayne on Twitter at @dlink7.

Mike Umphreys is a Sr. Technical Marketing Engineer on the Product and Technical Marketing team at Nutanix.

About Nutanix

Nutanix makes infrastructure invisible, elevating IT to focus on the applications and services that power their business. The Nutanix enterprise cloud software leverages web-scale engineering and consumer-grade design to natively converge compute, virtualization, and storage into a resilient, software-defined solution with rich machine intelligence. The result is predictable performance, cloud-like infrastructure consumption, robust security, and seamless application mobility for a broad range of enterprise applications. Learn more at www.nutanix.com or follow us on Twitter @nutanix.

© 2022 Nutanix, Inc. All rights reserved

List of Figures

Figure 1: Nutanix Enterprise Cloud OS Stack....9
Figure 2: Scalable Replication....10
Figure 3: Two-Way Mirroring....12
Figure 4: Many-to-One Architecture....13
Figure 5: Public Cloud as a Backup Destination....14
Figure 6: On-Prem, Cloud, and Xi-Based Availability Zones....15
Figure 7: Setup Options for a Remote Site....23
Figure 8: Max Bandwidth Settings....25
Figure 9: VStore Container Mappings for Replication....26
Figure 10: Multiple Schedules for a Production Domain....28
Figure 11: Network Mapping for Xi Leap....36
Figure 12: Example Snapshot Schedule: Snapshot at Noon and 6 PM....39
Figure 13: Cloning a VM for Testing Using the Remote Local Snapshot Browser....45

List of Tables

Table 1: Document Version History....7
Table 2: MS Failover Settings Adjusted for Using VSS....18
Table 3: Default Telescopic Schedule for One Month....29
Table 4: Local Full Snapshot Reserve Formula....38
Table 5: Remote Snapshot Reserve Formula....40
Table 6: LWS Reserve....41
Table 7: Cluster LWS Reserve....41
Table 8: Bandwidth Sizing....44

v2.3 | May 2021 | BP-2050

BEST PRACTICES

Physical Networking

NUTANIX
YOUR ENTERPRISE CLOUD

Copyright

Copyright 2021 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

Contents

1. Executive Summary 5

2. Introduction 6
- Audience 6
- Purpose 6

3. Nutanix Enterprise Cloud Overview 8
- Nutanix HCI Architecture 9

4. Choosing a Physical Switch 11

5. Network Design Requirements and Recommendations 14
- Maximum of Three Switch Hops 14
- Same Switch Fabric 14
- WAN Links 14
- VLANs 15
- Stretched Layer 2 Networking 15
- Rack Awareness and Block Awareness 16
- Oversubscription 16
- Host Connections 17

6. Nutanix-Recommended Network Designs 19
- Leaf-Spine 19
- Core-Aggregation-Access 20
- Multisite Designs 22

7. Conclusion 24

Appendix 25
- Best Practices Checklist 25
- References 26
- About the Authors 26

About Nutanix..... 26
List of Figures.....28
List of Tables.....29

1. Executive Summary

Nutanix maximizes modern datacenter network technology to provide a highly available virtualization, compute, and storage platform. The network is a key component in ensuring high performance and availability, and successful Nutanix deployments combine the right physical switches with the right network designs. At the same time, certain physical switches and network designs can lead to poor performance or poor availability, and this guide helps you avoid those risks.

© 2021 Nutanix, Inc. All rights reserved

2. Introduction

Well-designed Nutanix networks are central to Nutanix's resilience and performance. A Nutanix cluster can tolerate multiple simultaneous failures because it maintains a set replication factor and offers features like block and rack awareness. However, this level of resilience requires a highly available, redundant network between the cluster's nodes. Protecting the cluster's read and write storage capabilities also requires highly available connectivity between nodes. Even with intelligent data placement strategies, if network connectivity between more than the allowed number of nodes breaks down, VMs on the cluster could experience write failures and enter read-only mode.

To optimize I/O speed, Nutanix clusters choose to send each write to another node in the cluster. As a result, a fully populated cluster sends storage replication traffic in a full mesh, using network bandwidth between all Nutanix nodes. Because storage write latency directly correlates to the network latency between Nutanix nodes, any network latency increase adds to storage write latency.

Audience

This best practices guide is part of the Nutanix Solutions Library. We intend it for network and virtualization administrators and architects responsible for designing networks for a Nutanix environment. Readers of this document should already be familiar with top-of-rack networking concepts for datacenter environments and basic hypervisor networking.

Purpose

In this document, we cover the following topics:

- Choosing physical switches.

© 2021 Nutanix, Inc. All rights reserved

- Network design requirements and recommendations:
  - Switch fabric scale.
  - VLANs.
  - Stretched networks and WANs.
  - Oversubscription.
  - Rack awareness and block awareness.
- Recommended network designs:
  - Leaf-spine.
  - Core-aggregation-access (multi-tier).
  - Multisite designs.

Table 1: Document Version History

Version Number	Published	Notes
1.0	March 2019	Original publication.
2.0	November 2019	QoS and VLAN updates.
2.1	March 2020	Updated the Nutanix Enterprise Cloud Overview and Choosing a Physical Switch sections.
2.2	January 2021	Updated the Choosing a Physical Switch section.
2.3	May 2021	Added spanning tree warning.

© 2021 Nutanix, Inc. All rights reserved

3. Nutanix Enterprise Cloud Overview

Nutanix delivers a web-scale, hyperconverged infrastructure solution purpose-built for virtualization and both containerized and private cloud environments. This solution brings the scale, resilience, and economic benefits of web-scale architecture to the enterprise through the Nutanix enterprise cloud platform, which combines the core HCI product families—Nutanix AOS and Nutanix Prism management—along with other software products that automate, secure, and back up cost-optimized infrastructure.

Available attributes of the Nutanix enterprise cloud OS stack include:

- Optimized for storage and compute resources.
- Machine learning to plan for and adapt to changing conditions automatically.
- Intrinsic security features and functions for data protection and cyberthreat defense.
- Self-healing to tolerate and adjust to component failures.
- API-based automation and rich analytics.
- Simplified one-click upgrades and software life cycle management.
- Native file services for user and application data.
- Native backup and disaster recovery solutions.
- Powerful and feature-rich virtualization.
- Flexible virtual networking for visualization, automation, and security.
- Cloud automation and life cycle management.

Nutanix provides services and can be broken down into three main components: an HCI-based distributed storage fabric, management and operational intelligence from Prism, and AHV virtualization. Nutanix Prism furnishes one-click infrastructure management for virtual environments running on AOS. AOS is hypervisor agnostic, supporting two third-party hypervisors

© 2021 Nutanix, Inc. All rights reserved

—VMware ESXi and Microsoft Hyper-V—in addition to the native Nutanix hypervisor, AHV.

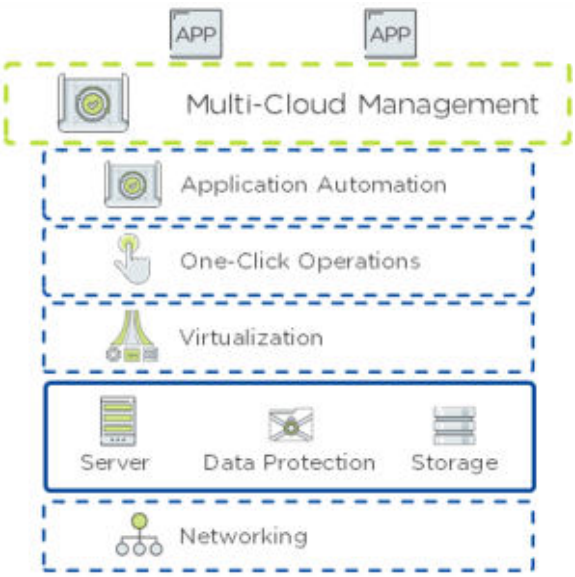


Figure 1: Nutanix Enterprise Cloud OS Stack

Nutanix HCI Architecture

Nutanix does not rely on traditional SAN or network-attached storage (NAS) or expensive storage network interconnects. It combines highly dense storage and server compute (CPU and RAM) into a single platform building block. Each building block delivers a unified, scale-out, shared-nothing architecture with no single points of failure.

The Nutanix solution requires no SAN constructs, such as LUNs, RAID groups, or expensive storage switches. All storage management is VM-centric, and I/O is optimized at the VM virtual disk level. The software solution runs on nodes from a variety of manufacturers that are either entirely solid-state storage with NVMe for optimal performance or a hybrid combination of SSD and HDD storage that provides a combination of performance and additional capacity. The storage fabric automatically tiers data across the cluster to different classes of storage devices using intelligent data placement algorithms. For best

performance, algorithms make sure the most frequently used data is available in memory or in flash on the node local to the VM.

To learn more about Nutanix enterprise cloud software, visit the Nutanix Bible and Nutanix.com.

© 2021 Nutanix, Inc. All rights reserved

4. Choosing a Physical Switch

A Nutanix environment should use datacenter switches designed for transmitting large amounts of server and storage traffic at low latency. Don't use switches meant for deployment at the campus access layer. Campus access switches may have 10 Gbps ports like datacenter switches, but they aren't usually made to transport a large amount of bidirectional storage replication traffic.

The deployment size and purpose also influence physical switch choice. Datacenter switches with large buffers are critical in a large AOS cluster that is going to grow beyond eight nodes or host storage-intensive applications. In smaller clusters or ROBO deployments that have fewer than eight nodes or don't host write-intensive applications, the switch may not experience buffer contention and you can relax these switch restrictions. There are also some switch types you should never use for any Nutanix deployment because of oversubscription or other architecture choices; we list examples of these as well.

Datacenter switches have the following characteristics:

- Line rate: Ensures that all ports can simultaneously achieve advertised throughput.
- Low latency: Minimizes port-to-port latency, measured in microseconds or nanoseconds.
- Large per-port buffers: Handle speed mismatch from uplinks without dropping frames.
- Nonblocking, with low or no oversubscription: Reduces chance of drops during peak traffic periods.
- 10 Gbps or faster links for Nutanix CVM traffic: Only use 1 Gbps links either for additional user VM traffic or when 10 Gbps connections are not available, such as in a ROBO deployment. Limit Nutanix clusters using 1 Gbps links to eight nodes.

© 2021 Nutanix, Inc. All rights reserved

The switch manufacturer's datasheets, specifications, and white papers can help identify these characteristics. For example, a common datacenter switch datasheet may show a per-port buffer of 1 MB, while an access layer or fabric extension device has a per-port buffer of around 150 KB. During periods of high traffic or when using links with a speed mismatch (such as 40 Gbps uplinks to 10 Gbps edge ports), a smaller buffer can lead to frame drops, increasing storage latency.

The following list is not exhaustive, but it gives some examples of model lines that meet the above requirements for high-performance or large clusters. Models similar to the ones listed are also great choices and can be used for smaller clusters and ROBO.

- Arista 7050X3, 7160, 7170, 7280: Larger buffer models
- Aruba CX 8325 and CX 8360
- Cisco Nexus 9000, 7000, and 5000
- Dell S5200-ON
- HPE FM3810, FM3132Q
- Juniper QFX5100, QFX5200, QFX10K
- Lenovo NE2580O
- Mellanox SN2410, SN2100, and SN2010

The following are examples of switches that don't meet the high-performance datacenter switch requirements but are acceptable for ROBO clusters and clusters with fewer than eight nodes or low performance requirements:

- Arista 7050 and 7150s: Smaller buffer models
- Cisco Nexus 3000: Smaller buffer model
- Cisco Catalyst 9300 and 9500: Campus access switch
- Cisco Catalyst 3850: Stackable multigigabit switch
- HPE FM2072

The following are examples of switches that are never acceptable for any Nutanix storage connection, but are acceptable for out-of-band management:

© 2021 Nutanix, Inc. All rights reserved

- Cisco Nexus 2000 (Fabric Extender): Highly oversubscribed with small per-port buffers.
- 10 Gbps expansion cards in a 1 Gbps access switch: These 10 Gbps expansion cards provide uplink bandwidth for the switch, not server connectivity.

Each Nutanix node also has an out-of-band connection for IPMI, iLO, iDRAC, or similar management. Because out-of-band connections don't have the latency or throughput requirements of VM networking or storage connections, they can use any access layer switch.

> Tip: Nutanix recommends an out-of-band management switch network separate from the primary network for high availability.

Nutanix maintains close partnerships with several networking vendors. For more information on the network integration and automation capabilities available with Nutanix, check out our list of partners in the Networking and Security section of our Technology Alliances page.

Regardless of the switch vendor you choose, follow the general recommendations in this document. For vendor-specific configuration recommendations refer to the References section in the Appendix.

© 2021 Nutanix, Inc. All rights reserved

5. Network Design Requirements and Recommendations

Maximum of Three Switch Hops

Nutanix nodes send storage replication traffic to each other in a distributed fashion over the top-of-rack network. One Nutanix node can therefore send replication traffic to any other Nutanix node in the cluster. The network should provide low and predictable latency for this traffic. Ensure that there are no more than three switches between any two Nutanix nodes in the same cluster.

Same Switch Fabric

A switch fabric is either a single leaf-spine topology or all switches connected to the same switch aggregation layer. The Nutanix VLAN shares a common broadcast domain within the fabric. Connect all Nutanix nodes that form a cluster to the same switch fabric. Don't stretch a single Nutanix cluster across multiple, disconnected switch fabrics.

Every Nutanix node in a cluster should be in the same L2 broadcast domain and share the same IP subnet.

WAN Links

A WAN (wide area network) or metro link connects different physical sites over a distance. As an extension of the switch fabric requirement, don't place Nutanix nodes in the same cluster if they're separated by a WAN.

© 2021 Nutanix, Inc. All rights reserved

VLANs

To protect the Nutanix CVM and hypervisor traffic, place them together in their own dedicated VLAN, separate from other VM traffic. Don't place the CVM and hypervisor hosts in a VLAN shared with other VMs.

Nutanix recommends that you configure the CVM and hypervisor host VLAN as the native, or untagged, VLAN on the connected switch ports. This native VLAN configuration allows easy node addition and cluster expansion. By default, new Nutanix nodes send and receive untagged traffic. If you use a tagged VLAN for the CVM and hypervisor hosts instead, you must configure that VLAN while you provision the new node, before you add that node to the Nutanix cluster.

Use tagged VLANs for all guest VM traffic and add the required guest VM VLANs to all connected switch ports for hosts in the Nutanix cluster. Limit guest VLANs for guest VM traffic to the smallest number of physical switches and switch ports possible to reduce broadcast network traffic load. If a VLAN is no longer needed, remove it. Nutanix AHV network automation makes VLAN provisioning for guest VMs automatic. For a list of partners that support network automation, visit the Technology Alliances site and filter on networking and security.

Stretched Layer 2 Networking

Datacenters have traditionally shared a VLAN with a layer 2 (L2) broadcast domain between different racks. Modern designs sometimes terminate this L2 boundary at the top of the rack, so that each rack is a separate L2 domain and a different IP subnet. Because all nodes in a Nutanix cluster must share the same L2 broadcast domain, this approach presents a challenge.

This scenario—stretching an L2 network between two racks in the same datacenter—is the only case when using a stretched L2 network over a layer 3 (L3) network is acceptable, because the Nutanix cluster is still in the same switch fabric or aggregation layer.

> Note: Don't place Nutanix nodes in the same Nutanix cluster if the stretched L2 network spans multiple datacenters or availability zones or if there is a remote link between the two locations.

© 2021 Nutanix, Inc. All rights reserved

When you need compute and storage at multiple sites, use separate Nutanix clusters at each physical location. Use replication tools such as asynchronous disaster recovery, NearSync, and Metro Availability to share data between Nutanix clusters at different sites. These recommendations apply even if the two availability zones share close proximity or highly available network paths. Each site should be a cluster boundary to protect against network, power, or other failures. The boundary could be a building, a firewall, or even different racks in the datacenter, depending on your availability requirements.

Rack Awareness and Block Awareness

Block awareness and rack awareness provide smart placement of Nutanix cluster services, metadata, and VM data to help maintain data availability even if you lose an entire block or rack. The same network requirements for low latency and high throughput between servers in the same cluster still apply when you use block and rack awareness.

> Note: Don't use features like block or rack awareness to stretch a Nutanix cluster between different physical sites.

Oversubscription

Oversubscription occurs when an intermediate network device or link doesn't have enough capacity to allow line rate communication between the systems connected to it. For example, if a 10 Gbps link connects two switches and four hosts connect to each switch at 10 Gbps, the connecting link is oversubscribed. Oversubscription is often expressed as a ratio—in this case 4:1, as the environment could potentially attempt to transmit 40 Gbps between the switches with only 10 Gbps available. Achieving a 1:1 ratio isn't always feasible, but you should keep the ratio as small as possible based on budget and available capacity.

In a typical deployment where Nutanix nodes connect to redundant top-of-rack switches, storage replication traffic between CVMs traverses multiple devices. To avoid packet loss caused by link oversubscription, ensure that the switch uplinks consist of multiple interfaces operating at a faster speed than

© 2021 Nutanix, Inc. All rights reserved

the Nutanix host interfaces. For example, for nodes connected at 10 Gbps, the interswitch connection should consist of multiple 10 or 40 Gbps links.

Host Connections

Connect Nutanix hosts to redundant top-of-rack switches. Use the active-active configuration available in the hypervisor when possible. With vSphere, follow the VMware vSphere Networking best practices, using Route based on Originating Virtual Port or Route based on Physical NIC Load. With AHV, follow the Nutanix AHV Networking best practices, using either the default active-backup configuration or LACP with balance-tcp.

On switch ports that face Nutanix hosts, designate the interfaces as spanning tree edge ports to minimize port downtime and prevent triggering spanning tree topology changes.

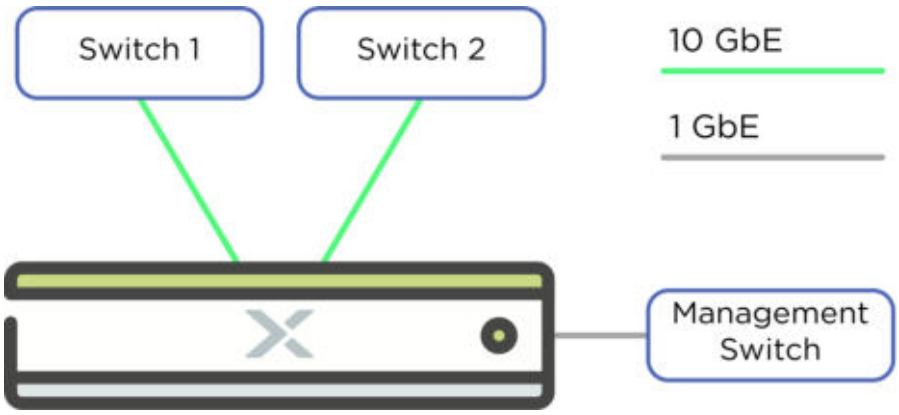


Figure 2: Host Network Connections

Quality of Service

When there is network traffic contention, quality of service (QoS) prioritizes traffic and reserves network resources for certain classes of traffic. QoS can mark important traffic with priority values or enable network switches to derive the traffic's class based on fields inside the traffic. Once you have defined traffic classes, each network device can reserve resources for certain classes or decide which classes to prioritize and which to drop.

© 2021 Nutanix, Inc. All rights reserved

With a fabric that eliminates contention for storage traffic, no QoS configuration is required. Nutanix doesn't recommend configuring any top-of-rack QoS for storage traffic. Instead, configure a nonblocking, low-latency switch fabric with no oversubscription between Nutanix nodes.

Nutanix doesn't recommend that you customize any QoS markings on traffic sourced from the CVM or the hypervisor host. If guest VMs pass frames with L2 CoS (class of service) or L3 DSCP (differentiated services code point) markers on them, these values pass from the hypervisor to the physical switch.

Nutanix AHV doesn't support configuration of enforcement for QoS prioritization or reservations.

With Nutanix on ESXi, refer to the Nutanix on VMware vSphere Networking guide for Network I/O Control (NIOC) recommendations. If you need QoS at the hypervisor level, Nutanix recommends using shares without hard limits, so QoS with NIOC only applies when there is network contention.

Spanning Tree

In traditional networks, using the spanning tree protocol prevents network loops and enforces a loop-free network switch topology. Not all networks use spanning tree, but if yours does, treat the switch ports facing Nutanix hosts as server or host ports, sometimes referred to as edge ports. A Nutanix server acts like an end host, not a switch. Nutanix recommends either disabling spanning tree on the ports facing Nutanix servers or configuring the Nutanix-facing ports with features such as type edge or portfast, so these ports skip the normal spanning tree phases and immediately forward traffic.

> Note: Configure switch ports facing Nutanix servers as portfast or type edge. Failure to do so may cause service interruption while switch ports transition through the listening and learning phases, dropping all production traffic for up to 30 seconds during spanning tree topology changes.

6. Nutanix-Recommended Network Designs

Leaf-Spine

The leaf-spine network design is popular in new datacenter deployments because it's easy to deploy and easy to scale after deployment. A leaf-spine topology requires at least two spine switches and two leaf switches. Every leaf connects to every spine using uplink ports. There are no connections between the spine switches or between the leaf switches in the conventional leaf-spine design.

Use uplinks that are a higher speed than the edge ports to reduce uplink oversubscription. To increase uplink capacity, add spine switches or uplink ports as needed.

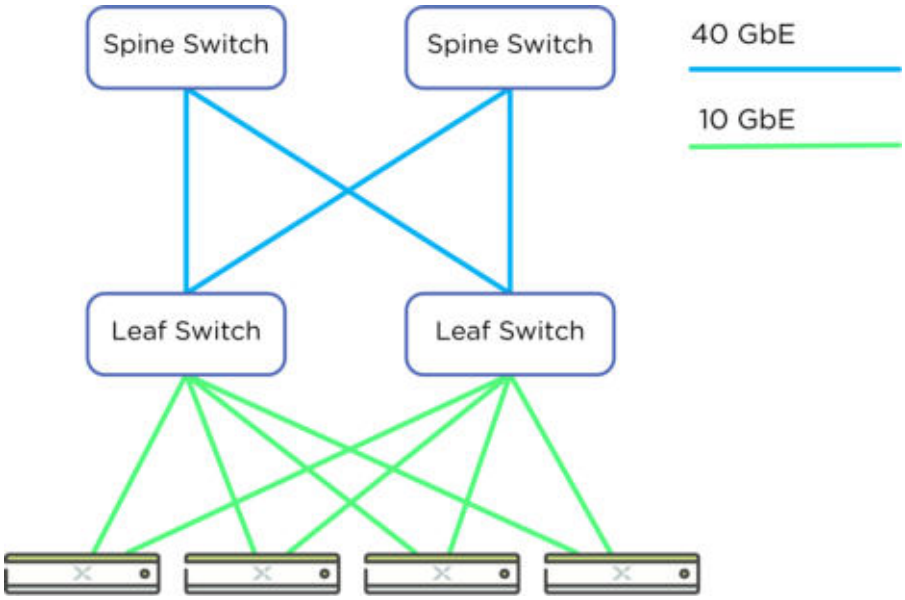


Figure 3: Leaf-Spine Network

© 2021 Nutanix, Inc. All rights reserved

To scale the leaf-spine network, add leaf and spine switches. Because there are no more than three switch hops between any two Nutanix nodes in this design, a Nutanix cluster can easily span multiple racks and still connect to the same switch fabric.

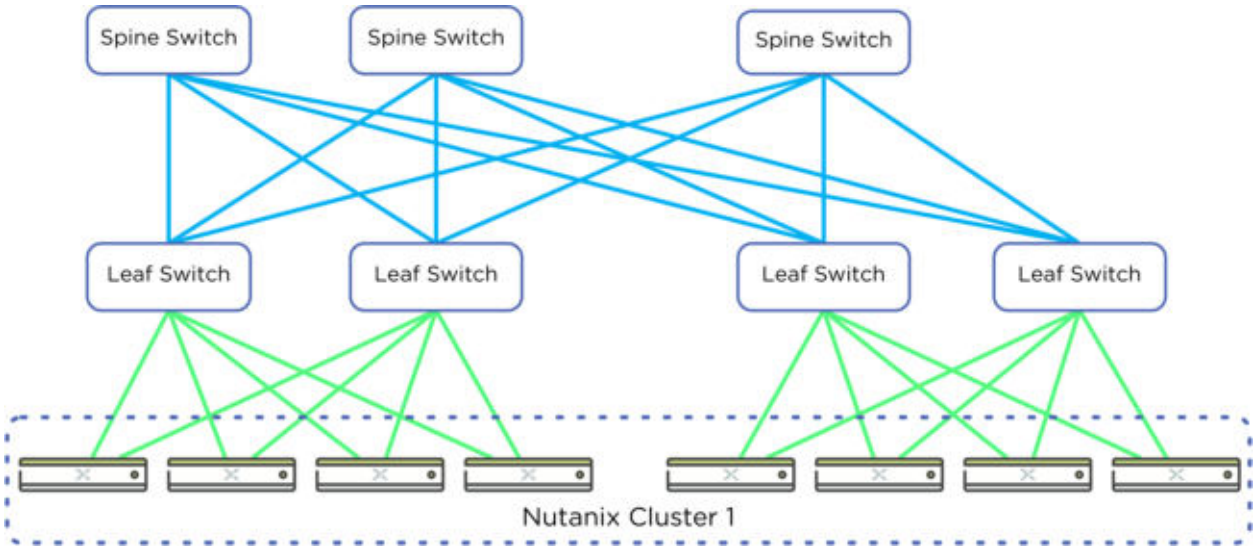


Figure 4: Scaling the Leaf-Spine Network

Core-Aggregation-Access

The core-aggregation-access (or three-tier) design is a modular layout that allows you to upgrade and scale layers independently. Ensure that all nodes in a Nutanix cluster share the same aggregation layer to meet the three-switch-hop rule.

© 2021 Nutanix, Inc. All rights reserved

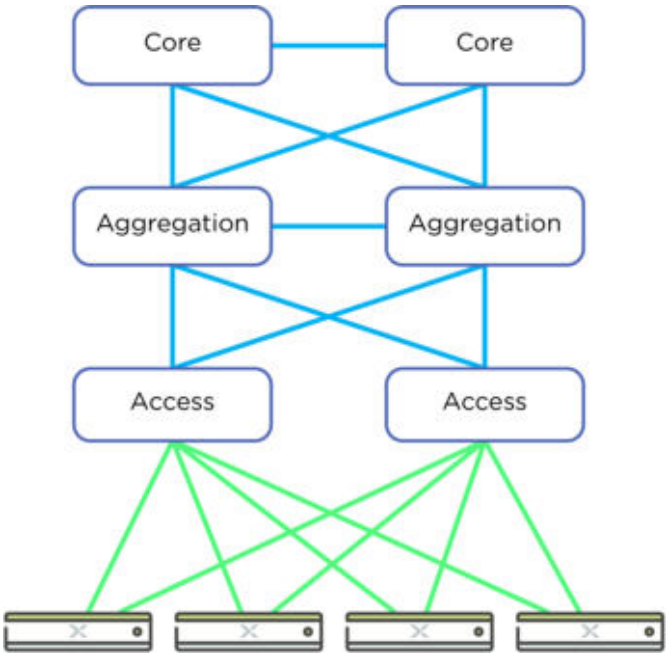


Figure 5: Core-Aggregation-Access Network

Scaling the three-tier network design may require adding another aggregation and access layer to the core. In this case, there would be more than three switch hops between the two access layers. Ensure that you add Nutanix nodes in separate aggregation and access layers to separate clusters to keep the number of switch hops between nodes in the same cluster to three or fewer. In the following example, Cluster 1 connects to one aggregation layer and Cluster 2 connects to another.

© 2021 Nutanix, Inc. All rights reserved

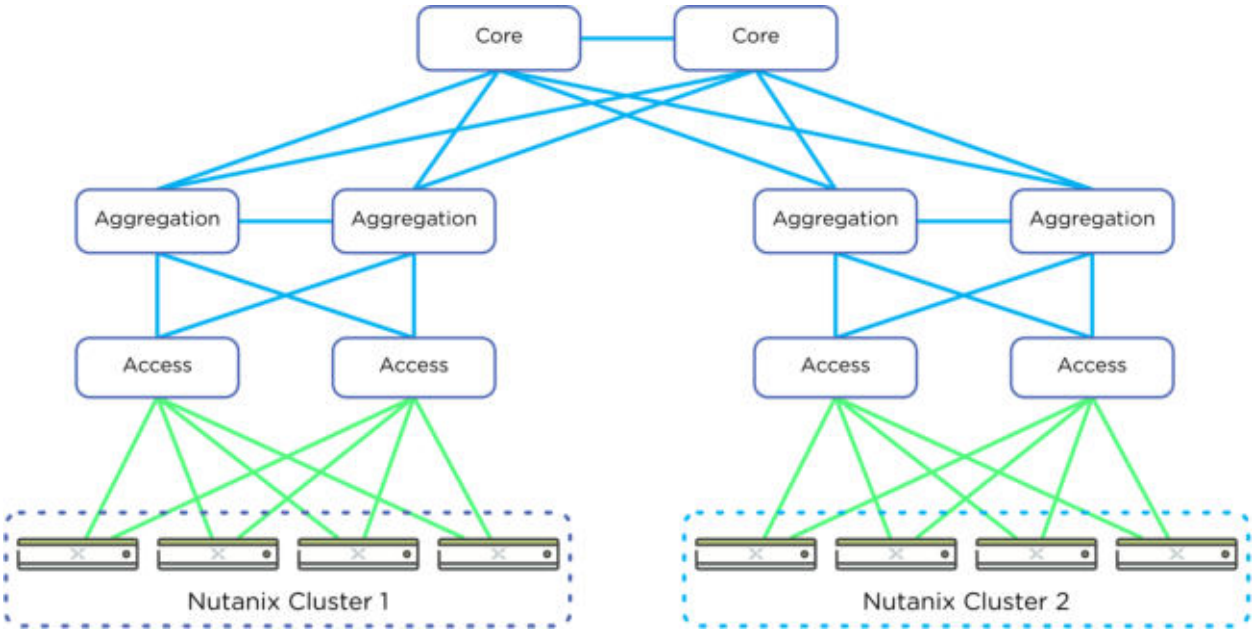


Figure 6: Scaling the Core-Aggregation-Access Network

There are many ways to connect switches in the three-tier design, so your deployment may look slightly different than the one shown above.

Multisite Designs

When two or more physical sites or physical availability zones exist, a single Nutanix cluster shouldn't span them. Instead, create multiple Nutanix clusters (one per availability zone) and connect them with tools such as asynchronous disaster recovery, NearSync, and Metro Availability. This design provides high availability for data and applications and eliminates the possibility of a split-brain scenario, in which a Nutanix cluster is partitioned when the two sites lose connectivity.

© 2021 Nutanix, Inc. All rights reserved

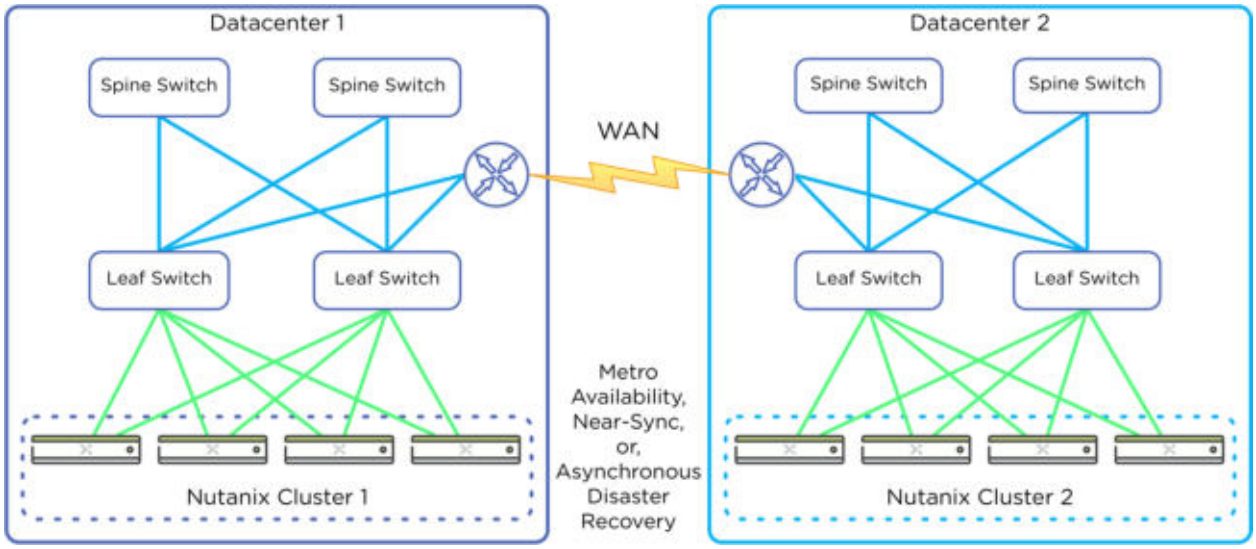


Figure 7: Multisite Network Design Using Replication

7. Conclusion

We often talk to customers concerned about the network for their datacenters in general as well as for their individual Nutanix clusters. Although networking may seem daunting, it really only requires attention to three key points:

1. Procure the right switches.
2. Select the right network design.
3. Use the right replication technologies.

If you connect your Nutanix cluster to a well-designed datacenter network, your storage platform can be highly available while still providing you with excellent performance.

For feedback or questions, please contact us using the Nutanix NEXT Community forums.

© 2021 Nutanix, Inc. All rights reserved

Appendix

Best Practices Checklist

A single Nutanix cluster must meet the following requirements:

- Maximum of three switch hops between any two Nutanix nodes in the same cluster.
- Connect all Nutanix nodes in the same cluster to the same switch fabric (leaf-spine network) or aggregation layer.
- No WAN or remote links between Nutanix nodes in the same Nutanix cluster.
- Separate Nutanix CVM and hypervisor hosts into a dedicated VLAN that doesn't include any VM traffic.
- Don't place Nutanix nodes in the same Nutanix cluster if the stretched layer 2 (L2) network spans multiple datacenters or availability zones or if there is a remote link between the two locations.
  - Using a stretched L2 network over layer 3 (L3) is only acceptable when the Nutanix cluster remains in the same switch fabric or aggregation layer, such as a L2 network stretched between two racks in the same datacenter.
- Don't use features like block or rack awareness to stretch a Nutanix cluster between different physical sites.
- Configure adequate uplinks between switches or interswitch links for east-west storage traffic to minimize port-to-port oversubscription. For example, use multiple 40 Gbps uplinks (or interswitch links).
- Connect hosts using redundant links.
- Configure switch ports facing Nutanix servers as spanning tree portfast or edge to skip the listening and learning phases and prevent cluster outages due to spanning tree topology changes.

© 2021 Nutanix, Inc. All rights reserved

These requirements and recommendations keep latency between nodes minimal and predictable. Networks that have too many switch hops or introduce WAN links between only some of the nodes introduce variable latency, which has a negative impact on storage latency, and variable throughput. Overloaded networks also add components (datacenters, zones, switches, and switch fabrics) that can all impact availability.

References

1. Data Protection and Disaster Recovery
2. Data Protection for AHV-Based VMs
3. Metro Availability
4. AHV Networking
5. VMware vSphere Networking
6. VMware NSX for vSphere
7. Cisco ACI™ with Nutanix
8. Mellanox Networking with Nutanix
9. Networking: Arista EOS Recommended Practices
10. Cisco Nexus Recommended Practices

About the Authors

Jason Burns is a Principal Technical Marketing Engineer at Nutanix. Follow Jason on Twitter @bbbburns.

Magnus Andersson is a Principal Solutions Architect at Nutanix. Follow Magnus on Twitter @magander_se.

About Nutanix

Nutanix makes infrastructure invisible, elevating IT to focus on the applications and services that power their business. The Nutanix enterprise cloud software leverages web-scale engineering and consumer-grade design to natively converge compute, virtualization, and storage into a resilient, software-defined solution with rich machine intelligence. The result is predictable

performance, cloud-like infrastructure consumption, robust security, and seamless application mobility for a broad range of enterprise applications. Learn more at www.nutanix.com or follow us on Twitter @nutanix.

© 2021 Nutanix, Inc. All rights reserved

List of Figures

Figure 1: Nutanix Enterprise Cloud OS Stack..... 9

Figure 2: Host Network Connections.....17

Figure 3: Leaf-Spine Network.....19

Figure 4: Scaling the Leaf-Spine Network..... 20

Figure 5: Core-Aggregation-Access Network..... 21

Figure 6: Scaling the Core-Aggregation-Access Network..... 22

Figure 7: Multisite Network Design Using Replication.....23

List of Tables

Table 1: Document Version History 7

AOS 6.1

Command Reference

February 24, 2022

Contents

1. Acropolis Command-Line Interface (aCLI) 4
- CLI Reference Conventions 4
- ads 4
- core 5
- ha 6
- host 7
- image 9
- iscsi_client 12
- net 13
- nf 26
- parcel 27
- snapshot 27
- vg 28
- vm 36
- vm_group 61

2. Nutanix Command-Line Interface (nCLI) 64
- Installing the nCLI on Your Local System 64
- Initiating an nCLI Session From Your Local System 66
- Command Format 67
- Embedded Help 67
- nCLI Entities 67
- Nutanix Command-Line Interface Reference 68
  - CLI Reference Conventions 68
  - alerts: Alert 68
  - authconfig: Authentication Configuration 70
  - cluster: Cluster 77
  - container: Storage Container 88
  - data-at-rest-encryption: Data At Rest Encryption 88
  - data-at-rest-encryption-certificate: Data At Rest Encryption Certificate 89
  - disk: Physical Disk 90
  - events: Event 91
  - host: Physical Host 92
  - http-proxy: HTTP Proxy 92
  - license: License 94
  - managementserver: Management Server 95
  - multicluster: Multicluster 96
  - network: Network 97
  - pulse-config: Pulse Configuration 97
  - rsyslog-config: TLS configuration for RSyslog service 98
  - snmp: SNMP 101
  - software: Software 105
  - ssl-certificate: SSL Certificate 107
  - storagepool: Storage Pool 108
  - tag: Tag 108
  - user: User 110
  - virtualmachine: Virtual Machine 113

3. Controller VM Commands..... 116
Specifying Credentials..... 116
cluster..... 116
genesis..... 202
ncc..... 208
setup_hyperv.py..... 245

Copyright..... 248
License..... 248
Conventions..... 248
Version..... 248

1

ACROPOLIS COMMAND-LINE INTERFACE (ACLI)

Acropolis provides a command-line interface for managing hosts, networks, snapshots, and VMs.

Accessing the Acropolis CLI

To access the Acropolis CLI, log on to a Controller VM in the cluster with SSH and type `acli` at the shell prompt.

To exit the Acropolis CLI and return to the shell, type `exit` at the `<acropolis>` prompt.

CLI Reference Conventions

This command-line interface reference uses the following conventions.

- Parameters in italic are unique to your environment.

  ```
  value
  ```

- Parameters in square brackets are optional.

  ```
  [ value ]
  ```

- Parameters in curly brackets must be one of a limited set of values.

  ```
  { value1 | value2 }
  ```

  One example is boolean parameters: `{ true | false }`

- The `keyword` is a literal string required by the command, and the *`value`* is the unique value for your environment.

  ```
  keyword=value
  ```

ads

Operations

- Get current Acropolis Dynamic Scheduling (ADS) configuration : `ads.get`
- Enable or disable Acropolis Dynamic Scheduling (ADS) : `ads.update`

Get current Acropolis Dynamic Scheduling (ADS) configuration

```
<acropolis> ads.get
```

Required arguments

None

Enable or disable Acropolis Dynamic Scheduling (ADS)

If this behavior is not desirable, then the only option is to disable this feature completely.

`<acropolis> ads.update [ enable="{ true | false }" ]`

Required arguments

None

Optional arguments

`enable`

Enable ADS.

Type: boolean

Default: true

core

Operations

- Exits the CLI : `core.exit`
- Gets the current value of the given configuration options : `core.get`
- Provides help text for the named object : `core.help`
- Sets the value of the given configuration options : `core.set`

Exits the CLI

`<acropolis> core.exit`

Required arguments

None

Gets the current value of the given configuration options

`<acropolis> core.get [ options ]`

Required arguments

None

Optional arguments

`options`

Comma-delimited list

Type: list of configuration options

Provides help text for the named object

`<acropolis> core.help [ name ]`

Required arguments

None

Optional arguments

`name`

Command or namespace to describe

Type: command or namespace name

Sets the value of the given configuration options

`<acropolis> core.set [ assume_default="{ true | false }" ][ inline_help="{ true | false }" ][ json="{ true | false }" ][ page_size="page_size" ][ pretty="{ true | false }" ][ use_dns="{ true | false }" ]`

Required arguments

None

Optional arguments

`assume_default`

Reply to all prompts with default values.

Type: boolean

`inline_help`

Show inline help for tab completion.

Type: boolean

`json`

Output JSON.

Type: boolean

`page_size`

Number of lines before pagination.

Type: int

`pretty`

Show pretty output.

Type: boolean

`use_dns`

Perform DNS lookups.

Type: boolean

ha

Operations

- Get current HA configuration : `ha.get`
- Enable, disable or modify VM availability configuration : `ha.update`

Get current HA configuration

`<acropolis> ha.get`

Required arguments

None

Enable, disable or modify VM availability configuration

All VMs are evacuated from the host that is picked as reserved.

`<acropolis> ha.update [ num_host_failures_to_tolerate="num_host_failures_to_tolerate" ][ wait="{ true | false }" ]`

Required arguments

None

Optional arguments

`num_host_failures_to_tolerate`

Number of host failures to tolerate. Setting to 0 will restart VMs on a best-effort basis.

Type: int

`wait`

If True, wait for the host evacuation attempt to finish

Type: boolean

Default: true

host

Operations

- Puts a host into maintenance mode : `host.enter_maintenance_mode`
- Check if AHV host can enter maintenance mode : `host.enter_maintenance_mode_check`
- Takes a host out of maintenance mode : `host.exit_maintenance_mode`
- Retrieves scheduler information about a Host : `host.get`
- Lists hosts in the cluster : `host.list`
- Lists VMs currently running on the host : `host.list_vms`

Puts a host into maintenance mode

The user should use the host.get command to determine the host's current maintenance mode state.

`<acropolis> host.enter_maintenance_mode` *host* `[ mode="`*mode*`" ][ non_migratable_vm_action="`*non_migratable_vm_action*`" ][ wait="`*{ true | false }*`" ]`

Required arguments

`host`

Host identifier

Type: host

Optional arguments

`mode`

Evacuation mode ('live', 'cold', 'power_off')

Type: string

Default: live

`non_migratable_vm_action`

Action for non migratable VMs ('block', 'acpi_shutdown')

Type: string

Default: block

`wait`

If True, wait for the host evacuation attempt to finish

Type: boolean

Default: true

Check if AHV host can enter maintenance mode

User can proceed to 'host.enter_maintenance_mode' if check succeeds.

`<acropolis> host.enter_maintenance_mode_check` *host* `[ non_migratable_vm_action="`*non_migratable_vm_action*`" ]`

Required arguments

`host`

Host identifier

Type: host

Optional arguments

`non_migratable_vm_action`

Action for non migratable VMs ('block', 'acpi_shutdown')

Type: string

Default: block

Takes a host out of maintenance mode

If the host is no longer in maintenance mode, this command has no effect.

`<acropolis> host.exit_maintenance_mode` *host*

Required arguments

`host`

Host identifier

Type: host

Retrieves scheduler information about a Host

`<acropolis> host.get` *host_list*

Required arguments

`host_list`

Host identifier

Type: list of hosts

Lists hosts in the cluster

`<acropolis> host.list`

Required arguments

None

Lists VMs currently running on the host

`<acropolis> host.list_vms` *host*

Required arguments

`host`

Host UUID

Type: host

image

Operations

- Create an image : `image.create`
- Delete an image(s) : `image.delete`
- Retrieves information about an image : `image.get`
- List all Images : `image.list`
- Update an image : `image.update`

Create an image

Optionally, a checksum may also be specified if we are creating an image from a source_url in order to verify the correctness of the image.

`<acropolis> image.create` *name* `[ annotation="`*annotation*`" ][ architecture="`*architecture*`" ][ clone_from_vmdisk="`*clone_from_vmdisk*`" ][ compute_checksum="`*{ true | false }*`" ][ container="`*container*`" ][ image_type="`*{raw|vhd|vmdk|vdi|iso|qcow2|vhdx}*`" ][ product_name="`*product_name*`" ][ product_version="`*product_version*`" ][ sha1_checksum="`*sha1_checksum*`" ][ sha256_checksum="`*sha256_checksum*`" ][ source_url="`*source_url*`" ][ wait="`*{ true | false }*`" ]`

Required arguments

name

Comma-delimited list of image names

Type: list of strings with expansion wildcards

Optional arguments

annotation

Image description

Type: string

architecture

Disk image CPU architecture

Type: image architecture

clone_from_vmdisk

UUID of the source vmdisk

Type: VM disk

compute_checksum

If True, we will compute the checksum of the image

Type: boolean

Default: false

container

Destination Storage Container

Type: container

`image_type`

Image type

Type: image type

`product_name`

Name of the producer/distributor of the image

Type: string

`product_version`

Version string for the image

Type: string

`sha1_checksum`

SHA-1 checksum

Type: hex checksum

`sha256_checksum`

SHA-256 checksum

Type: hex checksum

`source_url`

URL location of the source image

Type: image URL

`wait`

If True, we will wait for the image creation to complete

Type: boolean

Default: true

Examples

1. Create an image named 'foo' from an image located at http://test.com/disk_image.

```
<acropolis> image.create foo source_url=http://test.com/image_iso container=default
 image_type=kIsoImage architecture=kX86_64
```

2. Create an image named 'bar' from a vmdisk 0b4fc60b-cc56-41c6-911e-67cc8406d096.

```
<acropolis> image.create bar clone_from_vmdisk=0b4fc60b-cc56-41c6-911e-67cc8406d096
 image_type=kDiskImage
```

Delete an image(s)

`<acropolis> image.delete` *image_list*

Required arguments

`image_list`

Image identifiers

Type: list of images

Retrieves information about an image

`<acropolis> image.get` *image_list* `[ include_vmdisk_paths="`*{ true | false }*`" ][ include_vmdisk_sizes="`*{ true | false }*`" ]`

Required arguments

image_list

Image identifiers

Type: list of images

Optional arguments

include_vmdisk_paths

Fetch vmdisk paths

Type: boolean

Default: false

include_vmdisk_sizes

Fetch vmdisk sizes (in bytes)

Type: boolean

Default: true

List all Images

`<acropolis> image.list`

Required arguments

None

Update an image

`<acropolis> image.update [ annotation="`*annotation*`" ][ architecture="`*architecture*`" ][ image="`*image*`" ][ image_type="`*{raw|vhd|vmdk|vdi|iso|qcow2|vhdx}*`" ][ name="`*name*`" ][ product_name="`*product_name*`" ][ product_version="`*product_version*`" ]`

Required arguments

None

Optional arguments

annotation

Image description

Type: string

architecture

Disk image CPU architecture

Type: image architecture

image

Image identifier

Type: image

image_type

Image type

Type: image type

name

Image name

Type: string

`product_name`

Name of the producer/distributor of the image

Type: string

`product_version`

Version string for the image

Type: string

Examples

1. Update the name of an image named 'foo'.

```
<acropolis> image.update foo name=bar
```

iscsi_client

Operations

- Retrieves information about an ISCSI client : `iscsi_client.get`
- Lists all ISCSI clients : `iscsi_client.list`
- Update an ISCSI client's parameters : `iscsi_client.update`

Retrieves information about an ISCSI client

`<acropolis> iscsi_client.get [ client_uuid_list="`*client_uuid_list*`" ][ name_list="`*name_list*`" ][ network_id_list="`*network_id_list*`" ]`

Required arguments

None

Optional arguments

`client_uuid_list`

Comma-delimited list of iscsi initiator client uuids

Type: List of iSCSI clients

`name_list`

Comma-delimited list of iscsi initiator names (IQNs)

Type: List of iSCSI clients

`network_id_list`

Comma-delimited list of iscsi initiator network identifiers

Type: List of iSCSI clients

Lists all ISCSI clients

`<acropolis> iscsi_client.list`

Required arguments

None

Update an ISCSI client's parameters

`<acropolis> iscsi_client.update [ client_secret="`*client_secret*`" ][ client_uuid="`*client_uuid*`" ][ name="`*name*`" ][ network_id="`*network_id*`" ]`

Required arguments

None

Optional arguments

`client_secret`

Client secret to be used in CHAP authentication

Type: string

`client_uuid`

iscsi client UUID

Type: iSCSI client name or network id

`name`

iscsi initiator name (IQNs)

Type: iSCSI client name or network id

`network_id`

iscsi initiator network identifier

Type: iSCSI client name or network id

net

Operations

- Add a DHCP pool to a managed network : `net.add_dhcp_pool`
- Blacklists IP addresses for a managed network : `net.add_to_ip_blacklist`
- Clear the DHCP DNS configuration for a managed network : `net.clear_dhcp_dns`
- Clear the DHCP TFTP configuration for a managed network : `net.clear_dhcp_tftp`
- Clears the network function chain for this network : `net.clear_network_function_chain`
- Creates a new virtual network for VMs : `net.create`
- Creates a SPAN session : `net.create_span_session`
- Creates a Virtual Switch : `net.create_virtual_switch`
- Deletes a network : `net.delete`
- Delete a DHCP pool from a managed network : `net.delete_dhcp_pool`
- Removes IP addresses from a managed network's blacklist : `net.delete_from_ip_blacklist`
- Deletes a SPAN session : `net.delete_span_session`
- Deletes a virtual switch : `net.delete_virtual_switch`
- Disable virtual switch (delete all virtual switches) : `net.disable_virtual_switch`
- Retrieves information about a network : `net.get`
- Retrieves information about a SPAN session configuration and statistics : `net.get_span_session`
- Retrieves information about a specific virtual switch configuration : `net.get_virtual_switch`

- Lists all networks : `net.list`
- List blacklisted IPs for a managed network : `net.list_ip_blacklist`
- Lists all SPAN session configurations : `net.list_span_session`
- Lists all virtual switch configurations : `net.list_virtual_switch`
- Lists VMs configured on the network : `net.list_vms`
- Migrates an OVS bridge to Virtual Switch : `net.migrate_br_to_virtual_switch`
- Updates network metadata : `net.update`
- Configure the DHCP DNS configuration for a managed network : `net.update_dhcp_dns`
- Configure the DHCP TFTP configuration for a managed network : `net.update_dhcp_tftp`
- Sets the network function chain for this network : `net.update_network_function_chain`
- Updates a SPAN session : `net.update_span_session`
- Updates a Virtual Switch configuration : `net.update_virtual_switch`

Add a DHCP pool to a managed network

That is, a user may manually specify an address belonging to the pool when creating a virtual adapter.

`<acropolis> net.add_dhcp_pool` *network* `[ end="`*end*`" ][ start="`*start*`" ]`

Required arguments

network

Network identifier

Type: network

Optional arguments

end

Last IPv4 address

Type: IPv4 address

start

First IPv4 address

Type: IPv4 address

Examples

1. Auto-assign addresses from the inclusive range 192.168.1.16 - 192.168.1.32.

```
<acropolis> net.add_dhcp_pool vlan.16 start=192.168.1.16 end=192.168.1.32
```

Blacklists IP addresses for a managed network

This property may be useful for avoiding conflicts between VMs and other hosts on the physical network.

`<acropolis> net.add_to_ip_blacklist` *network* `[ cookie="`*cookie*`" ][ ip_list="`*ip_list*`" ][ num_ip_addresses="`*num_ip_addresses*`" ]`

Required arguments

network

Network identifier

Type: network

Optional arguments

`cookie`

Opaque data

Type: string

`ip_list`

Comma-delimited list of IP addresses

Type: list of IPv4 addresses

`num_ip_addresses`

Number of IP addresses to reserve

Type: int

Clear the DHCP DNS configuration for a managed network

`<acropolis> net.clear_dhcp_dns` *network*

Required arguments

`network`

Network identifier

Type: network

Examples

1. Clear DNS servers and search domains.

```
<acropolis> net.clear_dhcp_dns vlan.123
```

Clear the DHCP TFTP configuration for a managed network

`<acropolis> net.clear_dhcp_tftp` *network*

Required arguments

`network`

Network identifier

Type: network

Examples

1. Clear TFTP server name and boot filename.

```
<acropolis> net.clear_dhcp_tftp vlan.123
```

Clears the network function chain for this network

`<acropolis> net.clear_network_function_chain` *network*

Required arguments

`network`

Network identifier

Type: network

Creates a new virtual network for VMs

For more about managed networks, see the following commands: net.add_dhcp_pool net.add_to_ip_blacklist net.clear_dhcp_dns net.clear_dhcp_tftp net.delete_dhcp_pool net.delete_from_ip_blacklist net.list_ip_blacklist net.update_dhcp_dns net.update_dhcp_tftp

`<acropolis> net.create` *name* `[ annotation="`*annotation*`" ][ dhcp_address="`*dhcp_address*`" ][ ip_config="`*ip_config*`" ][ mtu="`*mtu*`" ][ virtual_switch="`*virtual_switch*`" ][ vlan="`*vlan*`" ][ vswitch_name="`*vswitch_name*`" ]`

Required arguments

name

Network name

Type: string

Optional arguments

annotation

Annotation string

Type: string

dhcp_address

DHCP server address (for managed networks)

Type: IPv4 address

ip_config

IP configuration in CIDR notation ("default_gateway/prefix")

Type: string

mtu

MTU setting

Type: int

virtual_switch

Virtual Switch name

Type: string

vlan

VLAN ID

Type: int

vswitch_name

OVS bridge name. This option is deprecated, please use virtual_switch option

Type: string

Examples

1. Create an unmanaged network on VLAN 66.

```
<acropolis> net.create mynet vlan=66
```

2. Create an unmanaged network on VLAN 66 with MTU 9000.

```
<acropolis> net.create mynet vlan=66 mtu=9000
```

3. Create a managed network on VLAN 99, bound to virtual switch br1.IPv4 range is 10.1.1.0 - 10.1.1.255, the default gateway is 10.1.1.1, and the DHCP server is 10.1.1.254

```
<acropolis> net.create mynet vlan=99 virtual_switch=vs1 ip_config=10.1.1.1/24
```

4. Create an untagged managed network. The managed IPv4 range is 192.168.0.0 - 192.168.3.255, and the default gateway is 192.168.5.254. In this example, the DHCP server will be automatically configured as 192.168.5.253 to avoid collision with the default gateway.

```
<acropolis> net.create mynet vlan=0 ip_config=192.168.5.254/22
```

5. Create a cloud network.

```
<acropolis> net.create mynet ip_config=10.192.20.0/24
```

Creates a SPAN session

`<acropolis> net.create_span_session` *name* `[ description="`*description*`" ][ dest_list="`*dest_list*`" ][ source_list="`*source_list*`" ]`

Required arguments

name

SPAN session name

Type: string

Optional arguments

description

Description of SPAN session

Type: string

dest_list

List of destination ports, in the format of [port1_config;port2_config], each port configuration in the format of {uuid=host_uuid/vm_uuid,type=kVmNic,identifier=host_nic_name/vnic_mac_addr}

Type: List of SPAN source/dest ports

source_list

List of source ports, in the format of [port1_config;port2_config], each port configuration in the format of {uuid=host_uuid/vm_uuid,type=kHostNic,identifier=host_nic_name/vnic_mac_addr,direction=kIngress/kEgress/kBiDir}

Type: List of SPAN source/dest ports

Examples

1. Create a SPAN session with one host NIC as source and one vNIC as destination

```
<acropolis> net.create_span_session span-example
 source_list=[{uuid=host_uuid,type=kHostNic,identifier=eth1,direction=kBiDir}]
 dest_list=[{uuid=vm_uuid,type=kVmNic,identifier=vnic_mac_addr}]
```

Creates a Virtual Switch

`<acropolis> net.create_virtual_switch` *vs_name* `[ bond_type="`*bond_type*`" ][ cluster="`*cluster*`" ][ description="`*description*`" ][ enable_igmp_querier="`*{ true | false }*`" ][ enable_igmp_snooping="`*{ true | false }*`" ][ host_uplink_config="`*host_uplink_config*`" ][ igmp_snooping_timeout="`*igmp_snooping_timeout*`" ][ lacp_fallback="`*{ true | false }*`" ][ lacp_timeout="`*lacp_timeout*`" ][ mtu="`*mtu*`" ][ quick_mode="`*{ true | false }*`" ]`

Required arguments

vs_name

Virtual switch name

Type: string

Optional arguments

`bond_type`

Type of uplink bond

Type: NIC Team policy

`cluster`

UUID of cluster (optional)

Type: string

`description`

Description of virtual switch

Type: string

`enable_igmp_querier`

Whether to enable IGMP querier on this virtual switch, default is false

Type: boolean

`enable_igmp_snooping`

Whether to enable IGMP snooping on this virtual switch, default is false

Type: boolean

`host_uplink_config`

List of host uplinks, in the format of {host1-uuid:[host1_uplink_port_list];host2-uuid:[host2_uplink_port_list]}

Type: List of host uplinks

`igmp_snooping_timeout`

Timeout value in seconds for entries in OVS multicast DB, default is 300, and must be a value in the range between 15 and 3600

Type: int

`lacp_fallback`

Whether to enable LACP fallback to active-backup on LACP negotiation failure

Type: boolean

`lacp_timeout`

LACP timeout value

Type: LACP timeout value

`mtu`

MTU of uplink interfaces, default is 1500

Type: int

`quick_mode`

With quick mode, hosts are not put into maintenance mode during OVS configuration changes

Type: boolean

Examples

1. Create virtual switch configuration on a cluster with 3 nodes, set uplink bond mode to active-backup, mtu to 1600, hosts are not put into maintenance mode during OVS configuration

```
<acropolis> net.create_virtual_switch vs-example mtu=1600 bond_type=kActiveBackup
 host_uplink_config={host1-uuid:[eth0,eth1];host2-uuid:[eth2,eth3];host3-uuid:
[eth0,eth1,eth2,eth3]} quick_mode=true
```

2. Create virtual switch configuration on a cluster with 3 nodes, set uplink bond mode to use LACP, and tune LACP parameters

```
<acropolis> net.create_virtual_switch vs-example bond_type=kBalanceTcp lacp_fallback=true
 lacp_timeout=kSlow host_uplink_config={host1-uuid:[eth0,eth1];host2-uuid:[eth2,eth3];host3-
uuid:[eth0,eth1,eth2,eth3]}
```

3. Create virtual switch configuration on a cluster with 3 nodes, use single uplink on node 1, and no uplink on remained nodes, explicitly specify the cluster UUID

```
<acropolis> net.create_virtual_switch vs-example mtu=1500 bond_type=kNoBond
 cluster=cluster_uuid host_uplink_config={host1-uuid:[eth0];host2-uuid:[];host3-uuid:[]}
```

4. Create virtual switch configuration on a cluster with 3 nodes, no uplink used on any node, enable IGMP snooping and querier, set snooping timeout to 500 second

```
<acropolis> net.create_virtual_switch vs-example bond_type=kNoBond host_uplink_config={host1-
uuid:[];host2-uuid:[];host3-uuid:[]} enable_igmp_snooping=true enable_igmp_querier=true
 igmp_snooping_timeout=500
```

Deletes a network

To determine which VMs are on a network, use net.list_vms.

`<acropolis> net.delete` *network*

Required arguments

network

Network identifier

Type: network

Delete a DHCP pool from a managed network

See network.add_dhcp_pool for more information.

`<acropolis> net.delete_dhcp_pool` *network* `[ start="`*start*`" ]`

Required arguments

network

Network identifier

Type: network

Optional arguments

start

First IPv4 address

Type: IPv4 address

Removes IP addresses from a managed network's blacklist

`<acropolis> net.delete_from_ip_blacklist` *network* `[ cookie="`*cookie*`" ][ ip_list="`*ip_list*`" ]`

Required arguments

network

Network identifier

Type: network

Optional arguments

cookie

Opaque data

Type: string

ip_list

Comma-delimited list of IP addresses

Type: list of IPv4 addresses

Deletes a SPAN session

`<acropolis> net.delete_span_session` *span_session*

Required arguments

span_session

SPAN session identifier

Type: SPAN Session Name

Deletes a virtual switch

`<acropolis> net.delete_virtual_switch` *vs_name*

Required arguments

vs_name

Virtual switch name

Type: string

Disable virtual switch (delete all virtual switches)

`<acropolis> net.disable_virtual_switch`

Required arguments

None

Retrieves information about a network

`<acropolis> net.get` *network_list*

Required arguments

network_list

Network identifier

Type: list of networks

Retrieves information about a SPAN session configuration and statistics

`<acropolis> net.get_span_session` *span_session_name_list*

Required arguments

span_session_name_list

List of SPAN session names.

Type: list of SPAN Session Names

Retrieves information about a specific virtual switch configuration

`<acropolis> net.get_virtual_switch` *virtual_switch_name_list*

Required arguments

virtual_switch_name_list

List of virtual switch names.

Type: list of Virtual Switch Names

Lists all networks

`<acropolis> net.list`

Required arguments

None

List blacklisted IPs for a managed network

`<acropolis> net.list_ip_blacklist` *network* `[ cookie="`*cookie*`" ]`

Required arguments

network

Network identifier

Type: network

Optional arguments

cookie

Opaque data

Type: string

Lists all SPAN session configurations

`<acropolis> net.list_span_session`

Required arguments

None

Lists all virtual switch configurations

`<acropolis> net.list_virtual_switch`

Required arguments

None

Lists VMs configured on the network

`<acropolis> net.list_vms` *network*

Required arguments

network

Network identifier

Type: network

Migrates an OVS bridge to Virtual Switch

`<acropolis> net.migrate_br_to_virtual_switch` *bridge* `[ cluster="`*cluster*`" ][ description="`*description*`" ][ vs_name="`*vs_name*`" ]`

Required arguments

bridge

OVS bridge name

Type: string

Optional arguments

cluster

UUID of cluster (optional)

Type: string

description

Description of virtual switch

Type: string

vs_name

Name of virtual switch

Type: string

Examples

1. Migrate br1 to virtual switch vs1. To make this work, br1 needs to exist on every node of this cluster, except storage-only nodes, and its configuration on every node needs to be consistent.

```
<acropolis> net.migrate_br_to_virtual_switch br1 vs_name=vs1
```

Updates network metadata

`<acropolis> net.update` *network* `[ annotation="`*annotation*`" ][ name="`*name*`" ][ virtual_switch="`*virtual_switch*`" ]`

Required arguments

network

Network identifier

Type: network

Optional arguments

annotation

Annotation string

Type: string

name

Network name

Type: string

virtual_switch

Virtual switch name

Type: string

Configure the DHCP DNS configuration for a managed network

However, the DHCP server hands out infinite leases, so clients will need to manually renew to pick up the new settings.

`<acropolis> net.update_dhcp_dns` *network* `[ domains="`*domains*`" ][ servers="`*servers*`" ]`

Required arguments

network

Network identifier

Type: network

Optional arguments

domains

Comma-delimited list of search domains

Type: list of DNS domains

servers

Comma-delimited list of DNS server IP addresses

Type: list of IPv4 addresses

Examples

1. Configure DNS servers and search domains.

```
<acropolis> net.update_dhcp_dns vlan.123 servers=10.1.1.1,10.1.1.2
 domains=eng.nutanix.com,corp.nutanix.com
```

Configure the DHCP TFTP configuration for a managed network

However, the TFTP server hands out infinite leases, so clients will need to manually renew to pick up the new settings.

`<acropolis> net.update_dhcp_tftp` *network* `[ bootfile_name="`*bootfile_name*`" ][ server_name="`*server_name*`" ]`

Required arguments

network

Network identifier

Type: network

Optional arguments

bootfile_name

Boot file name

Type: string

server_name

TFTP server name

Type: string

Examples

1. Configure TFTP server and bootfile.

```
<acropolis> net.update_dhcp_tftp vlan.123 server_name=10.1.1.1 bootfile_name=ARDBP32.BIN
```

Sets the network function chain for this network

`<acropolis> net.update_network_function_chain` *network chain*

Required arguments

network

Network identifier

Type: network

chain

Network function chain identifier

Type: network function chain

Updates a SPAN session

`<acropolis> net.update_span_session` *span_session* `[ description="`*description*`" ][ dest_list="`*dest_list*`" ][ name="`*name*`" ][ source_list="`*source_list*`" ]`

Required arguments

span_session

SPAN session identifier

Type: SPAN Session Name

Optional arguments

description

Description of SPAN session

Type: string

dest_list

List of destination ports, in the format of [port1_config;port2_config], each port configuration in the format of {uuid=host_uuid/vm_uuid,type=kVmNic,identifier=host_nic_name/vnic_mac_addr}

Type: List of SPAN source/dest ports

name

SPAN session name

Type: string

source_list

List of source ports, in the format of [port1_config;port2_config], each port configuration in the format of {uuid=host_uuid/vm_uuid,type=kHostNic,identifier=host_nic_name/vnic_mac_addr,direction=kIngress/kEgress/kBiDir}

Type: List of SPAN source/dest ports

Examples

1. Update SPAN session name

```
<acropolis> net.update_span_session span-example name=span-demo
```

2. Update a SPAN session with one host NIC as source and one vNIC as destination

```
<acropolis> net.update_span_session span-example
 source_list=[{uuid=host_uuid,type=kHostNic,identifier=eth1,direction=kBiDir}]
 dest_list=[{uuid=vm_uuid,type=kVmNic,identifier=vnic_mac_addr}]
```

Updates a Virtual Switch configuration

`<acropolis> net.update_virtual_switch` *virtual_switch* `[ bond_type="`*bond_type*`" ][ description="`*description*`" ][ enable_igmp_querier="`*{ true | false }*`" ][ enable_igmp_snooping="`*{ true | false }*`" ][ host_uplink_config="`*host_uplink_config*`" ][ igmp_query_vlan_list="`*igmp_query_vlan_list*`" ][ igmp_snooping_timeout="`*igmp_snooping_timeout*`" ][ lacp_fallback="`*{ true | false }*`"`

][lacp_timeout="*lacp_timeout*"][mtu="*mtu*"][quick_mode="*{ true | false }*"][vs_name="*vs_name*"]

Required arguments

virtual_switch

Virtual switch identifier

Type: Virtual Switch Name

Optional arguments

bond_type

Type of uplink bond

Type: NIC Team policy

description

Description of virtual switch

Type: string

enable_igmp_querier

Whether to enable IGMP querier on this virtual switch, default is false

Type: boolean

enable_igmp_snooping

Whether to enable IGMP snooping on this virtual switch, default is false

Type: boolean

host_uplink_config

List of host uplinks, in the format of {host1-uuid:[host1_uplink_port_list];host2-uuid:[host2_uplink_port_list]}

Type: List of host uplinks

igmp_query_vlan_list

List of VLAN IDs mapped to the virtual switch for which IGMP querier is enabled. When it's not set or set as an empty list, querier is enabled for all VLANs of the virtual switch.

Type: list of VLAN ID

igmp_snooping_timeout

Timeout value in seconds for entries in OVS multicast DB, default is 300, and must be a value in the range between 15 and 3600

Type: int

lacp_fallback

Whether to enable LACP fallback to active-backup on LACP negotiation failure

Type: boolean

lacp_timeout

LACP timeout value

Type: LACP timeout value

mtu

MTU of uplink interfaces, default is 1500

Type: int

`quick_mode`

With quick mode, hosts are not put into maintenance mode during OVS configuration changes

Type: boolean

`vs_name`

Virtual switch name

Type: string

Examples

1. Update name and uplink mtu of virtual switch vs1

```
<acropolis> net.update_virtual_switch vs1 vs_name=new-name mtu=1600
```

2. Update bond type of virtual switch vs1 from active-backup to balance-tcp, without changing uplink ports

```
<acropolis> net.update_virtual_switch vs1 bond_type=kBalanceTcp
```

3. Update LACP parameters of virtual switch vs1, only valid when bond type is balance-tcp

```
<acropolis> net.update_virtual_switch vs1 lacp_fallback=true lacp_timeout=kSlow
```

4. Update uplink of virtual switch vs1, change uplink on host1

```
<acropolis> net.update_virtual_switch vs1 host_uplink_config={host1-uuid:[eth2,eth3]}
```

5. Update uplink of virtual switch vs1, remove all uplinks, and hosts are not put into maintenance mode during this change

```
<acropolis> net.update_virtual_switch vs1 bond_type=kNoBond host_uplink_config={host1-uuid:
[];host2-uuid:[];host3-uuid:[]} quick_mode=true
```

6. Enable IGMP snooping and querier on virtual switch vs1, set snooping timeout to 500 second, send querier in VLAN 100 & 200

```
<acropolis> net.update_virtual_switch vs1 enable_igmp_snooping=true enable_igmp_querier=true
 igmp_snooping_timeout=500 igmp_query_vlan_list=[100,200]
```

nf

Operations

- Gets specific network function chains : `nf.chain_get`
- Lists all network function chains : `nf.chain_list`

Gets specific network function chains

`<acropolis> nf.chain_get` *chains*

Required arguments

`chains`

Network function chain

Type: list of network function chains

Lists all network function chains

`<acropolis> nf.chain_list`

Required arguments

None

parcel

Operations

- Retrieves information about a parcel : `parcel.get`
- List all Parcels : `parcel.list`

Retrieves information about a parcel

`<acropolis> parcel.get` *parcel_list*

Required arguments

parcel_list

Parcel identifiers

Type: list of parcels

List all Parcels

`<acropolis> parcel.list`

Required arguments

None

snapshot

Operations

- Deletes one or more snapshots : `snapshot.delete`
- Retrieves information about a snapshot : `snapshot.get`
- Lists all snapshots : `snapshot.list`

Deletes one or more snapshots

`<acropolis> snapshot.delete` *snapshot_list*

Required arguments

snapshot_list

Comma-delimited list of snapshot identifiers

Type: list of snapshots

Retrieves information about a snapshot

`<acropolis> snapshot.get` *snapshot_list*

Required arguments

snapshot_list

Snapshot identifier

Type: list of snapshots

Lists all snapshots

```
<acropolis> snapshot.list
```

Required arguments

None

vg

Operations

- Allow volume group to be accessed from an iSCSI intiator : `vg.attach_external`
- Attach a VG to the specified VM : `vg.attach_to_vm`
- Clone Volume Group : `vg.clone`
- Create one or more VGs : `vg.create`
- Delete one or more VGs : `vg.delete`
- Stop allowing volume group to be accessed from an iSCSI intiator : `vg.detach_external`
- Detach a VG from a VM : `vg.detach_from_vm`
- Add a disk to a VG : `vg.disk_create`
- Remove a disk from a VG : `vg.disk_delete`
- Update the properties or replace the volume group disk : `vg.disk_update`
- Retrieve information about a VG : `vg.get`
- List all VGs : `vg.list`
- Update the VGs : `vg.update`
- Update the iSCSI initiator properties for a VG : `vg.update_external`

Allow volume group to be accessed from an iSCSI intiator

```
<acropolis> vg.attach_external vg [ initiator_name="initiator_name" ][
initiator_network_id="initiator_network_id" ]
```

Required arguments

vg

VG identifier

Type: volume group

Optional arguments

`initiator_name`

Name of iSCSI initiator as a valid IQN

Type: iSCSI IQN

`initiator_network_id`

Network identifier of iSCSI initiator as a valid IPv4 address

Type: string

Examples

1. Attach an iSCSI client identified by name, iqn.1994-05.com.redhat:71eef92fe6c, to the VG, vg1,:

```
<acropolis> vg.attach_external vg1.attach_external
 initiator_name=iqn.1994-05.com.redhat:71eef92fe6c
```

2. Attach an iSCSI client identified by network id, 10.1.1.1, to the VG, vg2,:

```
<acropolis> vg.attach_external vg2.attach_external initiator_network_id=10.1.1.1
```

Attach a VG to the specified VM

`<acropolis> vg.attach_to_vm` *vg vm* `[ index="`*index*`" ]`

Required arguments

vg

VG identifier

Type: volume group

vm

VM identifier

Type: VM

Optional arguments

index

Device index on the scsi bus

Type: int

Clone Volume Group

If the ISCSI target names for the clones are not specified through the 'iscsi_target_prefix_list' argument, then default values will be used.

`<acropolis> vg.clone` *name_list* `[ clone_from_vg="`*clone_from_vg*`" ][ created_by="`*created_by*`" ][ iscsi_target_prefix_list="`*iscsi_target_prefix_list*`" ][ load_balance_vm_attachments="`*{ true | false }*`" ][ target_secret_list="`*target_secret_list*`" ]`

Required arguments

name_list

Comma-delimited list of VG names

Type: list of strings with expansion wildcards

Optional arguments

clone_from_vg

VG from which to clone

Type: volume group

created_by

Service/user who created these VGs

Type: string

iscsi_target_prefix_list

Comma-delimited list of iscsi target prefixes for each of the VGs

Type: list of strings

load_balance_vm_attachments

Whether to enable/disable VG load balance VM attachments

Type: boolean

target_secret_list

Comma delimited CHAP secrets associated with each of the VGs. To delete the secret, set target_secret="" using vg.update

Type: list of strings

Examples

1. Clone two VGs vg1 and vg2 with iscsi targets vgt1 and vgt2 from source-vg

```
<acropolis> vg.clone vg1,vg2 clone_from_vg=source-vg iscsi_target_prefix_list=vgt1,vgt2
```

2. Clone two VGs vg1 and vg2 with target_secrets vg1_target_secret and vg2_target_secret

```
<acropolis> vg.clone vg1,vg2 clone_from_vg=source-vg
 target_secret_list=vg1_target_secret,vg2_target_secret
```

3. Clone VG vg2 from vg1 with load_balance_vm_attachments set

```
<acropolis> vg.clone vg.clone vg2 clone_from_vg=vg1 load_balance_vm_attachments=True
```

Create one or more VGs

`<acropolis> vg.create` *name_list* `[ annotation="`*annotation*`" ][ created_by="`*created_by*`" ][ flash_mode="`*{ true | false }*`" ][ is_hidden="`*{ true | false }*`" ][ iscsi_target_prefix_list="`*iscsi_target_prefix_list*`" ][ load_balance_vm_attachments="`*{ true | false }*`" ][ shared="`*{ true | false }*`" ][ target_secret_list="`*target_secret_list*`" ][ usage_type="`*usage_type*`" ]`

Required arguments

name_list

Comma-delimited list of VG names

Type: list of strings with expansion wildcards

Optional arguments

annotation

Annotation string

Type: string

created_by

Service/user who created these VGs

Type: string

flash_mode

Whether to enable/disable flash mode

Type: boolean

is_hidden

Whether the VG is meant to be hidden or not

Type: boolean

iscsi_target_prefix_list

Comma-delimited list of iscsi target prefixes for each of the VGs

Type: list of strings

load_balance_vm_attachments

Whether to enable/disable VG load balance VM attachments

Type: boolean

`shared`

Allow VG to be attached to multiple VMs simultaneously?

Type: boolean

`target_secret_list`

Comma delimited CHAP secrets associated with each of the VGs. To delete the secret, set target_secret="" using vg.update

Type: list of strings

`usage_type`

Expected usage type for the VG

Type: usage type

Delete one or more VGs

`<acropolis> vg.delete` *vg_list*

Required arguments

`vg_list`

Comma-delimited VG identifiers

Type: list of volume groups

Stop allowing volume group to be accessed from an iSCSI intiator

`<acropolis> vg.detach_external` *vg* `[ initiator_name="`*initiator_name*`" ][ initiator_network_id="`*initiator_network_id*`" ]`

Required arguments

`vg`

VG identifier

Type: volume group

Optional arguments

`initiator_name`

Name of iSCSI initiator as a valid IQN

Type: VG external initiator name

`initiator_network_id`

Network identifier of iSCSI initiator as a valid IPv4 address

Type: VG external initiator name

Detach a VG from a VM

`<acropolis> vg.detach_from_vm` *vg* *vm*

Required arguments

`vg`

VG identifier

Type: volume group

`vm`

VM identifier

Type: VM attached to VG

Add a disk to a VG

This can be used to expand a disk image at clone time.

`<acropolis> vg.disk_create` *vg* `[ annotation="`*annotation*`" ][ clone_from_adsf_file="`*clone_from_adsf_file*`" ][ clone_from_vmdisk="`*clone_from_vmdisk*`" ][ clone_min_size="`*clone_min_size*`" ][ container="`*container*`" ][ create_size="`*create_size*`" ][ index="`*index*`" ]`

Required arguments

vg

VG identifier

Type: volume group

Optional arguments

`annotation`

Annotation string

Type: string

`clone_from_adsf_file`

Path to an ADSF file

Type: ADSF path

`clone_from_vmdisk`

A vmdisk UUID

Type: VM disk

`clone_min_size`

Minimum size of the resulting clone (only applies to cloned disks)

Type: size with cskKmMgGtT suffix

`container`

Storage Container (only applies to newly-created disks)

Type: container

`create_size`

Size of new disk

Type: size with cskKmMgGtT suffix

`index`

Device index on bus

Type: int

Examples

1. Create a blank 5GiB disk on ctr, and add it to my_vg at index 3

```
<acropolis> vg.disk_create my_vg create_size=5G container=ctr index=3
```

2. Clone a disk from the ADSF file /ctr/plan9.iso, and add it to a volume group my_vg

```
<acropolis> vg.disk_create my_vg clone_from_adsf_file=/ctr/plan9.iso
```

3. Clone a disk from the existing vmdisk, and add it to the first open slot

```
<acropolis> vg.disk_create my_vg clone_from_vmdisk=0b4fc60b-cc56-41c6-911e-67cc8406d096
```

Remove a disk from a VG

`<acropolis> vg.disk_delete` *vg index*

Required arguments

vg

VG identifier

Type: volume group

index

Disk index

Type: VG disk index

Update the properties or replace the volume group disk

For other options, the existing disk is deleted and replaced by an existing image, file or a vDisk.

`<acropolis> vg.disk_update` *vg index* `[ annotation="`*annotation*`" ][ clone_from_adsf_file="`*clone_from_adsf_file*`" ][ clone_from_image="`*clone_from_image*`" ][ clone_from_vmdisk="`*clone_from_vmdisk*`" ][ clone_min_size="`*clone_min_size*`" ][ container="`*container*`" ][ create_size="`*create_size*`" ][ flash_mode="`*{ true | false }*`" ][ new_annotation="`*new_annotation*`" ][ new_size="`*new_size*`" ]`

Required arguments

vg

VG identifier

Type: volume group

index

Disk index

Type: VG disk index

Optional arguments

`annotation`

Annotation string

Type: string

`clone_from_adsf_file`

Path to an ADSF file

Type: ADSF path

`clone_from_image`

An image name/UUID

Type: image

`clone_from_vmdisk`

A vmdisk UUID

Type: VM disk

`clone_min_size`

Minimum size of the resulting clone(only applies to cloned disks)

Type: size with cskKmMgGtT suffix

`container`

Storage Container (only applies to newly-created disks)

Type: container

`create_size`

Size of new disk

Type: size with cskKmMgGtT suffix

`flash_mode`

Enable/Disable flash mode on this disk

Type: boolean

`new_annotation`

New Annotation string

Type: string

`new_size`

New size for the existing disk

Type: size with cskKmMgGtT suffix

Examples

1. Replace the disk at index 0 with blank 5GiB disk on ctr.

```
<acropolis> vg.disk_update my_vg 0 create_size=5G container=ctr
```

2. Replace the disk at index 0 with blank 5GiB disk on ctr with annotation.

```
<acropolis> vg.disk_update my_vg 0 create_size=5G container=ctr annotation=foo
```

3. Replace the disk at index 0 with a clone of /ctr/plan9.iso.

```
<acropolis> vg.disk_update my_vg 0 clone_from_adsf_file=/ctr/plan9.iso
```

4. Replace the disk at index 0 with a clone of the existing vmdisk.

```
<acropolis> vg.disk_update my_vg 0 clone_from_vmdisk=0b4fc60b-cc56-41c6-911e-67cc8406d096
```

5. Update the size of disk at index 0 to 5GiB.

```
<acropolis> vg.disk_update my_vg 0 new_size=5G
```

6. Disable flash mode for disk at index 0.

```
<acropolis> vg.disk_update my_vg 0 flash_mode=false
```

7. Update annotation string for disk at index 0.

```
<acropolis> vg.disk_update my_vg 0 new_annotation="Updated annotation"
```

Retrieve information about a VG

`<acropolis> vg.get` *vg_list* `[ include_vmdisk_paths="{` *true | false* `}" ][ include_vmdisk_sizes="{` *true | false* `}" ]`

Required arguments

`vg_list`

VG identifier

Type: list of volume groups

Optional arguments

`include_vmdisk_paths`

Fetch disk paths

Type: boolean

Default: false

`include_vmdisk_sizes`

Fetch disk sizes (in bytes)

Type: boolean

Default: true

List all VGs

`<acropolis> vg.list`

Required arguments

None

Update the VGs

`<acropolis> vg.update` *vg_list* `[ annotation="`*annotation*`" ][ cbr_not_capable_reason="`*cbr_not_capable_reason*`" ][ created_by="`*created_by*`" ][ flash_mode="`*{ true | false }*`" ][ iscsi_target_prefix_list="`*iscsi_target_prefix_list*`" ][ load_balance_vm_attachments="`*{ true | false }*`" ][ name="`*name*`" ][ shared="`*{ true | false }*`" ][ target_secret_list="`*target_secret_list*`" ]`

Required arguments

`vg_list`

Comma-delimited list of VG identifiers

Type: list of volume groups

Optional arguments

`annotation`

Annotation string

Type: string

`cbr_not_capable_reason`

If set, marks the VG incapable of CBR workflows

Type: string

`created_by`

Service/user who created these VGs

Type: string

`flash_mode`

Whether to enable/disable flash mode

Type: boolean

`iscsi_target_prefix_list`

Comma-delimited list of iscsi target prefixes for each of the VGs

Type: list of strings

`load_balance_vm_attachments`

Whether to enable/disable VG load balance VM attachments

Type: boolean

`name`

VG name

Type: string

`shared`

Allow VG to be attached to multiple VMs simultaneously?

Type: boolean

`target_secret_list`

Comma delimited CHAP secrets associated with each of the VGs. Enter "" to delete the secret.

Type: list of strings

Examples

1. Update target_secret for VG vg1 and delete target_secret for VG vg2

```
<acropolis> vg.update vg1,vg2 target_secret_list=new_vg1_secret,""
```

Update the iSCSI initiator properties for a VG

`<acropolis> vg.update_external` *vg* `[ initiator_name="`*initiator_name*`" ][ initiator_network_id="`*initiator_network_id*`" ]`

Required arguments

vg

VG identifier

Type: volume group

Optional arguments

`initiator_name`

Name of iSCSI initiator as a valid IQN

Type: VG external initiator name

`initiator_network_id`

Network identifier of iSCSI initiator as a valid IPv4 address

Type: VG external initiator name

vm

Operations

- Retrieves affinity nodes for a given VM : `vm.affinity_nodeget`
- Enable VM-host affinity : `vm.affinity_set`
- Unsets affinity setting of specified VMs : `vm.affinity_unset`
- Clones a VM : `vm.clone`
- Creates one or more VMs : `vm.create`
- Deletes one or more VMs : `vm.delete`
- Attaches a new disk drive to a VM : `vm.disk_create`
- Detaches a disk drive from a VM and deletes the underlying disk : `vm.disk_delete`
- Gets details about the disks attached to a VM : `vm.disk_get`

- Lists the disks attached to a VM : `vm.disk_list`
- Updates the backing for the specified disk drive : `vm.disk_update`
- Force VM into the powered off state : `vm.force_off`
- Retrieves information about a VM : `vm.get`
- Attaches a new GPU to a VM : `vm.gpu_assign`
- Detaches a GPU from a VM : `vm.gpu_deassign`
- Initiates a Guest level Reboot of the VMs : `vm.guest_reboot`
- Initiates a Guest level Shutdown of the VMs : `vm.guest_shutdown`
- Lists all VMs : `vm.list`
- Live migrates a VM to another host : `vm.migrate`
- Clear the network function chain for this NIC : `vm.nic_clear_network_function_chain`
- Attaches a network adapter to a VM : `vm.nic_create`
- Deletes a NIC from a VM : `vm.nic_delete`
- Gets details about the NICs attached to a VM : `vm.nic_get`
- Lists the NICs attached to a VM : `vm.nic_list`
- Updates a network adapter, specified by the MAC address, on a VM : `vm.nic_update`
- Updates the network function chain for this NIC : `vm.nic_update_network_function_chain`
- Powers off the specified VMs : `vm.off`
- Powers on the specified VMs : `vm.on`
- Power cycles the specified VMs : `vm.power_cycle`
- Initiates a reboot by issuing an ACPI event : `vm.reboot`
- Resets the specified VMs : `vm.reset`
- Restores a VM to a snapshotted state : `vm.restore`
- Attaches a new serial port to a VM : `vm.serial_port_create`
- Detaches a serial port from a VM : `vm.serial_port_delete`
- Initiates a shutdown by issuing an ACPI event : `vm.shutdown`
- Creates one or more snapshots in a single consistency group : `vm.snapshot_create`
- Prints the graph representation of the snapshot history for a VM : `vm.snapshot_get_tree`
- Gets a list of all snapshots associated with a VM : `vm.snapshot_list`
- Updates the specified VMs : `vm.update`
- Updates a VM's boot device : `vm.update_boot_device`
- Updates storage container of all disks of the VM or just the specified disks : `vm.update_container`

Retrieves affinity nodes for a given VM

`<acropolis> vm.affinity_nodeget` *vm_list*

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Enable VM-host affinity

Host_list is the list of hosts that VM can affine to.

`<acropolis> vm.affinity_set` *vm_list* `[ host_list="`*host_list*`" ]`

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Optional arguments

host_list

Host list

Type: list of hosts

Unsets affinity setting of specified VMs

This will unset a VM affinity configuration, including policy, constraint, and binding entities.

`<acropolis> vm.affinity_unset` *vm_list*

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Clones a VM

The following suffixes are valid: M=2^20, G=2^30.

`<acropolis> vm.clone` *name_list* `[ clone_affinity="`*{ true | false }*`" ][ clone_from_snapshot="`*clone_from_snapshot*`" ][ clone_from_vm="`*clone_from_vm*`" ][ clone_ip_address="`*clone_ip_address*`" ][ memory="`*memory*`" ][ num_cores_per_vcpu="`*num_cores_per_vcpu*`" ][ num_threads_per_core="`*num_threads_per_core*`" ][ num_vcpus="`*num_vcpus*`" ]`

Required arguments

name_list

Comma-delimited list of VM names

Type: list of strings with expansion wildcards

Optional arguments

clone_affinity

Clone source VM's affinity rules.

Type: boolean

clone_from_snapshot

Snapshot from which to clone

Type: snapshot

clone_from_vm

VM from which to clone

Type: VM

clone_ip_address

IP addresses to assign to clones

Type: list of IPv4 addresses

memory

Memory size

Type: size with MG suffix

num_cores_per_vcpu

Number of cores per vCPU

Type: int

num_threads_per_core

Number of threads per core

Type: int

num_vcpus

Number of vCPUs

Type: int

Creates one or more VMs

Enabling metrics allows host-specific metrics to be percolated to this VM.

```
<acropolis> vm.create name_list [ agent_vm="{ true | false }" ][ disable_branding="{ true
| false }" ][ enable_metrics="{ true | false }" ][ extra_flags="extra_flags"
][ flash_mode="{ true | false }" ][ hardware_virtualization="{ true
| false }" ][ machine_type="machine_type" ][ memory="memory" ][
memory_overcommit="{ true | false }" ][ num_cores_per_vcpu="num_cores_per_vcpu"
][ num_threads_per_core="num_threads_per_core" ][ num_vcpus="num_vcpus" ][
num_vnuma_nodes="num_vnuma_nodes" ][ nvram_container="nvram_container" ][
secure_boot="{ true | false }" ][ uefi_boot="{ true | false }" ][ vcpu_hard_pin="{ true | false }"
][ windows_credential_guard="{ true | false }" ]
```

Required arguments

name_list

Comma-delimited list of VM names

Type: list of strings with expansion wildcards

Optional arguments

agent_vm

Agent VM

Type: boolean

disable_branding

Disable Nutanix branding

Type: boolean

`enable_metrics`

Enable host metrics for this VM

Type: boolean

`extra_flags`

Additional VM flags as key=value pairs, separated by semicolon

Type: string

`flash_mode`

Enable/Disable flash mode for all VM disks.

Type: boolean

`hardware_virtualization`

Virtual machine extention for hardware virtualization

Type: boolean

`machine_type`

Machine architecture type

Type: Machine Architecture Type

`memory`

Memory size

Type: size with MG suffix

Default: 2G

`memory_overcommit`

If True, the configured memory for the VM may reside outside of the host physical memory.

Type: boolean

`num_cores_per_vcpu`

Number of cores per vCPU

Type: int

`num_threads_per_core`

Number of threads per core

Type: int

`num_vcpus`

Number of vCPUs

Type: int

Default: 1

`num_vnuma_nodes`

Number of vNUMA nodes

Type: int

`nvram_container`

Container to store NVRAM disk if uefi_boot=True

Type: container

secure_boot

UVM Secure boot

Type: boolean

uefi_boot

UEFI boot

Type: boolean

vcpu_hard_pin

Enable hard pinning vcpu to pcpus

Type: boolean

windows_credential_guard

Windows UVM credential guard

Type: boolean

Deletes one or more VMs

If the VM is powered on, it will be powered off and then deleted.

`<acropolis> vm.delete` *vm_list* `[ delete_snapshots="`*{ true | false }*`" ]`

Required arguments

vm_list

Comma-delimited VM identifiers

Type: list of VMs

Optional arguments

delete_snapshots

Delete snapshots?

Type: boolean

Default: false

Attaches a new disk drive to a VM

Note that certain buses, like IDE, are not hot-pluggable.

`<acropolis> vm.disk_create` *vm* `[ bus="`*bus*`" ][ cdrom="`*{ true | false }*`" ][ clone_from_adsf_file="`*clone_from_adsf_file*`" ][ clone_from_external_url="`*clone_from_external_url*`" ][ clone_from_image="`*clone_from_image*`" ][ clone_from_vmdisk="`*clone_from_vmdisk*`" ][ clone_min_size="`*clone_min_size*`" ][ container="`*container*`" ][ create_size="`*create_size*`" ][ device_uuid="`*device_uuid*`" ][ empty="`*{ true | false }*`" ][ flash_mode="`*{ true | false }*`" ][ index="`*index*`" ][ passthru_from_external_url="`*passthru_from_external_url*`" ][ scsi_passthru="`*{ true | false }*`" ]`

Required arguments

vm

VM identifier

Type: VM

Optional arguments

bus

Device bus

Type: Bus Type

`cdrom`

Indicates if the disk is a CDROM drive

Type: boolean

`clone_from_adsf_file`

Path to an ADSF file

Type: ADSF path

`clone_from_external_url`

URL of the external datasource to clone from

Type: string

`clone_from_image`

An image name/UUID

Type: image

`clone_from_vmdisk`

A vmdisk UUID

Type: VM disk

`clone_min_size`

Minimum size of the resulting clone (only applies to cloned disks)

Type: size with cskKmMgGtT suffix

`container`

Container (only applies to newly-created disks and clones from external datasources)

Type: container

`create_size`

Size of new disk

Type: size with cskKmMgGtT suffix

`device_uuid`

Device UUID

Type: UUID

`empty`

Whether the disk is empty (only applies to CDROMs)

Type: boolean

`flash_mode`

If true, enable flash mode on the disk.

Type: boolean

`index`

Device index on bus

Type: int

`passthru_from_external_url`

URL of the external datasource for passthrough disk

Type: string

scsi_passthru

Passthrough disk?

Type: boolean

Examples

1. Create a blank 5GiB disk on ctr, and attach it as SCSI:3.

```
<acropolis> vm.disk_create my_vm create_size=5G container=ctr bus=scsi index=3
```

2. Clone a disk from the ADSF file /ctr/plan9.iso, and use it as the backing image for a newly-created CD-ROM drive on the first available IDE slot.

```
<acropolis> vm.disk_create my_vm clone_from_adsf_file=/ctr/plan9.iso cdrom=1
```

3. Clone a disk from the existing vmdisk, and attach it to the first available SCSI slot.

```
<acropolis> vm.disk_create my_vm clone_from_vmdisk=0b4fc60b-cc56-41c6-911e-67cc8406d096
```

4. Create a disk from an Acropolis image and attach it to the first SCSI slot.

```
<acropolis> vm.disk_create my_vm clone_from_image=my_image
```

5. Create a new empty CD-ROM drive, and attach it to the first available IDE slot.

```
<acropolis> vm.disk_create my_vm empty=1 cdrom=1
```

6. Create a disk in 'ctr' with data cloned from the datasource specified by clone_from_external_url.

```
<acropolis> vm.disk_create my_vm container=ctr clone_from_external_url=nfs://server.fqdn/
export1/sql_server/data_disk1
```

7. Create a passthrough disk in 'ctr' that forwards both reads and writes to the datasource specified by passthru_from_external_url.

```
<acropolis> vm.disk_create my_vm container=ctr passthru_from_external_url=nfs://server.fqdn/
export1/temp/data_disk1
```

Detaches a disk drive from a VM and deletes the underlying disk

Note that certain buses, like IDE, are not hot-pluggable.

`<acropolis> vm.disk_delete` *vm* `[ device_uuid="`*device_uuid*`" ][ disk_addr="`*disk_addr*`" ]`

Required arguments

vm

VM identifier

Type: VM

Optional arguments

device_uuid

Device UUID

Type: UUID

disk_addr

Disk address ("bus.index")

Type: VM disk

Gets details about the disks attached to a VM

`<acropolis> vm.disk_get` *vm* `[ device_uuid="`*device_uuid*`" ][ disk_addr="`*disk_addr*`" ][ include_vmdisk_datasource_info="{`*true* | *false*`}" ][ include_vmdisk_paths="{`*true* | *false*`}" ][ include_vmdisk_sizes="{`*true* | *false*`}" ]`

Required arguments

vm

VM identifier

Type: VM

Optional arguments

`device_uuid`

Device UUID

Type: UUID

`disk_addr`

Disk address ("bus.index")

Type: VM disk

`include_vmdisk_datasource_info`

Fetch the backing external datasource details, if applicable

Type: boolean

Default: false

`include_vmdisk_paths`

Fetch vmdisk paths

Type: boolean

Default: false

`include_vmdisk_sizes`

Fetch vmdisk sizes (in bytes)

Type: boolean

Default: true

Lists the disks attached to a VM

`<acropolis> vm.disk_list` *vm*

Required arguments

vm

VM identifier

Type: VM

Updates the backing for the specified disk drive

The existing disk image will be deleted and replaced by the new image (which may be a clone of the existing image).

`<acropolis> vm.disk_update` *vm* `[ clone_from_adsf_file="`*clone_from_adsf_file*`" ][ clone_from_image="`*clone_from_image*`" ][ clone_from_vmdisk="`*clone_from_vmdisk*`" ][ clone_min_size="`*clone_min_size*`" ][ container="`*container*`" ][ create_size="`*create_size*`" ][ device_uuid="`*device_uuid*`" ][ disk_addr="`*disk_addr*`" ][ empty="{`*true* | *false*`}" ][ flash_mode="{`*true* | *false*`}" ][ new_size="`*new_size*`" ]`

Required arguments

`vm`

VM identifier

Type: VM

Optional arguments

`clone_from_adsf_file`

Path to an ADSF file

Type: ADSF path

`clone_from_image`

An image name/UUID

Type: image

`clone_from_vmdisk`

A vmdisk UUID

Type: VM disk

`clone_min_size`

Minimum size of the resulting clone (only applies to cloned disks)

Type: size with cskKmMgGtT suffix

`container`

Container (only applies to newly-created disks)

Type: container

`create_size`

Size of new disk

Type: size with cskKmMgGtT suffix

`device_uuid`

Device UUID

Type: UUID

`disk_addr`

Disk address ("bus.index")

Type: VM disk

`empty`

Whether the disk is empty (only applies to CDROMs)

Type: boolean

`flash_mode`

Set flash mode on the specified disk.

Type: boolean

`new_size`

New size for the existing disk

Type: size with cskKmMgGtT suffix

Examples

1. Replace the disk at SCSI:0 with blank 5GiB disk on ctr.

```
<acropolis> vm.disk_update my_vm disk_addr=scsi.0 create_size=5G container=ctr
```

2. Replace the disk at IDE:0 with a clone of /ctr/plan9.iso. Note that if IDE:0 is a CD-ROM drive, it remains such.

```
<acropolis> vm.disk_update my_vm disk_addr=ide.0 clone_from_adsf_file=/ctr/plan9.iso
```

3. Replace the disk at SCSI:0 with a clone of the existing vmdisk.

```
<acropolis> vm.disk_update my_vm disk_addr=scsi.0 clone_from_vmdisk=0b4fc60b-
cc56-41c6-911e-67cc8406d096
```

4. Eject the image from the CD-ROM drive at IDE:0.

```
<acropolis> vm.disk_update my_vm disk_addr=ide.0 empty=1
```

5. Update the size of disk at SCSI:1 to 5GiB.

```
<acropolis> vm.disk_update my_vm disk_addr=scsi.1 new_size=5G
```

6. Enable flash mode for an existing disk at SCSI:0.

```
<acropolis> vm.disk_update my_vm disk_addr=scsi.0 flash_mode=True
```

7. Enable flash mode for an existing disk at SCSI:0 and resize the disk.

```
<acropolis> vm.disk_update my_vm disk_addr=scsi.0 new_size=100G flash_mode=True
```

Force VM into the powered off state

Therefore, the user should take adequate precautions to ensure that the old instance is really gone.

`<acropolis> vm.force_off` *vm*

Required arguments

vm

VM identifier

Type: VM

Retrieves information about a VM

`<acropolis> vm.get` *vm_list* `[ include_address_assignments="`*{ true | false }*`" ][ include_vmdisk_datasource_info="`*{ true | false }*`" ][ include_vmdisk_paths="`*{ true | false }*`" ][ include_vmdisk_sizes="`*{ true | false }*`" ]`

Required arguments

vm_list

VM identifier

Type: list of VMs

Optional arguments

include_address_assignments

Fetch configured IP addresses

Type: boolean

Default: true

include_vmdisk_datasource_info

Fetch the backing external datasource details, if applicable

Type: boolean

Default: false

`include_vmdisk_paths`

Fetch vmdisk paths

Type: boolean

Default: false

`include_vmdisk_sizes`

Fetch vmdisk sizes (in bytes)

Type: boolean

Default: true

Attaches a new GPU to a VM

Changes to the GPU configuration can only be made when VM is powered off.

`<acropolis> vm.gpu_assign` *vm* `[ extra_param="`*extra_param*`" ][ gpu="`*gpu*`" ]`

Required arguments

`vm`

VM identifier

Type: VM

Optional arguments

`extra_param`

extra_parameters

Type: string

`gpu`

GPU

Type: gpu

Examples

1. Add a new GPU in passthrough mode

```
<acropolis> vm.gpu_assign my_vm gpu=Nvidia_Tesla_M60
```

Detaches a GPU from a VM

Changes to the GPU configuration can only be made when VM is powered off.

`<acropolis> vm.gpu_deassign` *vm* `[ gpu="`*gpu*`" ]`

Required arguments

`vm`

VM identifier

Type: VM

Optional arguments

`gpu`

GPU

Type: gpu

Examples

1. Remove a new GPU in passthrough mode

```
<acropolis> vm.gpu_deassign my_vm gpu=Nvidia_Tesla_M60
```

Initiates a Guest level Reboot of the VMs

Requires NGT.

`<acropolis> vm.guest_reboot` *vm_list* `[ enable_script_exec="`*{ true | false }*`" ][ fail_on_script_failure="`*{ true | false }*`" ]`

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Optional arguments

enable_script_exec

Whether to execute pre-reboot script.

Type: boolean

Default: false

fail_on_script_failure

Whether to abort reboot if script fails.

Type: boolean

Default: false

Initiates a Guest level Shutdown of the VMs

Requires NGT.

`<acropolis> vm.guest_shutdown` *vm_list* `[ enable_script_exec="`*{ true | false }*`" ][ fail_on_script_failure="`*{ true | false }*`" ]`

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Optional arguments

enable_script_exec

Whether to execute pre-shutdown script.

Type: boolean

Default: false

fail_on_script_failure

Whether to abort shutdown if script fails.

Type: boolean

Default: false

Lists all VMs

`<acropolis> vm.list [ memory="`*memory*`" ][ num_vcpus="`*num_vcpus*`" ][ power_state="`*power_state*`" ]`

Required arguments

None

Optional arguments

`memory`

List VMs with the given amount of memory

Type: size with MG suffix

`num_vcpus`

List VMs with the given number of VCPUs

Type: int

`power_state`

List VMs with the given power state

Type: Power State

Live migrates a VM to another host

This limit is applied to each of the migrations individually.

`<acropolis> vm.migrate` *vm_list* `[ bandwidth_mbps="`*bandwidth_mbps*`" ][ host="`*host*`" ][ live="`*{ true | false }*`" ][ max_downtime_ms="`*max_downtime_ms*`" ]`

Required arguments

`vm_list`

Comma-delimited list of VM identifiers

Type: list of VMs

Optional arguments

`bandwidth_mbps`

Maximum bandwidth in MiB/s

Type: int

Default: 0

`host`

Destination host

Type: host

`live`

Live migration or suspended migration?

Type: boolean

Default: true

`max_downtime_ms`

Maximum time in ms for which this Vm can be stunned while migrating

Type: int

Default: 0

Clear the network function chain for this NIC

`<acropolis> vm.nic_clear_network_function_chain` *vm mac_addr*

Required arguments

vm

VM identifier

Type: VM

mac_addr

NIC MAC address

Type: NIC address

Attaches a network adapter to a VM

If the VM is running, the NIC is hot-added to the VM.

```
<acropolis> vm.nic_create vm [ allow_unknown_macs="{ true | false }" ][
connected="{ true | false }" ][ ip="ip" ][ mac="mac" ][ model="model" ][
network="network" ][ network_function_nic_type="network_function_nic_type" ][
queues="queues" ][ request_ip="{ true | false }" ][ rx_queue_size="rx_queue_size"
][ secondary_ip_addresses="secondary_ip_addresses" ][
trunked_networks="trunked_networks" ][ type="type" ][ vlan_mode="vlan_mode" ]
```

Required arguments

vm

VM identifier

Type: VM

Optional arguments

allow_unknown_macs

If true, then allow unknown unicast traffic to be forwarded to this NIC

Type: boolean

connected

Whether or not the NIC is connected.

Type: boolean

ip

IPv4 address

Type: IPv4 address

mac

MAC address

Type: MAC address

model

Virtual hardware model. Defaults to 'virtio', can also specify 'e1000'.

Type: string

network

Network identifier

Type: network

network_function_nic_type

Network function nic type

Type: Network function NIC Type

queues

Maximum number of Tx/Rx queue pairs (default: 1)

Type: int

request_ip

If true, then try to request an IP from Acropolis (static if the 'ip=' field is set, otherwise dynamic if the network is managed.) If false, then don't request an IP from Acropolis at all (even if 'ip=' is set.

Type: boolean

rx_queue_size

Receive queue depth (default: 256)

Type: int

secondary_ip_addresses

List of secondary IP Addresses.

Type: list of IPv4 addresses

trunked_networks

List of trunked networks.

Type: list of ints

type

NIC Type

Type: NIC Type

Default: 1

vlan_mode

VLan Mode. Access by default.

Type: Vlan Type

Deletes a NIC from a VM

If the NIC to be removed is specified as the boot device in the boot configuration, the boot device configuration will be cleared as a side effect of removing the NIC.

`<acropolis> vm.nic_delete` *vm mac_addr*

Required arguments

vm

VM identifier

Type: VM

mac_addr

NIC MAC address

Type: NIC address

Gets details about the NICs attached to a VM

`<acropolis> vm.nic_get` *vm* `[ include_address_assignments="{` *true | false* `}" ][ mac_addr="`*mac_addr*`" ]`

Required arguments

vm

VM identifier

Type: VM

Optional arguments

`include_address_assignments`

Fetch configured IP addresses

Type: boolean

Default: true

`mac_addr`

NIC MAC address

Type: NIC address

Lists the NICs attached to a VM

`<acropolis> vm.nic_list` *vm* `[ include_address_assignments="`*{ true | false }*`" ]`

Required arguments

vm

VM identifier

Type: VM

Optional arguments

`include_address_assignments`

Fetch configured IP addresses

Type: boolean

Default: true

Updates a network adapter, specified by the MAC address, on a VM

4) List of Secondary IP Addresses.

`<acropolis> vm.nic_update` *vm mac_addr* `[ allow_unknown_macs="`*{ true | false }*`" ][ connected="`*{ true | false }*`" ][ ip="`*ip*`" ][ network="`*network*`" ][ queues="`*queues*`" ][ request_ip="`*{ true | false }*`" ][ rx_queue_size="`*rx_queue_size*`" ][ secondary_ip_addresses="`*secondary_ip_addresses*`" ][ trunked_networks="`*trunked_networks*`" ][ type="`*type*`" ][ update_vlan_trunk_info="`*{ true | false }*`" ][ vlan_mode="`*vlan_mode*`" ]`

Required arguments

vm

VM identifier

Type: VM

mac_addr

NIC MAC address

Type: NIC address

Optional arguments

`allow_unknown_macs`

If true, then allow unknown unicast traffic to be forwarded to this NIC

Type: boolean

connected

Whether or not the NIC is connected.

Type: boolean

ip

IPv4 address

Type: IPv4 address

network

Network identifier

Type: network

queues

Maximum number of Tx/Rx queue pairs (default: 1)

Type: int

request_ip

If true, request a new IP address.

Type: boolean

rx_queue_size

Receive queue depth (default: 256)

Type: int

secondary_ip_addresses

List of secondary IP Addresses.

Type: list of IPv4 addresses

trunked_networks

List of trunked networks.

Type: list of ints

type

NIC Type

Type: NIC Type

update_vlan_trunk_info

If true, then update vlan type and trunked network list.

Type: boolean

vlan_mode

VLan Mode. Access by default.

Type: Vlan Type

Updates the network function chain for this NIC

`<acropolis> vm.nic_update_network_function_chain` *vm mac_addr chain*

Required arguments

vm

VM identifier

Type: VM

mac_addr

NIC MAC address

Type: NIC address

chain

Network function chain

Type: network function chain

Powers off the specified VMs

`<acropolis> vm.off` *vm_list*

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Powers on the specified VMs

Note that no such host may be available.

`<acropolis> vm.on` *vm_list* `[ enable_migrations="{` *true | false* `}" ][ host="`*host*`" ]`

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Optional arguments

enable_migrations

Whether live migrations of other VMs should be performed to free up space in order to power on the VM on a node

Type: boolean

Default: false

host

Host on which to power on the VM

Type: host

Power cycles the specified VMs

`<acropolis> vm.power_cycle` *vm_list* `[ change_host="{` *true | false* `}" ][ host="`*host*`" ]`

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Optional arguments

change_host

Whether to power on on a different host

Type: boolean

Default: false

host

Host on which to power on the VM

Type: host

Initiates a reboot by issuing an ACPI event

`<acropolis> vm.reboot` *vm_list*

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Resets the specified VMs

`<acropolis> vm.reset` *vm_list*

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Restores a VM to a snapshotted state

In this case, the user may choose not to restore the VM's network adapters using the "restore_network_config" keyword argument.

`<acropolis> vm.restore` *vm snapshot* `[ restore_network_config="`*{ true | false }*`" ]`

Required arguments

vm

VM identifier

Type: VM

snapshot

Snapshot identifier

Type: snapshot

Optional arguments

restore_network_config

Whether to restore the VM's networking configuration

Type: boolean

Default: true

Attaches a new serial port to a VM

Changes to the serial port configuration only take effect after a full power cycle.

`<acropolis> vm.serial_port_create` *vm* `[ index="`*index*`" ][ type="`*type*`" ]`

Required arguments

vm

VM identifier

Type: VM

Optional arguments

`index`

Serial port index

Type: int

`type`

Serial port type

Type: serial port type

Examples

1. Add a new serial port at COM1 in server mode.

```
<acropolis> vm.serial_port_create my_vm index=0 type=kServer
```

Detaches a serial port from a VM

Changes to the serial port configuration only take effect after a full power cycle.

`<acropolis> vm.serial_port_delete` *vm index*

Required arguments

vm

VM identifier

Type: VM

index

Serial port index

Type: int

Examples

1. Remove the serial port at COM2.

```
<acropolis> vm.serial_port_delete my_vm 1
```

Initiates a shutdown by issuing an ACPI event

`<acropolis> vm.shutdown` *vm_list*

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Creates one or more snapshots in a single consistency group

If no snapshot name is provided, the snapshot will be referred to as "<vm_name>-<timestamp>", where the timestamp is in ISO 8601 format (YYYY-MM-DDTHH:MM:SS.mmmmmm).

`<acropolis> vm.snapshot_create` *vm_list* `[ snapshot_name_list="`*snapshot_name_list*`" ]`

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Optional arguments

snapshot_name_list

Comma-delimited list of names for each snapshot

Type: list of strings

Examples

1. Create a snapshot named 'dev-vm-gold' from a VM named 'dev-vm'.

```
<acropolis> vm.snapshot_create dev-vm snapshot_name_list=dev-vm-gold
```

2. Create a consistent snapshot across several VMs, using the default naming scheme.

```
<acropolis> vm.snapshot_create vm1,vm2,vm3
```

Prints the graph representation of the snapshot history for a VM

`<acropolis> vm.snapshot_get_tree` *vm*

Required arguments

vm

VM identifier

Type: VM

Gets a list of all snapshots associated with a VM

`<acropolis> vm.snapshot_list` *vm*

Required arguments

vm

VM identifier

Type: VM

Updates the specified VMs

Enabling metrics allows host-specific metrics to be percolated to this VM.

`<acropolis> vm.update` *vm_list* `[ agent_vm="`*{ true | false }*`" ][ annotation="`*annotation*`" ][ cbr_not_capable_reason="`*cbr_not_capable_reason*`" ][ cpu_passthrough="`*{ true | false }*`" ][ disable_branding="`*{ true | false }*`" ][ enable_metrics="`*{ true | false }*`" ][ extra_flags="`*extra_flags*`" ][ flash_mode="`*{ true | false }*`" ][ gpu_console="`*{ true | false }*`" ][ ha_priority="`*ha_priority*`" ][ hardware_virtualization="`*{ true | false }*`" ][ hwclock_timezone="`*hwclock_timezone*`" ][ machine_type="`*machine_type*`" ][ memory="`*memory*`" ][ memory_overcommit="`*{ true | false }*`" ][ name="`*name*`" ][ num_cores_per_vcpu="`*num_cores_per_vcpu*`" ][ num_threads_per_core="`*num_threads_per_core*`" ][ num_vcpus="`*num_vcpus*`" ][ num_vnuma_nodes="`*num_vnuma_nodes*`" ][ nvram_container="`*nvram_container*`" ][ secure_boot="`*{ true | false }*`" ][ uefi_boot="`*{ true | false }*`" ][ vcpu_hard_pin="`*{ true | false }*`" ][ vga_console="`*{ true | false }*`" ][ windows_credential_guard="`*{ true | false }*`" ]`

Required arguments

vm_list

Comma-delimited list of VM identifiers

Type: list of VMs

Optional arguments

agent_vm

Agent vm

Type: boolean

`annotation`

Annotation string

Type: string

`cbr_not_capable_reason`

If set, marks the VM incapable of CBR workflows

Type: string

`cpu_passthrough`

Pass through all host CPU features.

Type: boolean

`disable_branding`

Disable Nutanix branding

Type: boolean

`enable_metrics`

Enable host metrics

Type: boolean

`extra_flags`

Additional VM flags as key=value pairs, separated by semicolon

Type: string

`flash_mode`

Enable/Disable flash mode for all VM disks.

Type: boolean

`gpu_console`

Enable/Disable GPU console for the VM.

Type: boolean

`ha_priority`

Numeric priority for HA restart. Negative value indicates no restart.

Type: int

`hardware_virtualization`

Virtual machine extention for hardware virtualization

Type: boolean

`hwclock_timezone`

Hardware clock timezone

Type: timezone

`machine_type`

Machine architecture type

Type: Machine Architecture Type

`memory`

Memory size

Type: size with MG suffix

`memory_overcommit`

If True, the configured memory for the VM may reside outside of the host physical memory.

Type: boolean

`name`

VM name

Type: string

`num_cores_per_vcpu`

Number of cores per vCPU

Type: int

`num_threads_per_core`

Number of threads per core

Type: int

`num_vcpus`

Number of vCPUs

Type: int

`num_vnuma_nodes`

Number of vNUMA nodes

Type: int

`nvram_container`

Container to store NVRAM disk if uefi_boot=True

Type: container

`secure_boot`

Secure boot

Type: boolean

`uefi_boot`

UEFI boot

Type: boolean

`vcpu_hard_pin`

Enable hard pinning vcpu to pcpus

Type: boolean

`vga_console`

VGA console device

Type: boolean

`windows_credential_guard`

Windows UVM credential guard

Type: boolean

Updates a VM's boot device

Changes to the boot order only take effect after a full VM power cycle.

`<acropolis> vm.update_boot_device` *vm* `[ boot_device_order="`*boot_device_order*`" ][ disk_addr="`*disk_addr*`" ][ mac_addr="`*mac_addr*`" ]`

Required arguments

vm

VM identifier

Type: VM

Optional arguments

boot_device_order

Device boot order

Type: list of boot device types

disk_addr

Disk bus address

Type: VM disk

mac_addr

NIC MAC address

Type: NIC address

Updates storage container of all disks of the VM or just the specified disks

`<acropolis> vm.update_container` *vm* `[ container="`*container*`" ][ device_uuid_list="`*device_uuid_list*`" ][ disk_addr_list="`*disk_addr_list*`" ][ wait="`*{ true | false }*`" ]`

Required arguments

vm

VM identifier

Type: VM

Optional arguments

container

Container

Type: container

device_uuid_list

List of devices

Type: list of UUIDs

disk_addr_list

Disk address ("bus.index")

Type: VM disk list

wait

Wait for tasks to complete

Type: boolean

Default: true

Examples

1. Update container of all disks of the VM.

```
<acropolis> vm.update_container my_vm container=target_ctr
```

2. Update container of disk at SCSI:0 AND SCSI:1.

```
<acropolis> vm.update_container my_vm container=target_ctr disk_addr_list=scsi.0,scsi.1
```

3. Update container of disk by specified device UUID.

```
<acropolis> vm.update_container my_vm container=target_ctrdevice_uuid_list=99359b95-
f8b0-4100-8c4b-88b1ddacf676,e6d85ee8-3797-4bae-98f2-9e4511be908d
```

vm_group

Operations

- Add VMs to a VM group : `vm_group.add_vms`
- Enables vm-vm preferential anti-affinity : `vm_group.antiaffinity_set`
- Clears vm-vm anti-affinity : `vm_group.antiaffinity_unset`
- Creates one or more VM groups : `vm_group.create`
- Deletes one or more VM groups : `vm_group.delete`
- Retrieves information about a VM group : `vm_group.get`
- Lists all VM groups : `vm_group.list`
- Lists VMs configured to the VmGroup : `vm_group.list_vms`
- Remove VMs from a VM group : `vm_group.remove_vms`
- Updates the specified VM groups : `vm_group.update`

Add VMs to a VM group

`<acropolis> vm_group.add_vms` *vm_group* `[ vm_list="`*vm_list*`" ]`

Required arguments

vm_group

VmGroup identifier

Type: VM group type

Optional arguments

vm_list

Comma-delimited VM identifiers

Type: list of VMs

Enables vm-vm preferential anti-affinity

It is OK to violate the rule during VM placement if scheduler cannot honor the policy.

`<acropolis> vm_group.antiaffinity_set` *vm_group_list*

Required arguments

vm_group_list

Comma-delimited list of VM group identifiers

Type: list of VM group types

Clears vm-vm anti-affinity

`<acropolis> vm_group.antiaffinity_unset` *vm_group_list*

Required arguments

vm_group_list

Comma-delimited list of VM group identifiers

Type: list of VM group types

Creates one or more VM groups

`<acropolis> vm_group.create` *name_list*

Required arguments

name_list

Comma-delimited list of VM group names

Type: list of strings

Deletes one or more VM groups

`<acropolis> vm_group.delete` *vm_group_list*

Required arguments

vm_group_list

Comma-delimited VM group identifiers

Type: list of VM group types

Retrieves information about a VM group

`<acropolis> vm_group.get` *vm_group_list*

Required arguments

vm_group_list

VM group identifier

Type: list of VM group types

Lists all VM groups

`<acropolis> vm_group.list`

Required arguments

None

Lists VMs configured to the VmGroup

`<acropolis> vm_group.list_vms` *vm_group*

Required arguments

vm_group

VmGroup identifier

Type: VM group type

Remove VMs from a VM group

`<acropolis> vm_group.remove_vms` *vm_group* `[ vm_list="`*vm_list*`" ]`

Required arguments

vm_group

VmGroup identifier

Type: VM group type

Optional arguments

`vm_list`

Comma-delimited VM identifiers

Type: List of VM configured to VmGroup

Updates the specified VM groups

`<acropolis> vm_group.update` *vm_group_list* `[ name="`*name*`" ]`

Required arguments

vm_group_list

Comma-delimited list of VM group identifiers

Type: list of VM group types

Optional arguments

`name`

VM group name

Type: string

2

NUTANIX COMMAND-LINE INTERFACE (NCLI)

The Nutanix command-line interface (nCLI) allows you to run system administration commands against the Nutanix cluster from any of the following machines:

- Your local machine (preferred)
- Any Controller VM in the cluster

Installing the nCLI on Your Local System

About this task

Tip: Refer to *KB-1661* for the default credentials of all cluster components.

Procedure

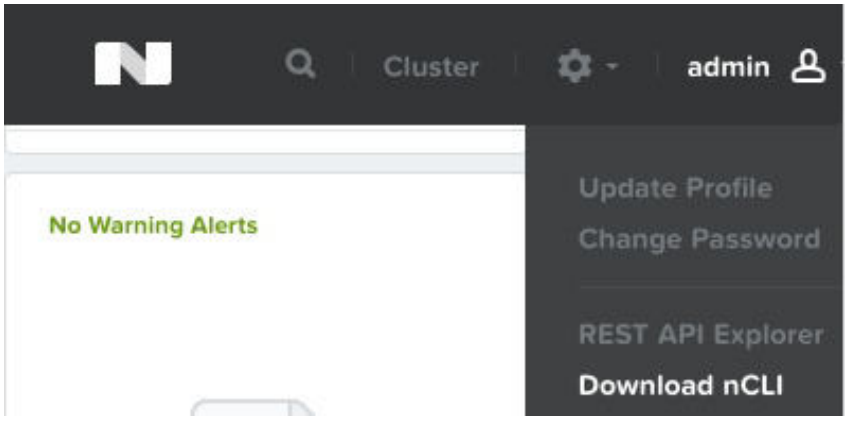
1. Verify that your system has Java Runtime Environment (JRE) version 5.0 or higher.
   To check which version of Java is installed on your system or to download the latest version, go to http://www.java.com/en/download/installed.jsp.

2. Download the nCLI.

   a. Connect to the Nutanix web console.

   b. Click the user icon

   at the top of the console.

   c. Select **Download nCLI**.

   d. Download and unzip the file on your local system.

3. Configure your Windows %PATH% or Linux $PATH environment variables.

   - `ncli` directory (for example, `C:\ncli`)
   - JRE `bin` directory (for example, `C:\Program Files\Java\jre1.6.0_21\bin`)

   The procedure to complete this step depends on your operating system. For more information, go to http://java.com/en/download/help/path.xml.

Initiating an nCLI Session From Your Local System

Before you begin

Install the nCLI on your local system by following Installing the nCLI on Your Local System on page 64.

Procedure

1. On your local system, open a command prompt (such as `bash` for Linux or `CMD` for Windows).
2. At the command prompt, start the nCLI by using one of the following commands.
   a. `ncli -s` *`management_ip_addr`* `-u '`*`username`*`' -p '`*`user_password`*`'`
      In this case, the console displays the password.
   b. `ncli -s` *`management_ip_addr`* `-u '`*`username`*`' -p`
      In this case, you are prompted to specify the password.

- Replace *`management_ip_addr`* with the IP address of any Nutanix Controller VM in the cluster.
- Replace *`username`* with the name of the user (if not specified, the default is admin).
- (Optional) Replace *`user_password`* with the password of the user.

> Note: When specifying the `-p '`*`user_password`*`'` command-line argument to establish an nCLI session interactively, the credentials are stored in the user's history file. This flag is intended for use in automated scripts or third-party application integration to nCLI.
>
> To prevent retention of the password for interactive login, specify the `-p` command flag alone without the password as described in step 2(b). An interactive prompt is provided to supply the password, which will not be stored in the user's history file.

Table 1: Troubleshooting

Error	Explanation/Resolution
`ncli` not found or not recognized as a command	The Windows %PATH% or Linux $PATH environment variable is not set.
`Error: Bad credentials`	The admin user password has been changed from the default and you did not specify the correct password. Type `exit` and start the nCLI again with the correct password.
`Error: Could not connect to Nutanix Gateway`	The cluster is not started. Log on to a Controller VM as the nutanix user and run the following command: `cluster start` When the cluster is up, exit the nCLI and start it again.

Results

The Nutanix CLI is now in interactive mode. To exit this mode, type `exit` at the `ncli>` prompt.

Command Format

Nutanix Command-Line Interface commands must match the following format:

```
ncli> entity action parameter1=value parameter2=value ...
```

entity can be replaced by any Nutanix entity, such as *cluster* or *disk*.

action can be replaced by any valid action for the preceding entity. Each entity has a unique set of actions, but a common action across all entities is list. For example, you can type the following command to request a list of all storage pools in the cluster.

```
ncli> storagepool list
```

Some actions require parameters at the end of the command. For example, when creating an NFS datastore, you need to provide both the name of the datastore as it will appear to the hypervisor and the name of the source storage container.

```
ncli> datastore create name="NTNX-NFS" ctr-name="nfs-ctr"
```

Parameter-value pairs can be listed in any order, as long as they are preceded by a valid entity and action.

> Tip: To avoid syntax errors, surround all string values with double-quotes, as demonstrated in the preceding example. This is particularly important when specifying parameters that accept a list of values.

Embedded Help

The nCLI provides assistance on all entities and actions. By typing **help** at the command line, you can request additional information at one of three levels of detail.

help

Provides a list of entities and their corresponding actions

***entity* help**

Provides a list of all actions and parameters associated with the entity, as well as which parameters are required, and which are optional

***entity action* help**

Provides a list of all parameters associated with the action, as well as a description of each parameter

The nCLI provides additional details at each level. To control the scope of the nCLI help output, add the detailed parameter, which can be set to either *true* or *false*.

For example, type the following command to request a detailed list of all actions and parameters for the **cluster** entity.

```
ncli> cluster help detailed=true
```

You can also type the following command if you prefer to see a list of parameters for the **cluster edit-params** action without descriptions.

```
ncli> cluster edit-params help detailed=false
```

nCLI Entities

alerts: An Alert
authconfig: Configuration information used to authenticate user
cluster: A Nutanix Complete Cluster
container: A Storage Container is a container for virtual disks
data-at-rest-encryption: Manage data-at-rest-encryption related operations

data-at-rest-encryption-certificate: Manage data-at-rest-encryption related digital certificates
disk: A Physical Disk
events: An Event
host: A Physical Host hosts Virtual Machines
http-proxy: An HTTP Proxy
license: License for a Nutanix cluster
managementserver: An infrastructure management server such as VCenter
multicluster: A Nutanix Management Console to manage multiple clusters
network: Network specific commands
pulse-config: Configuration information used for Pulse setup
rsyslog-config: TLS configuration for RSyslog service
snmp: An SNMP agent
software: NOS Software Release
ssl-certificate: Manage SSL certificates
storagepool: A Pool of Physical Disks
tag: A tag that groups entities
user: A User
virtualmachine: A Virtual Machine

Nutanix Command-Line Interface Reference

CLI Reference Conventions

This command-line interface reference uses the following conventions.

- Parameters in italic are unique to your environment.

  ```
  value
  ```

- Parameters in square brackets are optional.

  ```
  [ value ]
  ```

- Parameters in curly brackets must be one of a limited set of values.

  ```
  { value1 | value2 }
  ```

  One example is boolean parameters: `{ true | false }`

- The `keyword` is a literal string required by the command, and the *`value`* is the unique value for your environment.

  ```
  keyword=value
  ```

`alerts`: Alert

Description	An Alert
Alias	`alert`

Operations

- Acknowledge Alerts : `acknowledge` | `ack`
- Update Alert Configuration : `edit-alert-config` | `update-alert-config`
- List Alert Configuration : `get-alert-config`
- List history of Alerts : `history`
- List of unresolved Alerts : `list` | `ls`
- Resolve Alerts : `resolve`

Acknowledge Alerts

```
ncli> alerts { acknowledge | ack } ids="ids"
```

Required arguments

`ids`

A comma-separated list of ids of the Alerts

Update Alert Configuration

```
ncli> alerts { edit-alert-config | update-alert-config }[ enable="{ true | false }"
][ enable-default-nutanix-email="{ true | false }" ][ enable-email-digest="{ true | false }"
][ email-contacts="email_contacts" ]
```

Required arguments

None

Optional arguments

`enable`

Enable Alert emails

`enable-default-nutanix-email`

Enable default Nutanix email

`enable-email-digest`

Enable alert email digest

`email-contacts`

Comma-separated list of emails to be used while sending alerts. Set to '-' to clear all the existing emails.

List Alert Configuration

```
ncli> alerts { get-alert-config }
```

Required arguments

None

List history of Alerts

```
ncli> alerts { history } duration="duration" [ acknowledged="acknowledged" ][
resolved="resolved" ][ auto-resolved="auto_resolved" ][ max-alerts="max_alerts" ]
```

Required arguments

`duration`

Duration (in days) for getting the history of Alerts

Optional arguments

`acknowledged`

Acknowledged Alerts?

`resolved`

Resolved Alerts?

`auto-resolved`

Auto resolved Alerts?

`max-alerts`

Maximum number of Alerts to fetch (permitted max value = 1000)

Default: 100

List of unresolved Alerts

`ncli> alerts { list | ls }[ acknowledged="`*acknowledged*`" ][ max-alerts="`*max_alerts*`" ]`

Required arguments

None

Optional arguments

`acknowledged`

Acknowledged Alerts?

`max-alerts`

Maximum number of Alerts to fetch (permitted max value = 1000)

Default: 100

Resolve Alerts

`ncli> alerts { resolve } ids="`*ids*`"`

Required arguments

`ids`

A comma-separated list of ids of the Alerts

`authconfig`: Authentication Configuration

Description	Configuration information used to authenticate user
Alias	

Operations

- Add directory role mapping : `add-role-mapping`
- Comma-separated list of values to be added to the existing directory role mapping : `add-to-role-mapping-values`
- Clear Certificate Revocation information : `clear-certificate-revocation`
- Create Directory Service Configuration : `create-directory` | `add-directory`
- delete ca chain certificate : `delete-ca-chain-certificate` | `delete-ca-chain-cert`
- Delete directory role mappings : `delete-role-mapping`
- Update Authentication Configuration : `edit` | `update`
- Update Directory Service Configuration : `edit-directory` | `update-directory`
- Get Client Auth Configuration : `get-client-authentication-config`
- Get the list of entity values for the specified entity type and the directory name : `get-directory-values-by-type`
- List Authentication Configuration : `list` | `ls`
- List Authentication Configuration : `list-directory` | `ls-directory`
- Get the list of directory role mappings : `list-role-mappings` | `ls-role-mappings`
- Remove Authentication Configuration : `remove` | `rm`
- Remove Directory Service Configuration : `remove-directory` | `rm-directory`
- Comma-separated list of values to be removed from the existing directory role mapping : `remove-from-role-mapping-values`
- Set Certificate Revocation information : `set-certificate-revocation`
- Test LDAP Connection and authentication : `test-ldap-connection`
- Enable/Disable client authentication : `update-client-authentication`
- Update service account (CAC) : `update-service-account`

Add directory role mapping

`ncli> authconfig { add-role-mapping } role="`*role*`" entity-type="`*entity_type*`" entity-values="`*entity_values*`" name="`*name*`"`

Required arguments

`role`

Role of the directory.

`entity-type`

Type of directory entity.

`entity-values`

List of entity values.

`name`

Directory Name

Comma-separated list of values to be added to the existing directory role mapping

`ncli> authconfig { add-to-role-mapping-values } name="name" role="role" entity-type="entity_type" entity-values="entity_values"`

Required arguments

`name`

Name

`role`

Role

`entity-type`

Entity Type

`entity-values`

Values

Clear Certificate Revocation information

`ncli> authconfig { clear-certificate-revocation }`

Required arguments

None

Create Directory Service Configuration

`ncli> authconfig { create-directory | add-directory } directory-type="directory_type" connection-type="connection_type" directory-url="directory_url" domain="domain" name="name" service-account-username="service_account_username" [group-search-type="group_search_type"][user-object-class="user_object_class"][user-search-base="user_search_base"][username-attribute="username_attribute"][group-object-class="group_object_class"][group-search-base="group_search_base"][group-member-attribute="group_member_attribute"][group-member-attribute-value="group_member_attribute_value"][service-account-password="service_account_password"]`

Required arguments

`directory-type`

Type of the Directory Service

`connection-type`

Connection type for the Directory Service

`directory-url`

Url to connect to the Directory Service

`domain`

Domain of the Directory Service

`name`

Name of the Directory Service

`service-account-username`

User name of the directory administrator

Optional arguments

`group-search-type`

Type of the search whether RECURSIVE or NON_RECURSIVE

`user-object-class`

Object class for users

`user-search-base`

Search base for users

`username-attribute`

Unique identifier for a user that can be used for authentication

`group-object-class`

Object class for groups

`group-search-base`

Search base for groups

`group-member-attribute`

Attribute in group that associates users to group

`group-member-attribute-value`

User attributevalue that will be used in group entity to associate user to the group

`service-account-password`

Password of the directory administrator

delete ca chain certificate

`ncli> authconfig { delete-ca-chain-certificate | delete-ca-chain-cert } ca-chain-certificate-name="`*ca_chain_certificate_name*`"`

Required arguments

`ca-chain-certificate-name`

Name of the Certificate

Delete directory role mappings

`ncli> authconfig { delete-role-mapping } name="`*name*`" [ role="`*role*`" ][ entity-type="`*entity_type*`" ]`

Required arguments

`name`

Directory Name

Optional arguments

`role`

Role

`entity-type`

Entity Type

Update Authentication Configuration

`ncli> authconfig { edit | update } auth-type-list="`*auth_type_list*`" [ directory-list="`*directory_list*`" ]`

Required arguments

`auth-type-list`

Authentication Types

Optional arguments

`directory-list`

Id of an alert type.

Update Directory Service Configuration

```
ncli> authconfig { edit-directory | update-directory } directory-type="directory_type" connection-type="connection_type" directory-url="directory_url" domain="domain" name="name" [ group-search-type="group_search_type" ][ user-object-class="user_object_class" ][ user-search-base="user_search_base" ][ username-attribute="username_attribute" ][ group-object-class="group_object_class" ][ group-search-base="group_search_base" ][ group-member-attribute="group_member_attribute" ][ group-member-attribute-value="group_member_attribute_value" ][ service-account-username="service_account_username" ][ service-account-password="service_account_password" ]
```

Required arguments

`directory-type`

Type of the Directory Service

`connection-type`

Connection type for the Directory Service

`directory-url`

Url to connect to the Directory Service

`domain`

Domain of the Directory Service

`name`

Name of the Directory Service

Optional arguments

`group-search-type`

Type of the search whether RECURSIVE or NON_RECURSIVE

`user-object-class`

Object class for users

`user-search-base`

Search base for users

`username-attribute`

Unique identifier for a user that can be used for authentication

`group-object-class`

Object class for groups

`group-search-base`

Search base for groups

`group-member-attribute`

Attribute in group that associates users to group

`group-member-attribute-value`

User attributevalue that will be used in group entity to associate user to the group

`service-account-username`

User name of the directory administrator

service-account-password

Password of the directory administrator

Get Client Auth Configuration

```
ncli> authconfig { get-client-authentication-config }
```

Required arguments

None

Get the list of entity values for the specified entity type and the directory name.

```
ncli> authconfig { get-directory-values-by-type } name="name" directory-name="directory_name" username="username" password="password" entity-type="entity_type"
```

Required arguments

name

Name of the Directory

directory-name

Name of the directory.

username

User name.

password

Password for corresponding username.

entity-type

Type of the directory entity.

List Authentication Configuration

```
ncli> authconfig { list | ls }
```

Required arguments

None

List Authentication Configuration

```
ncli> authconfig { list-directory | ls-directory }[ name="name" ]
```

Required arguments

None

Optional arguments

name

Name of the Directory Service

Get the list of directory role mappings

```
ncli> authconfig { list-role-mappings | ls-role-mappings } name="name" [ role="role" ][ entity-type="entity_type" ]
```

Required arguments

name

Name

Optional arguments

`role`

Role

`entity-type`

Entity Type

Remove Authentication Configuration

```
ncli> authconfig { remove | rm }
```

Required arguments

None

Remove Directory Service Configuration

```
ncli> authconfig { remove-directory | rm-directory } name="name"
```

Required arguments

`name`

Name of the Directory

Comma-separated list of values to be removed from the existing directory role mapping

```
ncli> authconfig { remove-from-role-mapping-values } name="name" role="role"
entity-type="entity_type" entity-values="entity_values"
```

Required arguments

`name`

Name

`role`

Role

`entity-type`

Entity Type

`entity-values`

Values

Set Certificate Revocation information

```
ncli> authconfig { set-certificate-revocation }[ set-ocsp-
responder="set_ocsp_responder" ][ set-crl-uri="set_crl_uri" ][ set-
crl-refresh-interval="set_crl_refresh_interval" ][ set-crl-expiration-
interval="set_crl_expiration_interval" ]
```

Required arguments

None

Optional arguments

`set-ocsp-responder`

URI for OCSP responder

`set-crl-uri`

Comma separated list of CRL URI

`set-crl-refresh-interval`

Time between CRL refresh in seconds (default is 86400 seconds, 1 day)

`set-crl-expiration-interval`

Time interval, in seconds, after which the CRLs are expired if the DP has not been accessible (default is 604800 seconds, 7 days)

Test LDAP Connection and authentication

`ncli> authconfig { test-ldap-connection } directory-name="`*directory_name*`" password="`*password*`" username="`*username*`"`

Required arguments

`directory-name`

Directory name to test LDAP configuration.

`password`

Password to test LDAP configuration.

`username`

Username to test LDAP configuration.

Enable/Disable client authentication

`ncli> authconfig { update-client-authentication } enable-client-auth="`*{ true | false }*`"`

Required arguments

`enable-client-auth`

Enable client auth

Update service account (CAC)

`ncli> authconfig { update-service-account }[ enable="`*{ true | false }*`" ][ directory-name="`*directory_name*`" ][ username="`*username*`" ][ password="`*password*`" ]`

Required arguments

None

Optional arguments

`enable`

Enable Service Account (CAC)

`directory-name`

Service Account (CAC) Directory Name

`username`

Service Account (CAC) Username

`password`

Service Account (CAC) Password

`cluster`: Cluster

Description	A Nutanix Complete Cluster
Alias	

Operations

- Add public key to the cluster : `add-public-key`
- Add addresses to the list of name servers : `add-to-name-servers`
- Add addresses to the list of NTP servers : `add-to-ntp-servers`
- Clear configuration of SMTP Server used for transmitting alerts and report emails to Nutanix support : `clear-smtp-server`
- Edit the security params of a Cluster : `edit-cvm-security-params`
- Edit the hypervisor security compliance config of a Cluster : `edit-hypervisor-security-params`
- Edit params of a Cluster : `edit-params` | `edit-info`
- Edit the security params of a Multicluster : `edit-pcvm-security-params`
- Gets the Microsoft Azure Account details : `get-la-jolla-azure-info`
- Get the list of name servers : `get-name-servers`
- Get the list of NTP servers : `get-ntp-servers`
- Get params of a Cluster : `get-params` | `info`
- Get the security config for the Multicluster : `get-pcvm-security-config`
- Get remote support status on a Cluster : `get-remote-support-status`
- Get configuration of SMTP Server used for transmitting alerts and report emails to Nutanix support : `get-smtp-server`
- Get the list of public keys configured in the cluster : `list-public-keys` | `ls-public-keys`
- Remove all the public keys previously configured in the cluster : `remove-all-public-keys` | `rm-all-public-keys`
- Remove addresses from the list of name servers : `remove-from-name-servers`
- Remove addresses from the list of NTP servers : `remove-from-ntp-servers`
- Delete public key with the specified name from the cluster : `remove-public-key` | `rm-public-key`
- Send an email to test the SMTP Server configuration : `send-test-email`
- Sets the DVM configuration parameters for LaJolla cluster : `set-dvm-configuration`
- Set the redundancy state of the cluster : `set-redundancy-state`
- Set configuration of SMTP Server used for transmitting alert and report emails to Nutanix support : `set-smtp-server`
- Set the timezone of the Cluster : `set-timezone`
- Starts the CPS deployment in Lajolla cluster : `start-la-jolla-deployment`
- Start remote support on a Cluster : `start-remote-support`
- Get the status of all services in the cluster : `status`
- Stop remote support on a Cluster : `stop-remote-support`
- Get version of software running on a Cluster : `version`

Add public key to the cluster

`ncli> cluster { add-public-key } name="name" file-path="file_path"`

Required arguments

`name`

Name of the cluster public key

`file-path`

Full path of the public key file

Add addresses to the list of name servers

`ncli> cluster { add-to-name-servers }[ servers="servers" ][ name-servers="name_servers" ]`

Required arguments

None

Optional arguments

`servers`

Comma-separated list of IP addresses to be included in the name servers list

`name-servers`

Comma-separated list of IPV4/IPV6 addresses to be included in the name servers list

Add addresses to the list of NTP servers

`ncli> cluster { add-to-ntp-servers }[ servers="servers" ][ ntp-servers="ntp_servers" ]`

Required arguments

None

Optional arguments

`servers`

Comma-separated list of IP addresses/Host names to be included in the NTP servers list

`ntp-servers`

Comma-separated list of IPV4/IPV6/Hostname addresses to be included in the NTP servers list

Clear configuration of SMTP Server used for transmitting alerts and report emails to Nutanix support

`ncli> cluster { clear-smtp-server }`

Required arguments

None

Edit the security params of a Cluster

`ncli> cluster { edit-cvm-security-params }[enable-aide="{ true | false }"][enable-core="{ true | false }"][enable-high-strength-password="{ true | false }"][enable-banner="{ true | false }"][enable-snmpv3-only="{ true | false }"][schedule="schedule"][enable-kernel-mitigations="{ true | false }"][ssh-security-level="ssh_security_level"][enable-lock-status="{ true | false }"][enable-kernel-core="{ true | false }"]`

Required arguments

None

Optional arguments

`enable-aide`

Enable intrusion detection service.

`enable-core`

Enable user process core dumps.

`enable-high-strength-password`

Enable to set the minimum length for password to 14 and remember to 24 in PAM.

`enable-banner`

Enable DoD knowledge of consent banner for SSH and console logins.

`enable-snmpv3-only`

Enable to set the CVM host to either ignore or explicitly restrict v1 and v2 snmp based connections.

`schedule`

Set cron schedule to run Salt periodically.

`enable-kernel-mitigations`

Enable to mitigate some of the known CPU vulnerabilities which includes spectre, meltdown, etc. at cost of performance impact.

`ssh-security-level`

Different security levels for the nutanix user for ssh login to the Nutanix Cluster. It can be configured with one of the following level: 'default', 'limited' or 'restricted'

Default: default

`enable-lock-status`

Enable to lock the security config settings. If true, the config settings can not be edited by the user and a support call will need to be made to unlock this configuration.

`enable-kernel-core`

Enable kernel core dumps.

Edit the hypervisor security compliance config of a Cluster

`ncli> cluster { edit-hypervisor-security-params }[ enable-aide="`*{ true | false }*`" ][ enable-core="`*{ true | false }*`" ][ enable-high-strength-password="`*{ true | false }*`" ][ enable-banner="`*{ true | false }*`" ][ schedule="`*schedule*`" ]`

Required arguments

None

Optional arguments

`enable-aide`

Enable intrusion detection service.

`enable-core`

Enable user process core dumps.

enable-high-strength-password

Enable to set the minimum length for password to 14 and remember to 24 in PAM.

enable-banner

Enable DoD knowledge of consent banner for SSH and console logins.

schedule

Set cron schedule to run Salt periodically.

Edit params of a Cluster

```
ncli> cluster { edit-params | edit-info }[ new-name="new_name" ][ support-verbosity-level="support_verbosity_level" ][ external-ip-address="external_ip_address" ][ external-ip-addresses="external_ip_addresses" ][ cluster-fqdn="cluster_fqdn" ][ masquerading-ip-address="masquerading_ip_address" ][ masquerading-ip-addresses="masquerading_ip_addresses" ][ masquerading-port="masquerading_port" ][ cluster-usage-warning-alert-threshold-pct="cluster_usage_warning_alert_threshold_pct" ][ cluster-usage-critical-alert-threshold-pct="cluster_usage_critical_alert_threshold_pct" ][ external-data-services-ip-address="external_data_services_ip_address" ][ external-data-services-ip-addresses="external_data_services_ip_addresses" ][ enable-shadow-clones="{ true | false }" ][ enable-password-remote-login="{ true | false }" ][ enable-rf1-container="{ true | false }" ][ logon-name="logon_name" ][ password="password" ][ disable-degraded-node-monitoring="{ true | false }" ][ enable-on-disk-dedup="{ true | false }" ][ enable-common-criteria-mode="{ true | false }" ][ enable-rebuild-reservation="{ true | false }" ]
```

Required arguments

None

Optional arguments

new-name

Name of the cluster

support-verbosity-level

Verbosity level settings for populating support information

external-ip-address

External IP address to access the cluster. Set to '-' to clear the existing value

external-ip-addresses

External IP addresses to access the cluster. It accepts ipv4 and ipv6. Set to '-' to clear the existing value

cluster-fqdn

FQDN to access the cluster. Set to '-' to clear the existing value

masquerading-ip-address

Masquerading IP address to access the cluster. Set to '-' to clear the existing value

masquerading-ip-addresses

Masquerading IP address to access the cluster.It accepts ipv4 and ipv6. Set to '-' to clear the existing value

masquerading-port

Masquerading port to access the cluster. Set to '-' to clear the existing value

`cluster-usage-warning-alert-threshold-pct`

Cluster usage alert threshold percentage for warning.

`cluster-usage-critical-alert-threshold-pct`

Cluster usage alert threshold percentage for critical.

`external-data-services-ip-address`

External Data Services IP address of the cluster. Set to '-' to clear the existing value

`external-data-services-ip-addresses`

External Data Services IP address of the cluster.It accepts ipv4 and ipv6. Set to '-' to clear the existing value

`enable-shadow-clones`

Enable/Disable shadow clones

`enable-password-remote-login`

Enable/Disable password remote login to cluster

`enable-rf1-container`

Enable/Disable container with replication factor 1 support

`logon-name`

Logon name (domain\username) of a domain user/administrator account that has privileges to perform the operation

`password`

Password for the account specified by the logon account name

`disable-degraded-node-monitoring`

Whether to disable detection of partially degraded nodes.

`enable-on-disk-dedup`

Whether to enable on-disk deduplication for any new writes at cluster level. It might be disabled again when storage issue detected.

`enable-common-criteria-mode`

Enables Common Criteria mode on the cluster when set to 'true'. Disables when set to 'false'

`enable-rebuild-reservation`

Enable/Disable rebuild capacity reservation

Edit the security params of a Multicluster

`ncli> cluster { edit-pcvm-security-params }[ enable-aide="`*{ true | false }*`" ][ enable-core="`*{ true | false }*`" ][ enable-high-strength-password="`*{ true | false }*`" ][ enable-banner="`*{ true | false }*`" ][ enable-snmpv3-only="`*{ true | false }*`" ][ schedule="`*schedule*`" ]`

Required arguments

None

Optional arguments

`enable-aide`

Enable intrusion detection service.

`enable-core`

Enable user process core dumps.

`enable-high-strength-password`

Enable to set the minimum length for password to 14 and remember to 24 in PAM.

`enable-banner`

Enable DoD knowledge of consent banner for SSH and console logins.

`enable-snmpv3-only`

Enable to set the CVM host to either ignore or explicitly restrict v1 and v2 snmp based connections.

`schedule`

Set cron schedule to run Salt periodically.

Gets the Microsoft Azure Account details

```
ncli> cluster { get-la-jolla-azure-info }
```

Required arguments

None

Get the list of name servers

```
ncli> cluster { get-name-servers }
```

Required arguments

None

Get the list of NTP servers

```
ncli> cluster { get-ntp-servers }
```

Required arguments

None

Get params of a Cluster

```
ncli> cluster { get-params | info }
```

Required arguments

None

Get the security config for the Multicluster

```
ncli> cluster { get-pcvm-security-config }
```

Required arguments

None

Get remote support status on a Cluster

```
ncli> cluster { get-remote-support-status }
```

Required arguments

None

Get configuration of SMTP Server used for transmitting alerts and report emails to Nutanix support

```
ncli> cluster { get-smtp-server }
```

Required arguments

None

Get the list of public keys configured in the cluster

```
ncli> cluster { list-public-keys | ls-public-keys }[ name="name" ]
```

Required arguments

None

Optional arguments

`name`

Name of the cluster public key

Remove all the public keys previously configured in the cluster

```
ncli> cluster { remove-all-public-keys | rm-all-public-keys }
```

Required arguments

None

Remove addresses from the list of name servers

```
ncli> cluster { remove-from-name-servers }[ servers="servers" ][ name-servers="name_servers" ]
```

Required arguments

None

Optional arguments

`servers`

Comma-separated list of IP addresses to be removed from the name servers list

`name-servers`

Comma-separated list of IPV4/IPV6 addresses to be removed in the name servers list

Remove addresses from the list of NTP servers

```
ncli> cluster { remove-from-ntp-servers }[ servers="servers" ][ ntp-servers="ntp_servers" ]
```

Required arguments

None

Optional arguments

`servers`

Comma-separated list of IP addresses/Host names to be removed from the NTP servers list

`ntp-servers`

Comma-separated list of IPV4/IPV6/Hostname addresses to be removed in the NTP servers list

Delete public key with the specified name from the cluster

```
ncli> cluster { remove-public-key | rm-public-key } name="name"
```

Required arguments

`name`

Name of the cluster public key

Send an email to test the SMTP Server configuration

`ncli> cluster { send-test-email } recipient="`*recipient*`" subject="`*subject*`"`

Required arguments

`recipient`

Recipient of the test email

`subject`

Subject of the test email

Sets the DVM configuration parameters for LaJolla cluster

`ncli> cluster { set-dvm-configuration } ip-start="`*ip_start*`" ip-end="`*ip_end*`" ip-mask="`*ip_mask*`" gateway="`*gateway*`" company-name="`*company_name*`" [ license-key="`*license_key*`" ][ data-telemetry="`*data_telemetry*`" ]`

Required arguments

`ip-start`

IP Address range start.

`ip-end`

IP Address range end.

`ip-mask`

Subnet mask.

`gateway`

Default Gateway IP.

`company-name`

Company name setting dvm configuraiton

Optional arguments

`license-key`

License key of Microsoft

`data-telemetry`

True/False indicating whether you want your data to be sent to Microsoft.

Set the redundancy state of the cluster

`ncli> cluster { set-redundancy-state } desired-redundancy-factor="`*desired_redundancy_factor*`"`

Required arguments

`desired-redundancy-factor`

Desired redundancy factor of the cluster

Set configuration of SMTP Server used for transmitting alert and report emails to Nutanix support

`ncli> cluster { set-smtp-server }[ port="`*port*`" ][ username="`*username*`" ][ password="`*password*`" ][ security-mode="`*security_mode*`" ][ from-email-address="`*from_email_address*`" ][ address="`*address*`" ][ smtp-address="`*smtp_address*`" ]`

Required arguments

None

Optional arguments

`port`

Port number of the SMTP Server. By default, port 25 is used

`username`

Username to access the SMTP Server

`password`

Password to access the SMTP Server

`security-mode`

Security mode used by SMTP Server for data encryption and authentication. SMTP Server in Nutanix cluster can be configured with one of the following mode: 'none', 'ssl' or 'starttls'

Default: none

`from-email-address`

From email address to be used while sending emails (Set to '-' to clear the existing value)

`address`

Fully Qualified Domain Name(FQDN) or IPv4 address of the SMTP Server

`smtp-address`

Fully Qualified Domain Name(FQDN) or IPv6/IPv4 address of the SMTP Server

Set the timezone of the Cluster

`ncli> cluster { set-timezone } timezone="`*timezone*`"`

Required arguments

`timezone`

Timezone of the cluster

Starts the CPS deployment in Lajolla cluster

`ncli> cluster { start-la-jolla-deployment } ip-start="`*ip_start*`" ip-end="`*ip_end*`" ip-mask="`*ip_mask*`" gateway="`*gateway*`" company-name="`*company_name*`" username="`*username*`" password="`*password*`" storage_admin_password="`*storage_admin_password*`" azure_onboarding_enabled="`*azure_onboarding_enabled*`" [ license-key="`*license_key*`" ][ azure_subscription_name="`*azure_subscription_name*`" ][ azure_subscription_id="`*azure_subscription_id*`" ][ azure_site_recovery_region="`*azure_site_recovery_region*`" ][ azure_ops_insights_region="`*azure_ops_insights_region*`" ]`

Required arguments

`ip-start`

IP Address range start.

`ip-end`

IP Address range end.

`ip-mask`

Subnet mask.

gateway

Default Gateway IP.

company-name

Company name setting dvm configuraiton

username

Active Directory User name

password

Active directory password

storage_admin_password

Admin password for the cluster

azure_onboarding_enabled

Set to true or false based on whether Azure info details are provided or not

Optional arguments

license-key

License key of Microsoft

azure_subscription_name

Subscription name of Microsoft Account

azure_subscription_id

Subscription Id of Microsoft Account

azure_site_recovery_region

Site recovery region of Microsoft

azure_ops_insights_region

Ops insight region of Microsoft

Start remote support on a Cluster

```
ncli> cluster { start-remote-support }[ duration="duration" ]
```

Required arguments

None

Optional arguments

duration

Time (in minutes) to keep remote support enabled

Get the status of all services in the cluster

```
ncli> cluster { status }
```

Required arguments

None

Stop remote support on a Cluster

```
ncli> cluster { stop-remote-support }[ duration="duration" ]
```

Required arguments

None

Optional arguments

`duration`

Time (in minutes) to keep remote support disabled

Get version of software running on a Cluster

```
ncli> cluster { version }
```

Required arguments

None

`container`: Storage Container

Description	A Storage Container is a container for virtual disks
Alias	`ctrstorage-container`
Operations	• List Storage Containers : `list` | `ls` • Get stats data for Storage Containers : `list-stats` | `ls-stats`

List Storage Containers

```
ncli> container { list | ls }[ id="id" ][ name="name" ]
```

Required arguments

None

Optional arguments

`id`

ID of the Storage Container

`name`

Name of the Storage Container

Get stats data for Storage Containers

```
ncli> container { list-stats | ls-stats }[ id="id" ][ name="name" ]
```

Required arguments

None

Optional arguments

`id`

ID of the Storage Container

`name`

Name of the Storage Container

`data-at-rest-encryption`: Data At Rest Encryption

Description	Manage data-at-rest-encryption related operations
Alias	

Operations

- Generates a backup file for all the software encryption keys used on this cluster : `backup-software-encryption-keys`
- Generates new passwords for each container with software encryption enabled and makes a new password for the metadata encryption key : `rekey-software-encryption-keys`

Generates a backup file for all the software encryption keys used on this cluster

```
ncli> data-at-rest-encryption { backup-software-encryption-keys } file-path="file_path" [ password="password" ]
```

Required arguments

`file-path`

Path to store the backup

Optional arguments

`password`

Password to use to encrypt the backup file

Generates new passwords for each container with software encryption enabled and makes a new password for the metadata encryption key

```
ncli> data-at-rest-encryption { rekey-software-encryption-keys }
```

Required arguments

None

`data-at-rest-encryption-certificate`: Data At Rest Encryption Certificate

Description Manage data-at-rest-encryption related digital certificates

Alias

Operations

- Get certification information : `get-csr-information`
- Get the list of ca certificates stored in the cluster : `list-ca-certificates` | `ls-ca-certificates`
- Delete ca certificate with the specified name from the cluster : `remove-ca-certificate` | `rm-ca-certificate`
- Delete certificate installed on a node for a key management server : `remove-cvm-certificate` | `rm-cvm-certificate`

Get certification information

```
ncli> data-at-rest-encryption-certificate { get-csr-information }
```

Required arguments

None

Get the list of ca certificates stored in the cluster

```
ncli> data-at-rest-encryption-certificate { list-ca-certificates | ls-ca-certificates }
```

Required arguments

None

Delete ca certificate with the specified name from the cluster

```
ncli> data-at-rest-encryption-certificate { remove-ca-certificate | rm-ca-certificate } ca-name="ca_name"
```

Required arguments

`ca-name`

Certificate Authority name

Delete certificate installed on a node for a key management server

```
ncli> data-at-rest-encryption-certificate { remove-cvm-certificate | rm-cvm-certificate } host-id="host_id" key-management-server-name="key_management_server_name"
```

Required arguments

`host-id`

Id of the node on which certificate is installed

`key-management-server-name`

Key management Server name for which certificate is installed

`disk`: Physical Disk

Description	A Physical Disk
Alias	
Operations	• List Physical Disks : `list` | `ls` • List Physical Disks that are not assigned to any Storage Pool : `list-free` | `ls-free` • Get stats data for Physical Disks : `list-stats` | `ls-stats`

List Physical Disks

```
ncli> disk { list | ls }[ id="id" ][ sp-id="sp_id" ][ sp-name="sp_name" ][ tier-name="tier_name" ]
```

Required arguments

None

Optional arguments

`id`

ID of the Physical Disk

`sp-id`

ID of the Storage Pool for the Physical Disk

`sp-name`

Name of the Storage Pool for the Physical Disk

`tier-name`

Corresponding Storage Tier

List Physical Disks that are not assigned to any Storage Pool

```
ncli> disk { list-free | ls-free }
```

Required arguments

None

Get stats data for Physical Disks

```
ncli> disk { list-stats | ls-stats }[ id="id" ][ sp-id="sp_id" ][ tier-name="tier_name" ]
```

Required arguments

None

Optional arguments

`id`

ID of the Physical Disk

`sp-id`

ID of the Storage Pool for the Physical Disk

`tier-name`

Corresponding Storage Tier

`events`: Event

Description	An Event
Alias	`event`
Operations	• List history of Events : `history` • List of unacknowledged Events : `list` | `ls`

List history of Events

```
ncli> events { history } duration="duration" [ acknowledged="acknowledged" ][ max-events="max_events" ]
```

Required arguments

`duration`

Duration (in days) for getting the history of Events

Optional arguments

`acknowledged`

Acknowledged Events?

`max-events`

Maximum number of Events to fetch

Default: 100

List of unacknowledged Events

```
ncli> events { list | ls }[ max-events="max_events" ]
```

Required arguments

None

Optional arguments

`max-events`

Maximum number of Events to fetch

Default: 100

`host`: Physical Host

Description	A Physical Host hosts Virtual Machines
Alias	
Operations	• List Physical Hosts : `list` | `ls` • Get stats data for Physical Host : `list-stats` | `ls-stats`

List Physical Hosts

```
ncli> host { list | ls }[ id="id" ]
```

Required arguments

None

Optional arguments

`id`

ID of the Physical Host

Get stats data for Physical Host

```
ncli> host { list-stats | ls-stats }[ id="id" ]
```

Required arguments

None

Optional arguments

`id`

ID of the Physical Host

`http-proxy`: HTTP Proxy

Description	An HTTP Proxy
Alias	`proxy`
Operations	• Create a new HTTP Proxy : `add` • Add HTTP Proxy whitelist entry : `add-to-whitelist` • Delete HTTP Proxy whitelist entry : `delete-from-whitelist` • Edit an HTTP Proxy : `edit` | `update` • Get HTTP Proxy whitelist : `get-whitelist` • List HTTP Proxies : `list` | `ls` • Remove an HTTP Proxy : `remove` | `rm`

Create a new HTTP Proxy

```
ncli> http-proxy { add } name="name" port="port" [ address="address" ][ host="host"
][ username="username" ][ password="password" ][ proxy-types="proxy_types" ]
```

Required arguments

`name`

Name of the HTTP Proxy

`port`

Port number of the HTTP Proxy

Optional arguments

`address`

Address of the HTTP Proxy

`host`

Address of the HTTP Proxy

`username`

Username to access the HTTP Proxy

`password`

Password to access the HTTP Proxy

`proxy-types`

Proxy types to send applicable traffic

Add HTTP Proxy whitelist entry

`ncli> http-proxy { add-to-whitelist } target-type="`*target_type*`" target="`*target*`"`

Required arguments

`target-type`

Type of the target. Values can be {IPV4_ADDRESS, IPV6_ADDRESS, IPV4_NETWORK_MASK, DOMAIN_NAME_SUFFIX, HOST_NAME}

`target`

Http proxy whitelist target. If the target-type is IPV4_NETWORK_MASK, specify the netmask using traditional notation (for example, 172.16.0.0/255.240.0.0), not CIDR notation.

Delete HTTP Proxy whitelist entry

`ncli> http-proxy { delete-from-whitelist } target="`*target*`"`

Required arguments

`target`

Http proxy whitelist target. If the target-type is IPV4_NETWORK_MASK, specify the netmask using traditional notation (for example, 172.16.0.0/255.240.0.0), not CIDR notation.

Edit an HTTP Proxy

`ncli> http-proxy { edit | update } name="`*name*`" [ address="`*address*`" ][ host="`*host*`" ][ username="`*username*`" ][ password="`*password*`" ][ port="`*port*`" ]`

Required arguments

`name`

Name of the HTTP Proxy

Optional arguments

`address`

Address of the HTTP Proxy

host

Address of the HTTP Proxy

username

Username to access the HTTP Proxy

password

Password to access the HTTP Proxy

port

Port number of the HTTP Proxy

Get HTTP Proxy whitelist

```
ncli> http-proxy { get-whitelist }
```

Required arguments

None

List HTTP Proxies

```
ncli> http-proxy { list | ls }[ name="name" ]
```

Required arguments

None

Optional arguments

name

Name of the HTTP Proxy

Remove an HTTP Proxy

```
ncli> http-proxy { remove | rm } name="name"
```

Required arguments

name

Name of an HTTP Proxy

`license`: License

Description	License for a Nutanix cluster
Alias	
Operations	• Apply a license file to the cluster : `apply-license` • Download cluster info as a file : `download-cluster-info` • Get cluster info from the cluster : `generate-cluster-info` • Read allowances for features as listed in the license : `get-allowances` • Read license file from the cluster : `get-license`

Apply a license file to the cluster

```
ncli> license { apply-license } license-file="license_file"
```

Required arguments

`license-file`

License file

Download cluster info as a file

`ncli> license { download-cluster-info } file-path="`*file_path*`" [ csf-version="`*csf_version*`" ]`

Required arguments

`file-path`

File path

Optional arguments

`csf-version`

CSF Version

Default: 1

Get cluster info from the cluster

`ncli> license { generate-cluster-info }[ csf-version="`*csf_version*`" ]`

Required arguments

None

Optional arguments

`csf-version`

CSF Version

Default: 1

Read allowances for features as listed in the license

`ncli> license { get-allowances }[ feature-name="`*feature_name*`" ]`

Required arguments

None

Optional arguments

`feature-name`

Feature name

Read license file from the cluster

`ncli> license { get-license }[ showAllClusters="`*showAllClusters*`" ]`

Required arguments

None

Optional arguments

`showAllClusters`

Show PE Licenses

Default: false

`managementserver`: Management Server

Description	An infrastructure management server such as VCenter

Alias

Operations

- Create and register a management server extension for Nutanix : `register`
- Unregister the management server extension for Nutanix : `unregister`

Create and register a management server extension for Nutanix.

`ncli> managementserver { register } ip-address="`*ip_address*`" port="`*port*`" admin-username="`*admin_username*`" admin-password="`*admin_password*`"`

Required arguments

`ip-address`

Address of the management server.

`port`

Port of the management server.

`admin-username`

Administrator username of the management server.

`admin-password`

Administrator password of the management server.

Unregister the management server extension for Nutanix.

`ncli> managementserver { unregister } ip-address="`*ip_address*`" port="`*port*`" admin-username="`*admin_username*`" admin-password="`*admin_password*`"`

Required arguments

`ip-address`

Address of the management server.

`port`

Port of the management server.

`admin-username`

Administrator username of the management server.

`admin-password`

Administrator password of the management server.

`multicluster`: Multicluster

Description A Nutanix Management Console to manage multiple clusters

Alias

Operations

- Get cluster state : `get-cluster-state`

Get cluster state

`ncli> multicluster { get-cluster-state }[ cluster-id="`*cluster_id*`" ]`

Required arguments

None

Optional arguments

`cluster-id`

Id of the cluster

`network`: Network

Description	Network specific commands
Alias	`net`
Operations	• List all host interfaces : `list-host-nics` • List all host virtual interfaces : `list-host-vnics` • List all VM virtual Nics : `list-vm-nics`

List all host interfaces

`ncli> network { list-host-nics } host-id="`*host_id*`" [ pnic-id="`*pnic_id*`" ]`

Required arguments

`host-id`

ID of the Host

Optional arguments

`pnic-id`

ID of Host Nic

List all host virtual interfaces

`ncli> network { list-host-vnics } host-id="`*host_id*`" [ host-vnic-id="`*host_vnic_id*`" ]`

Required arguments

`host-id`

ID of the Host

Optional arguments

`host-vnic-id`

ID of Host Virtual Nic

List all VM virtual Nics

`ncli> network { list-vm-nics } vm-id="`*vm_id*`" [ vnic-id="`*vnic_id*`" ]`

Required arguments

`vm-id`

ID of the VM

Optional arguments

`vnic-id`

ID of VM Nic

`pulse-config`: Pulse Configuration

Description	Configuration information used for Pulse setup
Alias	

Operations

- Update Pulse Configuration : `edit` | `update`
- Update Pulse Configuration on registered PE's : `fanout_update`
- List Pulse Configuration : `list` | `ls`

Update Pulse Configuration

```
ncli> pulse-config { edit | update }[ enable="{ true | false }" ][ enable-default-nutanix-email="{ true | false }" ][ support-verbosity-level="support_verbosity_level" ][ email-contacts="email_contacts" ]
```

Required arguments

None

Optional arguments

`enable`

Enable Pulse emails

`enable-default-nutanix-email`

Enable Pulse default Nutanix email

`support-verbosity-level`

List Pulse Configuration

`email-contacts`

Comma-separated list of emails to be used while sending Pulse info. Set to '-' to clear all the existing emails.

Update Pulse Configuration on registered PE's

```
ncli> pulse-config { fanout_update } enable_on_pe="{ true | false }" [ cluster_list="cluster_list" ]
```

Required arguments

`enable_on_pe`

Enable/Disable Pulse on all registered PE's

Optional arguments

`cluster_list`

Comma-separated list of registered cluster uuid's on which pulse enable/disable operation is to be executed

List Pulse Configuration

```
ncli> pulse-config { list | ls }
```

Required arguments

None

`rsyslog-config`: TLS configuration for RSyslog service

Description	TLS configuration for RSyslog service
Alias	

Operations

- Create RSyslog Server Module : `create-module` | `add-module`
- Create RSyslog Configuration : `create-server` | `add-server`
- Remove RSyslog Server Module : `delete-module` | `remove-module`
- Remove RSyslog Server : `delete-server` | `remove-server`
- Update RSyslog Server : `edit-server` | `update-server`
- Returns the status of RSyslog service : `get-status`
- List RSyslog Servers : `list` | `ls`
- List RSyslog Server Modules : `list-modules` | `ls-modules`
- List RSyslog Servers : `list-servers` | `ls-servers`
- Sets the status of RSyslog service : `set-status`

Create RSyslog Server Module

`ncli> rsyslog-config { create-module | add-module } server-name="`*server_name*`" module-name="`*module_name*`" level="`*level*`" [ include-monitor-logs="`*{ true | false }*`" ]`

Required arguments

`server-name`

Log Server Name

`module-name`

Name of the RSyslog Server Module

`level`

Log level for RSyslog Server Module

Optional arguments

`include-monitor-logs`

Include monitor logs for the given RSyslog Server Module

Create RSyslog Configuration

`ncli> rsyslog-config { create-server | add-server } name="`*name*`" ip-address="`*ip_address*`" port="`*port*`" [ network-protocol="`*network_protocol*`" ][ relp-enabled="`*relp_enabled*`" ]`

Required arguments

`name`

Name of the RSyslog Server

`ip-address`

Ip address of the RSyslog Server

`port`

port number

Optional arguments

`network-protocol`

Protocol for RSyslog server configuration

`relp-enabled`

Reliable Event Logging Protocol option

Remove RSyslog Server Module

`ncli> rsyslog-config { delete-module | remove-module } server-name="`*server_name*`" module-name="`*module_name*`"`

Required arguments

`server-name`

Name of the RSyslog Server

`module-name`

Name of the RSyslog Server Module

Remove RSyslog Server

`ncli> rsyslog-config { delete-server | remove-server } name="`*name*`"`

Required arguments

`name`

Name of the log server

Update RSyslog Server

`ncli> rsyslog-config { edit-server | update-server } name="`*name*`" [ ip-address="`*ip_address*`" ][ port="`*port*`" ][ network-protocol="`*network_protocol*`" ][ relp-enabled="`*relp_enabled*`" ]`

Required arguments

`name`

Name of the RSyslog Server

Optional arguments

`ip-address`

Ip address of the RSyslog Server

`port`

port number

`network-protocol`

Protocol for RSyslog server configuration

`relp-enabled`

Reliable Event Logging Protocol option

Returns the status of RSyslog service

`ncli> rsyslog-config { get-status }`

Required arguments

None

List RSyslog Servers

`ncli> rsyslog-config { list | ls }`

Required arguments

None

List RSyslog Server Modules

`ncli> rsyslog-config { list-modules | ls-modules } server-name="`*server_name*`"`

Required arguments

`server-name`

Name of the log server

List RSyslog Servers

```
ncli> rsyslog-config { list-servers | ls-servers }
```

Required arguments

None

Sets the status of RSyslog service

```
ncli> rsyslog-config { set-status } enable="{ true | false }"
```

Required arguments

`enable`

Enable RSyslog Status

`snmp`: SNMP

Description	An SNMP agent
Alias	
Operations	• Add a transport to the list of snmp transports : `add-transport` • Add a trap sink to the list of trap sinks : `add-trap` • Add an snmp user along with its authentication and privacy keys : `add-user` • Edit one of the trap sinks from the list of trap sinks : `edit-trap` | `update-trap` • Modify the authentication and encrytption information of an existing snmp user : `edit-user` | `update-user` • Returns the status of the snmp service : `get-status` • List all the transports specified for the snmp agent : `list-transports` | `ls-transports` • List all the configured trap sinks along with their user information : `list-traps` | `ls-traps` • Lists all the snmp users along with their properties like authentication and privacy information : `list-users` | `ls-users` • Remove a transport from the list of snmp transports : `remove-transport` | `delete-transport` • Remove a trap from the list of snmp traps : `remove-trap` | `delete-trap` • Remove a user from the list of snmp users : `remove-user` | `delete-user` • Sets the status of the snmp service : `set-status`

Add a transport to the list of snmp transports. Each transport is a protocol:port pair

```
ncli> snmp { add-transport } protocol="protocol" port="port"
```

Required arguments

`protocol`

Protocol for the snmp agent or trap sink. Currently supported protocols are UDP, TCP and UDP_6

`port`

Port number on which an snmp agent listens for requests or on which a trap sink is waiting traps

Add a trap sink to the list of trap sinks. Each trap sink is a combination of trap sink address, username and authentication information

```
ncli> snmp { add-trap } address="address" [ username="username" ][ port="port" ][ protocol="protocol" ][ version="version" ][ community="community" ][ engine-id="engine_id" ][ inform="inform" ]
```

Required arguments

`address`

Address of an snmp trap sink. This should be an IP address or FQDN

Optional arguments

`username`

Identity of an snmp user. It is required for version snmpv3. It is not used for version snmpv2c.

`port`

Port number on which an snmp agent listens for requests or on which a trap sink is waiting traps

`protocol`

Protocol for the snmp agent or trap sink. Currently supported protocols are UDP, TCP and UDP_6

`version`

SNMP version [snmpv2c, snmpv3]

`community`

SNMP community string. Used for snmpv2c only. If not set, default to "public"

`engine-id`

Engine id of the snmp trap sink. This must be a hex string starting with 0x. It is set for snmpv3, not used for snmpv2c.

`inform`

Flag that specifies whether a trap sink is actually an inform sink

Default: false

Add an snmp user along with its authentication and privacy keys

```
ncli> snmp { add-user } username="username" auth-key="auth_key" [ auth-type="auth_type" ][ priv-key="priv_key" ][ priv-type="priv_type" ]
```

Required arguments

`username`

Identity of an snmp user. It is required for version snmpv3. It is not used for version snmpv2c.

`auth-key`

Authentication key for an snmp user

Optional arguments

`auth-type`

Authentication type for snmp user. Can be SHA

`priv-key`

Encryption key for an snmp user

`priv-type`

Encryption type for an snmp user. Can be AES

Edit one of the trap sinks from the list of trap sinks. Editable properties are username, authentication and privacy settings and protocol

```
ncli> snmp { edit-trap | update-trap } address="address" [ port="port" ][
protocol="protocol" ][ version="version" ][ community="community" ][ engine-
id="engine_id" ][ inform="inform" ][ username="username" ]
```

Required arguments

`address`

Address of an snmp trap sink. This should be an IP address or FQDN

Optional arguments

`port`

Port number on which an snmp agent listens for requests or on which a trap sink is waiting traps

`protocol`

Protocol for the snmp agent or trap sink. Currently supported protocols are UDP, TCP and UDP_6

`version`

SNMP version [snmpv2c, snmpv3]

`community`

SNMP community string. Used for snmpv2c only. If not set, default to "public"

`engine-id`

Engine id of the snmp trap sink. This must be a hex string starting with 0x. It is set for snmpv3, not used for snmpv2c.

`inform`

Flag that specifies whether a trap sink is actually an inform sink

`username`

Identity of an snmp user. It is required for version snmpv3. It is not used for version snmpv2c.

Modify the authentication and encrytption information of an existing snmp user

```
ncli> snmp { edit-user | update-user } username="username" [ auth-key="auth_key"
][ auth-type="auth_type" ][ priv-key="priv_key" ][ priv-type="priv_type" ]
```

Required arguments

`username`

Identity of an snmp user. It is required for version snmpv3. It is not used for version snmpv2c.

Optional arguments

`auth-key`

Authentication key for an snmp user

`auth-type`

Authentication type for snmp user. Can be SHA

`priv-key`

Encryption key for an snmp user

`priv-type`

Encryption type for an snmp user. Can be AES

Returns the status of the snmp service

```
ncli> snmp { get-status }
```

Required arguments

None

List all the transports specified for the snmp agent. Each transport is a protocol:port pair

```
ncli> snmp { list-transports | ls-transports }
```

Required arguments

None

List all the configured trap sinks along with their user information.

```
ncli> snmp { list-traps | ls-traps }[ address="address" ]
```

Required arguments

None

Optional arguments

`address`

Address of an snmp trap sink. This should be an IP address or FQDN

Lists all the snmp users along with their properties like authentication and privacy information

```
ncli> snmp { list-users | ls-users }[ username="username" ]
```

Required arguments

None

Optional arguments

`username`

Identity of an snmp user. It is required for version snmpv3. It is not used for version snmpv2c.

Remove a transport from the list of snmp transports

```
ncli> snmp { remove-transport | delete-transport } protocol="protocol" port="port"
```

Required arguments

`protocol`

Protocol for the snmp agent or trap sink. Currently supported protocols are UDP, TCP and UDP_6

`port`

Port number on which an snmp agent listens for requests or on which a trap sink is waiting traps

Remove a trap from the list of snmp traps

`ncli> snmp { remove-trap | delete-trap } address="`*address*`"`

Required arguments

`address`

Address of an snmp trap sink. This should be an IP address or FQDN

Remove a user from the list of snmp users

`ncli> snmp { remove-user | delete-user } username="`*username*`"`

Required arguments

`username`

Identity of an snmp user. It is required for version snmpv3. It is not used for version snmpv2c.

Sets the status of the snmp service

`ncli> snmp { set-status } enable="`*{ true | false }*`"`

Required arguments

`enable`

Enable or disable snmp agent on a cluster

`software`: Software

Description	NOS Software Release
Alias	
Operations	• Toggle automatic download of a Software : `automatic-download` • Download a Software : `download` • List Software : `list` | `ls` • Pause Downloading / Uploading a Software : `pause` • Delete a Software : `remove` | `rm` | `delete` • Upload a Software : `upload`

Toggle automatic download of a Software

`ncli> software { automatic-download }[ enable="`*{ true | false }*`" ]`

Required arguments

None

Optional arguments

`enable`

Enable/Disable automatic downloads

Download a Software

`ncli> software { download } name="name" software-type="software_type"`

Required arguments

`name`

Name of the software

`software-type`

Type of the software (NOS | HYPERVISOR | FIRMWARE_DISK | NCC | FILES | FILE_SERVER | PRISM_CENTRAL_DEPLOY | PRISM_CENTRAL | FILE_ANALYTICS)

List Software

`ncli> software { list | ls }[ name="name" ][ software-type="software_type" ]`

Required arguments

None

Optional arguments

`name`

Name of the software

`software-type`

Type of the software (NOS | HYPERVISOR | FIRMWARE_DISK | NCC | FILES | FILE_SERVER | PRISM_CENTRAL_DEPLOY | PRISM_CENTRAL | FILE_ANALYTICS)

Pause Downloading / Uploading a Software

`ncli> software { pause } name="name" software-type="software_type"`

Required arguments

`name`

Name of the software

`software-type`

Type of the software (NOS | HYPERVISOR | FIRMWARE_DISK | NCC | FILES | FILE_SERVER | PRISM_CENTRAL_DEPLOY | PRISM_CENTRAL | FILE_ANALYTICS)

Delete a Software

`ncli> software { remove | rm | delete } name="name" software-type="software_type"`

Required arguments

`name`

Name of the software

`software-type`

Type of the software (NOS | HYPERVISOR | FIRMWARE_DISK | NCC | FILES | FILE_SERVER | PRISM_CENTRAL_DEPLOY | PRISM_CENTRAL | FILE_ANALYTICS)

Upload a Software

`ncli> software { upload } file-path="file_path" software-type="software_type" [ hypervisor-type="hypervisor_type" ][ meta-file-path="meta_file_path" ]`

Required arguments

`file-path`

Path to the software to be uploaded

`software-type`

Type of the software (NOS | HYPERVISOR | FIRMWARE_DISK | NCC | FILES | FILE_SERVER | PRISM_CENTRAL_DEPLOY | PRISM_CENTRAL | FILE_ANALYTICS)

Optional arguments

`hypervisor-type`

Type of the Hypervisor

`meta-file-path`

Path to the metadata file of the software to be uploaded

`ssl-certificate`: SSL Certificate

Description	Manage SSL certificates
Alias	
Operations	• Import SSL Certificate, key and CA certificate or chain file : `import` • Change password for pfx file : `ssl-certificate-change-pfx-file-password` • Generates SSL Certificate with cipher Strength 2048 bits and replaces the existing certificate : `ssl-certificate-generate`

Import SSL Certificate, key and CA certificate or chain file. This import replaces the existing certificate

```
ncli> ssl-certificate { import } certificate-path="certificate_path" cacertificate-path="cacertificate_path" key-path="key_path" key-type="key_type"
```

Required arguments

`certificate-path`

Path of the SSL certificate

`cacertificate-path`

Path of the CA certificate or chain file

`key-path`

Path of the private key

`key-type`

Type of Private key. Must be either RSA_2048 or ECDSA_256 or ECDSA_384 or ECDSA_521

Change password for pfx file

```
ncli> ssl-certificate { ssl-certificate-change-pfx-file-password }
```

Required arguments

None

Generates SSL Certificate with cipher Strength 2048 bits and replaces the existing certificate

```
ncli> ssl-certificate { ssl-certificate-generate }
```

Required arguments

None

`storagepool`: Storage Pool

Description	A Pool of Physical Disks
Alias	`sp`
Operations	• List Storage Pools : `list` | `ls` • Get stats data for Storage Pools : `list-stats` | `ls-stats`

List Storage Pools

```
ncli> storagepool { list  | ls }[ id="id" ][ name="name" ]
```

Required arguments

None

Optional arguments

`id`

ID of the Storage Pool

`name`

Name of the Storage Pool

Get stats data for Storage Pools

```
ncli> storagepool { list-stats  | ls-stats }[ id="id" ][ name="name" ]
```

Required arguments

None

Optional arguments

`id`

ID of the Storage Pool

`name`

Name of the Storage Pool

`tag`: Tag

Description	A tag that groups entities
Alias	
Operations	• Create a new tag : `create` • Show an individual tag's details : `get` • Get tagged entities : `get-entities` • List all tags with filtering : `list` | `ls` • Remove a tag : `remove` • Update a tag : `update`

Create a new tag

`ncli> tag { create } name="`*name*`" entity-type="`*entity_type*`" [ uuid="`*uuid*`" ][ description="`*description*`" ]`

Required arguments

`name`

The friendly name for this tag.

`entity-type`

The entity type that the tag has been tagged with.

Optional arguments

`uuid`

The id for this tag.

`description`

The description for this tag.

Show an individual tag's details

`ncli> tag { get } uuid="`*uuid*`"`

Required arguments

`uuid`

UUID for tag requested.

Get tagged entities

`ncli> tag { get-entities } entity-type="`*entity_type*`" tag-ids="`*tag_ids*`"`

Required arguments

`entity-type`

The entity type

`tag-ids`

The list of tag uuids

List all tags with filtering

`ncli> tag { list | ls }[ filter-criteria="`*filter_criteria*`" ]`

Required arguments

None

Optional arguments

`filter-criteria`

Filtering on attributes of the tag.

Remove a tag

`ncli> tag { remove } uuid="`*uuid*`"`

Required arguments

`uuid`

UUID for tag to remove.

Update a tag

`ncli> tag { update } tag-uuid="`*tag_uuid*`" [ name="`*name*`" ][ description="`*description*`" ][ entity-type="`*entity_type*`" ]`

Required arguments

`tag-uuid`

Uuid of tag to update

Optional arguments

`name`

The friendly name for this tag.

`description`

The description for this tag.

`entity-type`

The entity type that the tag has been tagged with.

`user`: User

Description A User

Alias

Operations

- Change the password of a User : `change-password`
- Add a new User : `create` | `add`
- Delete a User : `delete` | `remove` | `rm`
- Disable a User : `disable`
- Edit a User : `edit` | `update`
- Enable a User : `enable`
- Get a list of all users who are currently logged in to the system along with their IP Addresses and browser details : `get-logged-in-users` | `get-logged-in-users`
- Grant backup administration role to a User : `grant-backup-admin-role`
- Grant cluster administration role to a User : `grant-cluster-admin-role`
- Grant user administration role to a User : `grant-user-admin-role`
- List Users : `list` | `ls`
- Reset the password of a User : `reset-password`
- Revoke backup administration role from a User : `revoke-backup-admin-role`
- Revoke cluster administration role from a User : `revoke-cluster-admin-role`
- Revoke user administration role from a User : `revoke-user-admin-role`
- Show profile of current User : `show-profile`

Change the password of a User

`ncli> user { change-password } current-password="`*current_password*`" new-password="`*new_password*`"`

Required arguments

`current-password`

Current password of the user

new-password

New password of the user

Add a new User

```
ncli> user { create | add } user-name="user_name" user-password="user_password"
first-name="first_name" last-name="last_name" email-id="email_id" [ middle-
initial="middle_initial" ]
```

Required arguments

user-name

User name of the user

user-password

Password of the user

first-name

First name of the user

last-name

Last name of the user

email-id

Email address of the user

Optional arguments

middle-initial

Middle Initial of the user

Delete a User

```
ncli> user { delete | remove | rm } user-name="user_name"
```

Required arguments

user-name

User name of the user

Disable a User

```
ncli> user { disable } user-name="user_name"
```

Required arguments

user-name

User name of the user

Edit a User

```
ncli> user { edit | update } user-name="user_name" [ first-name="first_name" ][
last-name="last_name" ][ middle-initial="middle_initial" ][ email-id="email_id" ]
```

Required arguments

user-name

User name of the user

Optional arguments

first-name

First name of the user

last-name

Last name of the user

`middle-initial`

Middle Initial of the user

`email-id`

Email address of the user

Enable a User

```
ncli> user { enable } user-name="user_name"
```

Required arguments

`user-name`

User name of the user

Get a list of all users who are currently logged in to the system along with their IP Addresses and browser details

```
ncli> user { get-logged-in-users | get-logged-in-users }
```

Required arguments

None

Grant backup administration role to a User

```
ncli> user { grant-backup-admin-role } user-name="user_name"
```

Required arguments

`user-name`

User name of the user

Grant cluster administration role to a User

```
ncli> user { grant-cluster-admin-role } user-name="user_name"
```

Required arguments

`user-name`

User name of the user

Grant user administration role to a User

```
ncli> user { grant-user-admin-role } user-name="user_name"
```

Required arguments

`user-name`

User name of the user

List Users

```
ncli> user { list | ls }[ user-name="user_name" ]
```

Required arguments

None

Optional arguments

`user-name`

User name of the user

Reset the password of a User

```
ncli> user { reset-password } user-name="user_name" password="password"
```

Required arguments

`user-name`

User name of the user

`password`

password

Revoke backup administration role from a User

`ncli> user { revoke-backup-admin-role } user-name="`*user_name*`"`

Required arguments

`user-name`

User name of the user

Revoke cluster administration role from a User

`ncli> user { revoke-cluster-admin-role } user-name="`*user_name*`"`

Required arguments

`user-name`

User name of the user

Revoke user administration role from a User

`ncli> user { revoke-user-admin-role } user-name="`*user_name*`"`

Required arguments

`user-name`

User name of the user

Show profile of current User

`ncli> user { show-profile }`

Required arguments

None

`virtualmachine`: Virtual Machine

Description	A Virtual Machine
Alias	`vm`
Operations	• Attach a disk from file level restore capable snapshot to a VM : `attach-flr-disk` • Detach a file level restore disk from a VM : `detach-flr-disk` • List Virtual Machine : `list` | `ls` • Get all file level restore capable snapshots attached to a VM : `list-attached-flr-snapshots` • Get file level restore capable snapshots of a VM : `list-flr-snapshots` | `ls-flr-snaps` • Get snapshots of a VM : `list-snapshots` | `ls-snaps` • Get stats data for Virtual Machine : `list-stats` | `ls-stats`

Attach a disk from file level restore capable snapshot to a VM

`ncli> virtualmachine { attach-flr-disk } vm-id="`*vm_id*`" snap-id="`*snap_id*`" disk-label="`*disk_label*`"`

Required arguments

`vm-id`

ID of the Virtual Machine

`snap-id`

Id of the Snapshot

`disk-label`

Label for disk to be attached

Detach a file level restore disk from a VM

`ncli> virtualmachine { detach-flr-disk } vm-id="`*vm_id*`" attached-disk-label="`*attached_disk_label*`"`

Required arguments

`vm-id`

ID of the Virtual Machine

`attached-disk-label`

Attached disk label

List Virtual Machine

`ncli> virtualmachine { list | ls }[ name="`*name*`" ][ id="`*id*`" ][ host-id="`*host_id*`" ][ get-unprotected-cbr-vms-only="`*get_unprotected_cbr_vms_only*`" ][ count="`*count*`" ][ pageNumnber="`*pageNumnber*`" ]`

Required arguments

None

Optional arguments

`name`

Name of the Virtual Machine

`id`

ID of the Virtual Machine

`host-id`

ID of the Physical Host running Virtual Machines

`get-unprotected-cbr-vms-only`

Get only unprotected Virtual Machines that can participate in Nutanix Converged Backup and Recovery?

Default: false

`count`

Count of the Virtual Machine

`pageNumnber`

PageNumber of the Virtual Machine

Default: false

Get all file level restore capable snapshots attached to a VM.

`ncli> virtualmachine { list-attached-flr-snapshots } vm-id="vm_id" [ snap-id="snap_id" ]`

Required arguments

`vm-id`

ID of the Virtual Machine

Optional arguments

`snap-id`

Id of the Snapshot

Get file level restore capable snapshots of a VM.

`ncli> virtualmachine { list-flr-snapshots | ls-flr-snaps } vm-id="vm_id" [ snapshot-count="snapshot_count" ]`

Required arguments

`vm-id`

ID of the Virtual Machine

Optional arguments

`snapshot-count`

Number of snapshots to be fetched. A value of 0 indicates all snapshots will be returned.

Get snapshots of a VM.

`ncli> virtualmachine { list-snapshots | ls-snaps } vm-id="vm_id" [ snap-id="snap_id" ]`

Required arguments

`vm-id`

ID of the Virtual Machine

Optional arguments

`snap-id`

Id of the Snapshot

Get stats data for Virtual Machine

`ncli> virtualmachine { list-stats | ls-stats }[ name="name" ][ id="id" ][ host-id="host_id" ]`

Required arguments

None

Optional arguments

`name`

Name of the Virtual Machine

`id`

ID of the Virtual Machine

`host-id`

ID of the Physical Host running Virtual Machines

3

CONTROLLER VM COMMANDS

Specifying Credentials

When specifying a password on the command line, always enclose the password in single quotes. For example:`--hypervisor_password='nutanix/4u'`

- To display all user name and password options for `diagnostics.py`, type `/home/nutanix/diagnostics/diagnostics.py --help | egrep -A1 'password|user'`

```
 --hypervisor_password: Default hypervisor password.
    (default: 'nutanix/4u')

  --hyperv_hypervisor_username: The username to use when logging into the local
    Hyper-V node.

  --hypervisor_username: The username to use when logging into the local
    hypervisor.

  --ipmi_password: The password to use when logging into the local IPMI device.
    (default: 'ADMIN')
  --ipmi_username: The username to use when logging into the local IPMI device.
    (default: 'ADMIN')

  --kmip_user_id: Username to insert into certificate signing request.
    (default: 'nutanix')

  --nsc_username: User name that provides services through SSH at Nutanix
    service center.
```

- You can find all user name and password options for `cluster`, `genesis`, and `setup_hyperv.py` by also typing `--help | egrep -A1 'password|user'` as part of the command. For example, `setup_hyperv.py --help | egrep -A1 'password|user'`

cluster

Usage

```
Usage: /usr/local/nutanix/cluster/bin/cluster [flags] [command]

commands:

 convert_cluster
 create
 destroy
 disable_auto_install
 disable_rate_limit_iops
 enable_auto_install
 enable_rate_limit_iops
 foundation_upgrade
 host_upgrade
```

```
ipconfig
lite_upgrade
migrate_zeus
pass_shutdown_token
reconfig
remove_all_public_keys
remove_public_key
restart_genesis
set_rack_aware
set_two_node_cluster_leader
start
status
stop
upgrade
upgrade_node
```

/usr/local/nutanix/cluster/bin/cluster

`--add_dependencies`

Include Dependencies.

Default: false

`--backplane_netmask`

Backplane netmask

`--backplane_network`

Backplane network config

Default: false

`--backplane_subnet`

Backplane subnet

`--backplane_vlan`

Backplane VLAN id

Default: -1

`--block_aware`

Set to True to enable block awareness. This workflow is unsupported, in favour of configuring block awareness on a created cluster

Default: false

`--bundle`

Bundle for upgrading host in cluster.

`--clean_debug_data`

If 'clean_debug_data' is True, then when we destroy a cluster we will also remove the logs, binary logs, cached packages, and core dumps on each node.

Default: false

`--cluster_external_ip`

Cluster ip to manage the entire cluster.

`--cluster_function_list`

List of functions of the cluster (use with create). Accepted functions are ['minerva', 'multicluster', 'two_node_cluster', 'jump_box_vm', 'ags_cluster', 'one_node_cluster', 'xi_vm', 'clusters_on_aws_cluster', 'iam_cluster', 'ndfs', 'extension_store_vm', 'witness_vm', 'cloud_data_gateway']

Default: ndfs

`--cluster_name`

Name of the cluster (use with create).

`--cluster_uuid`

Cluster uuid for cluster in string format.

`--config`

Path to the cluster configuration file.

`--container_name`

Name of the default container on the cluster.

`--dns_servers`

Comma separated list of one or more DNS servers.

`--domain_password`

Domain password of the hosts.

`--domain_username`

Domain username of the hosts.

`--enable_lite_upgrade`

Set to False to disable lite upgrade before it is ready.

Default: true

`--force_install_genesis`

Installs the infrastructure package on all of the nodes.

Default: true

`--foundation_upgrade`

Operation specified will be done in context of foundation upgrade

Default: false

`--hardware_device_type`

Type of hardware device. Please specify one of the following: disk nic hba bios bmc. (Currently supported: disk, bios, bmc.)

`--help`

show this help

Default: 0

`--helpfull`

Show flags for all modules

Default: false

`--helpshort`

show usage only for this module

Default: 0

`--helpxml`

like --help, but generates XML output

Default: false

`--host_upgrade`

Operation specified will be done in context of host_upgrade

Default: false

`--hyperv_sku`

Hypervisor sku to which the HyperV host is being upgraded.

`--hypervisor`

Hypervisor that needs to be upgraded. Possible values: esx, kvm, hyperv.

`--ignore_preupgrade_tests`

Perform preupgrade tests

Default: false

`--ignore_vm_conversion_warnings`

Ignore vm conversion errors during cluster conversion.

Default: false

`--installer_dir_path`

Location of the Nutanix installer directory.

`--ip_specification_json`

JSON file with IP configuration.

`--key_file`

Nutanix default SSH public key.

Default: `/home/nutanix/ssh_keys/nutanix.pub`

`--key_name`

Identifier of the public ssh key in the cluster.

Default: `legacy_nos_compatibility`

`--license_file_zknode`

Path to the zookeeper node that contains the cluster license information.

Default: `/appliance/logical/license/license_file`

`--lockdown_mode`

Flag for lockdown mode

Default: false

`--manual_upgrade`

Manual upgrade method.

Default: false

`--md5sum`

Md5sum of the bundle.

`--migrate_from`

The old zeus node IP address for Zeus migration.

`--migrate_to`

The new zeus node IP address for Zeus migration.

`--ncc_precheck_args`

Additional args to pass to NCC during precheck.

`--no_verification`

Skip verification for package integrity.

Default: true

`--ntp_servers`

Comma separated list of one or more NTP servers.

`--nutanix_default_password_salt`

Seed prism admin password only if the password hash does not contain this salt.

Default: `$6$Mkd8T74/$`

`--password_lockdown_mode`

Flag for password lockdown mode

Default: false

`--provided_root_certificate`

File to be used as the root certificate when creating or upgrading a cluster.

`--provided_root_certificate_key`

File to be used as the root CA private key.

`--python_proto_egg_path`

Path of protobuf egg relative to the install dir.

Default: `lib/py/protobuf-2.6.1-py2.6-linux-x86_64.egg`

`--rack_aware`

Set to True to enable rack awareness. This workflow is unsupported, in favour of configuring rack awareness on a created cluster

Default: false

`--rack_config_json_path`

Path to the json file containing svm_ips to rack namemapping. Json file will contain svm_ips as the keys and the name of the rack they belong to, as values

`--redundancy_factor`

Number of replicas of data as kept in cluster. Clusterwill tolerate node failure of 'redundancy_factor -1'nodes.

Default: -1

`--remove_installer_dir`

Whether or not to remove the installer directory automatically when finished.

Default: true

`--reset_gflags_on_destroy`

When performing cluster destroy, remove all gflag files.

Default: false

`--seed_prism_password`

Seed Prism admin password to be the same as the nutanix user if nutanix user has non-default password.

Default: true

`--shutdown_token_ip`

IP address of intended shutdown token holder SVM.

`--skip_discovery`

Skip discovery.

Default: false

`--skip_ncc_prechecks`

If set to true, skip NCC checks during preupgrade.

Default: false

`--skip_reconfig`

Skip CVM reconfig.

Default: false

`--skip_rf1_vms_check`

If set to true, skip checking if there are running RF1 VMs which will get rebooted during upgrade

Default: false

`--skip_upgrade`

Skip actual upgrade.

Default: false

`--stand_alone_upgrade_timeout`

Timeout for stand-alone upgrade.

Default: 600

`--svm_ips`

Comma separated list of IP addresses of one or more SVMs in the target cluster. Not required if cluster.cfg is being used.

`--svm_login`

User name for logging into SVM.

Default: `nutanix`

`--target_hypervisor`

Target hypervisor type for cluster conversion. Valid types esx/kvm.

`--timeout`

Number of seconds each command to SVMs should take.

Default: 180

`--upgrade_node_ip`

Ipv4 or IPv6 address of node to be upgraded.

`--vcenter_json_file`

File containing vcenter details for dial workflow. The json has host, username and password keys.

`--vcenter_not_required`

Set it to true if vcenter is not used to manage ESX cluster.

Default: false

`--verification_file`

Metadata file for package integrity, upgrade info.

`--version`

Version to which upgrade needs to be performed.

`--wait`

Wait for action to complete.

Default: true

cluster.client.ce_helper

`--ce_version_map_znode_path`

Zookeeper node containing the CE version mapping.

Default: `/appliance/logical/community_edition/version_map`

cluster.client.cluster_upgrade

`--svm_reboot_timeout`

Maximum time expected for SVM to reboot/shutdown.

Default: 420

cluster.client.consts

`--allow_hetero_sed_node`

Flag that can be set by an SRE to let a node have a mix of sed and non-sed disks.

Default: true

`--app_deployment_progress_zknode`

Zknode to use for deployment state machine

Default: `/appliance/logical/app_deployment_progress`

`--app_deployment_proto_zknode`

Zknode to use for deployment state machine

Default: `/appliance/logical/app_deployment_info`

`--authn_enabled`

Boolean signifying if the service authn feature is enabled in the current NOS

Default: true

`--auxiliary_config_json_path`

Path to the auxiliary_config.json file

Default: `/etc/nutanix/auxiliary_config.json`

`--backplane_physical_segmentation_support`

Boolean to control whether to enable/disable Physical segmentation for Backplane feature

Default: true

`--backplane_segmentation_on_dvs_support`

Boolean to control whether to enable/disable Backplane segmentation on DVS feature

Default: true

`--boot_size`

Size of the root partition in bytes.

Default: 20969472

`--boot_size_extra_alignment`

Extra alignment size of the root partition in bytes.

Default: 41943040

`--build_last_commit_date_path`

Path to the file that contains the local release version's last commit date.

Default: `/etc/nutanix/build_last_commit_date`

`--bulk_esx_vm_shutdown_time`

Time allowed for bulk concurrent ESX power operations

Default: 180

`--ca_trust_anchor_path`

Zookeeper path where pki CA certs are stored.

Default: `/appliance/logical/pki_ca_certs`

`--cassandra_health_znode`

Zookeeper node where each cassandra creates an ephmeral node indicating it is currently available.

Default: `/appliance/logical/health-monitor/cassandra`

`--cloud_disk_mount_name_prefix`

Default cloud disk mount name prefix.

Default: `cloud_disk_`

`--cloud_disk_size_bytes`

Size of cloud disk to be added to the cluster in bytes.

Default: 9007199254740991

`--cloud_init_network_conf_file`

Path to the cloud init network configuration file.

Default: `/etc/cloud/cloud.cfg.d/99_network_config.cfg`

`--cloud_storage_tier_random_io_priority`

Random IO priority on cloud disks.

Default: 6

`--cloud_storage_tier_sequential_io_priority`

Sequential IO priority on cloud disks.

Default: 6

`--cluster_disabled_services`

Zookeeper node where a service profile is represented asthe set of services to disable.

Default: `/appliance/logical/cluster_disabled_services`

`--cluster_hibernate_task_status_znode`

Location in Zookeeper where we keep the current status of cluster hibernation task as tracked by different components.

Default: `/appliance/logical/cluster_hibernate_task_status`

`--command_timeout_secs`

Number of seconds to spend retrying an RPC request.

Default: 180

`--compute_only_enabled`

Boolean signifying CO feature support in current NOS

Default: true

`--convert_cluster_zknode`

Holds information about cluster conversion operations and current status for each node.

Default: `/appliance/logical/genesis/convert_cluster`

`--csr_cn_entry`

Common name to use instead of <node_uuid>.nutanix.com

Default: `None`

`--csr_cn_suffix`

Suffix to use instead of nutanix.com when creating CSR

Default: `nutanix.com`

`--cvm_certs_file`

File containing SSL certs of all CVMs in the cluster.

Default: `/home/nutanix/ssh_keys/cvm_certs`

`--cx4_rdma_support`

Boolean to control whether to enable/disable CX4 RDMA feature

Default: true

`--default_cvm_password`

Default password for the CVM.

Default: `nutanix/4u`

`--default_disable_services_file`

Path to the default_disabled_services.json file.

Default: `/home/nutanix/config/genesis/service_profiles/default_disabled_services.json`

`--default_esx_password`

Default password for the ESX hypervisor.

Default: `nutanix/4u`

`--default_upgrade_info_zknode`

Location in a zookeeper where we keep the Upgrade node information.

Default: `/appliance/logical/upgrade_info/nos`

`--degraded_node_policy_dir`

Zookeeper directory where we store the degraded_node_policy

Default: `/appliance/logical/zookeeper-monitor`

`--degraded_node_policy_path`

Zookeeper location to store the degraded node shutdown policy.

Default: `/appliance/logical/zookeeper-monitor/degraded_node_policy`

`--deployment_container_name`

Name for the private container for app images.

Default: `NutanixManagementShare`

`--deployment_info_zknode`

Zknode to keep download info related to uvm deployments

Default: `/appliance/logical/deployment_info`

`--disable_cluster_sync_marker_path`

Path of the marker to disable Cluster Sync.

Default: `/home/nutanix/.disable_cluster_sync`

`--disable_external_ssh`

If True, external SSH access using the password or SSH keys added via prism will be disabled.

Default: false

`--disable_set_irq_affinities`

Boolean to control whether to enable/disable AOS setting of irq_affinities

Default: false

`--disable_xps_driver`

Boolean to control whether to enable/disable XPS driver

Default: true

`--discover_scsi_emulated_nvme`

When True, the SCSI emulated NVMe's are discovered.

Default: true

`--disk_diagnostics_asup_path`

Path for disk diagnostics ASUP data.

Default: `/home/nutanix/data/serviceability/disk_diagnostics_run.json`

`--disk_size_skew_pct_threshold`

Skew threshold (%) for max and min size of disk_size. If the skew computed during a disk addition is determined to be outside the threshold, the extent store size of the disk will be modified to avoid any performance issues.

Default: 15

`--dont_configure_hades_marker_path`

Path of the marker to disable call to Hades configure(), Hades will be configured when this file is deleted.

Default: `/home/nutanix/.dont_configure_hades`

`--dual_stack_network_support`

Flag to control IPv4/IPv6 dual stack support

Default: true

`--enable_cluster_hibernate_resume`

This flag will enable/disable Hibernate Resume feature on CVM. Set to False by default, thereby disabling the RPC in Genesis for initiating the Hibernation or Resume workflow.

Default: false

`--enable_concurrent_operations`

Flag to control the concurrent operations feature with respect to expand cluster for cloud substrates only.

Default: false

`--enable_dynamic_storate_tier_change`

Flag to allow dynamic storage tier change in the Configuration proto.

Default: true

`--enable_hugepages_only_for_spdk`

Enable hugespages only if SPDK is enabled. For non-SPDK clusters, do not enable hugepages.

Default: false

`--enable_hyperv_internal_switch_ha_monitoring`

Flag that can be set by an SRE to disable the HyperV HA algorithm of monitoring the internal switch health.

Default: false

`--enable_legacy_aplos_uwsgi_stack`

If True, then the legacy Aplos stack with uwsgi is started. If False, v3 APIs are routed through Mercury.

Default: false

`--enable_network_configuration_keeper`

Switch to enable/disable network configuration keeper

Default: true

`--enable_parallel_npe_support`

Switch to enable/disable Parallel NPE support

Default: true

`--ergon_register_name`

Name of component to be registered with ergon service.

Default: `Genesis`

`--esx_cvm_backplane_portgroup`

Portgroup for backplane interface created in CVM.

Default: `CVM Backplane Network`

`--esx_external_vswitch`

External vswitch name in ESX host.

Default: `vSwitch0`

`--esx_host_backplane_portgroup`

Portgroup for backplane interface created in ESXi host.

Default: `Backplane Network`

`--expand_cluster_info_znode`

Location in Zookeeper where we keep the info for nodes to be added.

Default: `/appliance/logical/expand_cluster_info`

`--expand_cluster_npe_status`

Zookeeper node to track NPE progress from Expand Cluster

Default: `/appliance/logical/expand_cluster/npe`

`--expand_cluster_status_znode`

Location in Zookeeper where the list for nodes which are currently being added by Genesis are kept.

Default: `/appliance/logical/genesis/expand_cluster_status`

`--expand_cluster_whitelist_znode`

Location in Zookeeper where we keep the whitelist containing allowed hypervisors for imaging new nodes.

Default: `/appliance/logical/expand_cluster_whitelist`

`--expand_npe_enabled`

Switch to enable/disable NPE for Expand Cluster

Default: true

`--experimental_allow_g4_g5_in_same_block`

Flag to disable the check for G4 and G5 nodes comingling in the same block.

Default: false

`--external_switch_rename_enabled`

Feature enable / disable flag for External Switch rename support on HyperV.

Default: `False`

`--firmware_config_json_path`

Path to json file storing information about firmware of hardware devices such as BIOS and BMC.

Default: `/etc/nutanix/firmware_config.json`

`--firmware_installer_storage`

Directory where the firmware installer is located.

Default: `/home/nutanix/data/installer/firmware`

`--firmware_upgrade_fatal_failure`

Firmware upgrade failed and hardware is not working.

Default: `fatal`

`--firmware_upgrade_params_znode`

Zookeeper location to store firmware upgrade parameters.

Default: `/appliance/logical/upgrade_info/firmware_upgrade_params`

`--firmware_upgrade_recoverable_failure`

Firmware upgrade failed but hardware is functioning ok.

Default: `recoverable`

`--firmware_znode`

Zookeeper node where we keep firmware upgrade information.

Default: `/appliance/logical/genesis/firmware`

`--foundation_installer_dir`

Path at which the foundation installer is present.

Default: `/home/nutanix/software_downloads/foundation`

`--foundation_port`

Port on which foundation service listens.

Default: 8000

`--foundation_uncompress_path`

Location for uncompressing foundation package.

Default: `/home/nutanix/software_uncompressed/foundation/`

`--foundation_upgrade_flock_path`

Path to the flock in order that must be acquired before running the foundation upgrade script.

Default: `/home/nutanix/foundation_upgrade_lock`

`--foundation_upgrade_info_znode`

Location in a zookeeper where we keep the foundationUpgrade node information.

Default: `/appliance/logical/upgrade_info/foundation`

`--foundation_upgrade_installer_dir`

Directory where all foundation upgrade related packages are stored.

Default: `/home/nutanix/data/installer/foundation`

`--foundation_upgrade_params_znode`

Zookeeper location to store foundation upgrade paramteters.

Default: `/appliance/logical/upgrade_info/foundation_upgrade_params`

`--foundation_url_port`

Foundation service port.

Default: `8000`

`--foundation_version`

Zookeeper location to store foundation version.

Default: `/appliance/logical/foundation/version`

`--frozen_mounts_file`

Path to the file that indicates that Nutanix mountpoints are currently frozen.

Default: `/tmp/frozen_nutanix_mounts`

`--genesis_bin_dir`

Directory where all of the Genesis scripts are located.

Default: `/home/nutanix/cluster/bin`

`--genesis_cluster_manager_path`

Zk path for healthy genesis node list

Default: `/appliance/logical/pyleaders/genesis_cluster_manager`

`--genesis_jsonrpc_url`

URL of the JSON RPC handler on the Genesis HTTP server.

Default: `/jsonrpc`

`--genesis_libs`

Directory where genesis libraries are located.

Default: `/home/nutanix/cluster/lib`

`--genesis_path`

Path to genesis command on SVMs.

Default: `/home/nutanix/cluster/bin/genesis`

`--genesis_port`

Port that Genesis listens on.

Default: 2100

`--genesis_restart_required_path`

Marker file to indicate that genesis restart is required during upgrade.

Default: `/home/nutanix/.genesis_restart_required_path`

`--genesis_rpc_timeout_secs`

Timeout for each Genesis RPC.

Default: 60

`--genesis_wal_zk_path`

Path to zknode that will be used as a wal to implement any genesis operations idempotently.

Default: `/appliance/logical/genesis/_genesis_wal`

`--hades_block_store_experimental_sleep_secs_before_disk_config_create`

Experimental gflag to sleep before persisting the DiskConfigProto at the specified offset on a disk.

Default: 0

`--hades_block_store_experimental_sleep_secs_before_zeus_config_update`

Experimental gflag to sleep before updating the blockstore info in zeus config for blockstore disks.

Default: 0

`--hades_config_znode_dir`

Parent zookeeper directory for hades config per CVM.

Default: `/appliance/physical/hades/configuration`

`--hades_location`

Location of Hades binary.

Default: `/usr/local/nutanix/bootstrap/bin/hades`

`--hades_mixed_capacity_enabled`

Flag to enable mixed capacity support on the node.

Default: true

`--hades_znode_dir`

Parent zookeeper directory for hades information.

Default: `/appliance/physical/hades`

`--hardware_config_json_path`

Path to the hardware_config.json file.

Default: `/etc/nutanix/hardware_config.json`

`--hardware_config_znode_path`

Zookeeper node base path containing the hardware configurations of each node.

Default: `/appliance/physical/hardware_configs`

`--hcl_json_path`

Path to the hcl.json file.

Default: `/etc/nutanix/hcl.json`

`--health_monitor_znode`

Zookeeper node where ephmeral nodes of services are created.

Default: `/appliance/logical/health-monitor`

`--hibernate_cloud_storage_tier_prefix`

Default storage tier prefix for cloud disks used for hibernate/resume.

Default: `CLOUD-Hibernate-`

`--hibernate_cluster_info_znode`

Location in Zookeeper where we persist the info required for cluster hibernation.

Default: `/appliance/logical/hibernate_cluster_info`

`--hibernate_workflow_status_znode`

Location in Zookeeper where we persist the status of cluster hibernation operation. This is where the hibernation workflow state machine details are saved.

Default: `/appliance/logical/genesis/hibernate_workflow_status`

`--home_nutanix_size`

Size of /home/nutanix partition in bytes.

Default: 83884032

`--home_nutanix_size_extra_alignment`

Extra alignment size of /home/nutanix partition in bytes.

Default: 83886080

`--host_bootdisk_repair_preprocess_znode`

Location in Zookeeper where the status of the host bootdisk repair preprocess is kept.

Default: `/appliance/logical/genesis/host_bootdisk_preprocess`

`--host_bootdisk_repair_status_znode`

Location in Zookeeper where the status of the host bootdisk repair is kept.

Default: `/appliance/logical/genesis/host_bootdisk_repair_status`

`--host_bundle_name`

Base name of the host bundle to be copied to other CVM and hosts.

Default: `host-bundle`

`--host_client_cert_location`

Location of the libvirt certificate to use when establishing a connection.

Default: `/etc/pki/libvirt/clientcert.pem`

`--host_client_key_location`

Location of the libvirt key used when establishing a connection.

Default: `/etc/pki/libvirt/private/clientkey.pem`

`--host_server_cert_location`

Location of the libvirt certificate used when accepting clients.

Default: `/etc/pki/libvirt/servercert.pem`

`--host_server_key_location`

Location of the libvirt key used when accepting clients.

Default: `/etc/pki/libvirt/private/serverkey.pem`

`--host_upgrade_flock_path`

Path to the flock in order that must be acquired before running host upgrade

Default: `/tmp/host_upgrade_lock`

`--host_upgrade_installer_dir`

Directory where all hypervisor upgrade related bundle are stored.

Default: `/home/nutanix/data/installer/hypervisor/`

`--host_znode`

Zookeeper node where we keep the host related dataof the cluster.

Default: `/appliance/logical/genesis/hypervisor`

`--hpssacli_location`

Location of hpssacli utility.

Default: `/usr/local/nutanix/cluster/lib/hp/hpssacli`

`--hyperv_details_znode`

Location in Zookeeper where we keep the HyperV details, like domain info, failover cluster info, used when adding new hyperv nodes to cluster.

Default: `/appliance/logical/genesis/hyperv_details`

`--hyperv_external_vswitch_name`

Default name of the network switch for Hyper-V's external network.

Default: `ExternalSwitch`

`--hyperv_failed_internal_switch_health_znode`

Zookeeper node where genesis on each node monitors the health of their internal switch and publishes if bad.

Default: `/appliance/logical/genesis/hyperv_failed_internal_switch_health`

`--hyperv_host_backplane_adapter`

Backplane Adapter created in Hyperv host.

Default: `Backplane`

`--hyperv_internal_adapter_macaddr`

MAC address of the internal network adapter connected to the CVM in Hyper-V nodes.

Default: `00:15:5D:00:00:80`

`--hyperv_internal_vswitch_name`

Default name of the network switch for Hyper-V's internal network.

Default: `InternalSwitch`

`--hyperv_lsiutil_relative_path`

Path of the LSIUtil utility in a Hyper-V node relative to the Nutanix directory.

Default: `\Utils\LSIUtil.exe`

`--hyperv_megacli64_relative_path`

Path of the MegaCli64 utility in a Hyper-V node relative to the Nutanix directory.

Default: `\Utils\MegaCli64.exe`

`--hyperv_powershell_path`

The path to the powershell binary on the Hyper-V host.

Default: `${env:windir}\System32\WindowsPowerShell\v1.0\powershell.exe`

`--hyperv_sas2ircu_relative_path`

Path of the sas2ircu utility in a Hyper-V node relative to the Nutanix directory.

Default: `\Utils\sas2ircu.exe`

`--hyperv_sas3ircu_relative_path`

Path of the sas3ircu utility in a Hyper-V node relative to the Nutanix directory.

Default: `\Utils\sas3ircu.exe`

`--hyperv_sg_raw_path`

Path to the utility to allow sg_raw on Hyper-V clusters.

Default: `/home/nutanix/bin/winsg_raw`

`--hyperv_sg_scan_relative_path`

Path of the sg_scan utility in a Hyper-V node relative to the Nutanix directory.

Default: `\Utils\sg_scan.exe`

`--hyperv_sg_ses_relative_path`

Path of the sg_ses utility in a Hyper-V node relative to the Nutanix directory.

Default: `\Utils\sg_ses.exe`

`--hyperv_upgrade_min_space`

Minimum space required for hyperv host upgrade.

Default: 6291456

`--hyperv_upgrade_status_znode`

Location where the status of the hyperv upgrade is kept.

Default: `/appliance/logical/genesis/hyperv_upgrade_status`

`--hypervisor_ca_certificate_path`

Location of the CA certificate used on the AHV host.

Default: `/etc/pki/CA/cacert.pem`

`--hypervisor_ilorest_bin_path`

Path of ilorest utility bin on hypervisor

Default: `/usr/bin/`

`--hypervisor_ilorest_lib_path`

Path of ilorest utility lib on hypervisor

Default: `/lib64/`

`--hypervisor_imaging_foundation_zknode`

Holds IP address of CVM which needs to be reimaged using foundation service.

Default: `/appliance/logical/genesis/hypervisor_imaging/foundation`

`--hypervisor_imaging_zknode`

Holds details of the node which needs to be reimaged using foundation service.

Default: `/appliance/logical/genesis/hypervisor_imaging`

`--hypervisor_installer_dir`

Path at which various hypervisor installer tarballs are kept.

Default: `/home/nutanix/software_downloads/hypervisor`

`--hypervisor_installer_info_znode`

Location in Zookeeper where we keep the info about the hypervisor installers.

Default: `/appliance/logical/software_info/hypervisor_installer`

`--hypervisor_release_version_path`

File on hypervisor that contains the local AHV version.

Default: `/etc/nutanix-release`

`--installer_foundation_version_path`

Path to file containing the foundation version that was used for imaging

Default: `/etc/nutanix/foundation_version`

`--installer_phoenix_version_path`

Path to file containing the phoenix version that was used for imaging

Default: `/etc/nutanix/phoenix_version`

`--installer_storage`

Directory where the installer is located.

Default: `/home/nutanix/data/installer`

`--ip6tables_backup_path`

Path to the backup of ip6tables file

Default: `/home/nutanix/config/salt_templates/ip6tables.backup`

`--ip6tables_path`

Path to the ip6tables file

Default: `/etc/sysconfig/ip6tables`

`--iptables_backup_path`

Path to the backup of iptables file

Default: `/home/nutanix/config/salt_templates/iptables.backup`

`--iptables_path`

Path to the iptables file

Default: `/etc/sysconfig/iptables`

`--journal_logs_fstab_entry`

Mount options while binding journal_logs_mount_path to /var/log/journal

Default: `/var/log/journal none defaults,noexec,rw,seclabel,bind,nosuid,noatime,nodev 0 2`

`--journal_logs_mount_path`

Path where /var/log/journal is bind in home partition

Default: `/home/log/journal`

`--key_management_server_test_result`

Path to zknode that will store key management server test results.

Default: `/appliance/logical/genesis/kms_test_result`

`--kmip_user_id`

Username to insert into certificate signing request.

Default: `nutanix`

`--kvm_external_virtual_switch`

External virtual switch name in KVM host.

Default: `vs0`

`--kvm_external_vswitch`

External vswitch name in KVM host.

Default: `br0`

`--kvm_host_backplane_interface`

Backplane interface created in KVM host.

Default: `br0-backplane`

`--large_partitions_flag`

If larger partitions are enabled on this installation.

Default: false

`--large_pc_memory_max_size_gb`

The upper limit of memory size to check if PC is large PC.

Default: 57

`--local_ahv_mnvcli_path`

Path of mnvcli utility

Default: `/usr/local/nutanix/cluster/lib/Linux/mnv_cli`

`--local_ahv_mvcli_path`

Path of mvcli utility

Default: `/usr/local/nutanix/cluster/lib/Linux/mvcli`

`--local_esx_mnvcli_path`

Path of mnvcli utility

Default: `/usr/local/nutanix/cluster/lib/esx5/mnv_cli`

`--local_esx_mvcli_path`

Path of mvcli utility

Default: `/usr/local/nutanix/cluster/lib/esx5/mvcli`

`--local_hyperv_mnvcli_path`

Path of mnv_cli utility

Default: `/usr/local/nutanix/cluster/lib/hyperv/mnv_cli.exe`

`--local_hyperv_mvcli_path`

Path of mvcli utility

Default: `/usr/local/nutanix/cluster/lib/hyperv/mvcli.exe`

`--local_ilorest_esx_bin_path`

ilorest utility binary path for esx

Default: `/usr/local/nutanix/cluster/lib/hpe_ilo/esx/ilorest`

`--local_ilorest_esx_lib_path`

ilorest utility lib path for esx

Default: `/usr/local/nutanix/cluster/lib/hpe_ilo/esx/ilorest_chif.so`

`--local_ilorest_kvm_bin_path`

ilorest utility binary path for kvm

Default: `/usr/local/nutanix/cluster/lib/hpe_ilo/kvm/ilorest`

`--local_ilorest_kvm_lib_path`

ilorest utility lib path for kvm

Default: `/usr/local/nutanix/cluster/lib/hpe_ilo/kvm/ilorest_chif.so`

`--local_timezone_path`

Path to the file that contains the local timezone.

Default: `/etc/localtime`

`--logpipe_path`

Path to logpipe.

Default: `/home/nutanix/bin/logpipe`

`--low_water_mark_file`

File where the time manager logs the time every minute

Default: `/home/nutanix/.low_water_mark.json`

`--manual_discovery_timeout_secs`

Timeout for manual discovery API.

Default: 30

`--max_concurrent_esx_powerops`

Max concurrent ESX VM power operations

Default: 50

`--max_foundation_active_sessions`

Max number of active foundation sessions allowed

Default: 2

`--mem_nvme_storage_tiering_enabled`

Flag to enable memory device storage tiering.

Default: true

`--minimum_disk_size_in_gb_for_extra_alignment`

Minimum disk size required in GB to add extra alignment.

Default: 2048

`--mount_dir`

Path of storage dir where disk should be mounted

Default: `/home/nutanix/data/stargate-storage/disks`

`--ndfs_path_prefix`

Prefix path of ndfs file

Default: `ndfs:///`

`--ndp_timeout_secs`

Timeout for hitting NDP manually.

Default: 20

`--node_in_firmware_upgrade_znode`

This node exists if some node is undergoing firmware upgrade of hardware components. It stores information about that node.

Default: `/appliance/logical/genesis/firmware/upgrade_node`

`--node_reconfigure_marker`

Path to marker file to indicate that this node should be reconfigured.

Default: `/home/nutanix/.node_reconfigure`

`--node_shutdown_token_priority_list`

Location in Zookeeper where we keep list of svm ips which are given preference for token.

Default: `/appliance/logical/genesis/node_shutdown_priority_list`

`--non_ha_vm_info`

Location where the non-HA VMs configuration file path is stored

Default: `/appliance/logical/genesis/non_ha_vm_info`

`--nos_installer_dir`

Path at which NOS installer tarballs are kept.

Default: `/home/nutanix/software_downloads/nos`

`--nos_tar_timeout_secs`

Timeout in secs for tarring nos package

Default: 3600

`--npe_enabled`

Flag to enable/disable NPE feature in the cluster

Default: true

`--npe_info_znode`

Intent znode to start NPE statemachine

Default: `/appliance/logical/npe_info`

`--npe_status_znode`

Location in Zookeeper to persist NPE current operation status for nodes

Default: `/appliance/logical/genesis/npe_status`

`--ns_backplane_config_zknode`

Location in Zookeeper where the backplane segmentation config is kept.

Default: `/appliance/logical/genesis/ns/configuration`

`--ns_backplane_dvs_portkey_map_znode`

Location in Zookeeper where the DVS portkeys for all the vmknics are stored.

Default: `/appliance/logical/genesis/ns/backplane_dvs_portkey_map`

`--ns_child_params_zknode`

Location in Zookeeper where the child params for service network segmentation is kept.

Default: `/appliance/logical/genesis/ns/child_params`

`--ns_config_zknode`

Location in Zookeeper where the services specific network segmentation config is kept.

Default: `/appliance/logical/genesis/ns/services_config`

`--ns_ip_pools_znode`

Location in Zookeeper where the IP Pools for Service Segmentation are stored.

Default: `/appliance/logical/genesis/ns/ip_pools`

`--ns_op_zknode`

Location in Zookeeper where the network segmentation operation parameters and status are kept.

Default: `/appliance/logical/genesis/ns/params`

`--ntp_external_netif`

Network interface of SVM for communicating with NTP server.

Default: `all`

`--ntpd_driftfile_path`

Path to the NTPD daemon's drift file.

Default: `/var/lib/ntp/drift`

`--num_retries_to_update_cloud_disks_to_zeus_proto`

Number of retries to commit the cloud disk info to zeus proto.

Default: 20

`--nutanix_bin_dir`

Path to the nutanix bin folder.

Default: `/home/nutanix/bin`

`--nutanix_binary_log_dir`

Directory containing binary logs.

Default: `/home/nutanix/data/binary_logs`

`--nutanix_core_dir`

Directory containing core dumps of all Nutanix binaries.

Default: `/home/nutanix/data/cores`

`--nutanix_data_dir`

Path to the nutanix data folder.

Default: `/home/nutanix/data`

`--nutanix_default_ssh_key`

Nutanix default SSH key used for logging into SVM.

Default: `/home/nutanix/ssh_keys/nutanix`

`--nutanix_home`

Nutanix home directory.

Default: `/home/nutanix`

`--nutanix_lock_dir`

Directory containing Nutanix service lock files.

Default: `/home/nutanix/data/locks`

`--nutanix_storage_dir`

Nutanix storage directory

Default: `/home/nutanix/data/stargate-storage/disks`

`--one_node_zookeeper_host_port_list`

Host port list for zookeeper in one node cluster.

Default: `zk1:9876`

`--pc_deployment_ephemeral`

Zknode to use for task tracking

Default: `/appliance/logical/pc_deployment_ephemeral`

`--pc_marker_file_path`

PC Marker file path

Default: `/etc/nutanix/pc-marker`

`--persistent_foundation_cvms_zk_path`

Zookeeper node to track which cvms have foundation started in a 'persistent' mode

Default: `/appliance/logical/genesis/persistent_foundation`

`--phoenix_firmware_installer_path`

Path to the module in phoenix image which does firmware install.

Default: `/phoenix/firmware_installer.py`

`--phoenix_firmware_installer_sh_path`

Path to the shell script in phoenix image which triggers firmware install.

Default: `/phoenix/firmware_installer.sh`

`--place_scsi_emulated_nvme_in_pcie_tier`

When True, the SCSI emulated NVMe's are placed in the SSD-PCIe tier.

Default: true

`--planned_outage_esx_enabled`

Switch to enable/disable support for ESX Hypervisor for Planned Outage

Default: false

`--planned_outage_manager_enabled`

Switch to enable/disable support for Planned Outage Manager

Default: true

`--planned_outage_zk_path`

Expect ZK child nodes under this path that are IP addresses of nodes that are going down.

Default: `/appliance/logical/genesis/planned_outage`

`--qos_enabled`

gflag to enable QoS marking for outgoing traffic

Default: true

`--qos_zknode`

Location in Zookeeper where qos mapping is stored.

Default: `/appliance/logical/genesis/qos_info`

`--rate_limit_marker_zknode`

Marker file to rate limit IO operations on CVM.

Default: `/appliance/logical/genesis/rate_limit_marker`

`--release_tag_path`

Path to the file that contains the local release tag.

Default: `/etc/nutanix/release_tag`

`--release_version_path`

Path to the file that contains the local release version.

Default: `/etc/nutanix/release_version`

`--resume_cluster_info_znode`

Location in Zookeeper where we persist the info required to resume cluster from hibernation.

Default: `/appliance/logical/resume_cluster_info`

`--resume_workflow_status_znode`

Location in Zookeeper where we persist the status of resuming a cluster from hibernation. This is where the resume workflow state machine details are saved.

Default: `/appliance/logical/genesis/resume_workflow_status`

`--rf1_vms_task_timeout_secs`

Default timeout value for tasks to handle RF1 VMs

Default: 1200

`--rf1_vms_zk_path`

zk path for rf1 vms

Default: `/appliance/logical/genesis/rf1_vms`

`--rolling_restart_znode`

Location in Zookeeper where the intent of rollingrestart is committed.

Default: `/appliance/logical/genesis/rolling_restart_znode`

`--rolling_svm_update_znode`

Location in Zookeeper where the intent of rollingsvm update is committed.

Default: `/appliance/logical/genesis/rolling_svm_update_znode`

`--run_parallel`

Run various operations in parallel.

Default: false

`--salt_command_path`

Path to the script to run salt related state changes.

Default: `/srv/salt/statechange`

`--salt_state_configuration_directory`

Directory where the current salt states are maintained.

Default: `/home/saltstates`

`--sasircu_location`

Location of sas2ircu utility.

Default: `/home/nutanix/cluster/lib/lsi-sas/sas2ircu`

`--secure_files_zkpath`

Path of the Zookeeper node storing Azure certificates.

Default: `/appliance/logical/secure_files`

`--sed_certs_file_name`

Name of the file containing list of signed certs of the node.

Default: `SEDCerts.txt`

`--sed_flock_path`

File to use as a lock to ensure orderly trusted_storage_device calls to the drives.

Default: `/tmp/sed_tool_lock`

`--sed_status_flush_seconds`

Number of seconds to wait before flushing the status of a self encrypting drive operation.

Default: 10

`--sed_tool_path`

Path to the self_encrypting_drive binary.

Default: `/usr/local/nutanix/bootstrap/bin/self_encrypting_drive`

`--self_encrypting_drive_config_zkpath`

Path to the zookeeper node where self encrypting drive configuration is stored.

Default: `/appliance/logical/genesis/sed_state`

`--self_encrypting_drives_operation_status`

Path to the zookeeper node where each drive will store it's status after an operation was issued

Default: `/appliance/logical/genesis/sed_operation_status`

`--service_segmentation_on_dvs_support`

Boolean to control whether to enable/disable Service segmentation on DVS feature

Default: true

`--service_segmentation_support`

Boolean to control whether to enable/disable Service Segmentation feature

Default: true

`--service_start_timeout_secs`

Default timeout for waitinig for a service to start.

Default: 600

`--service_status_check_interval_secs`

Default interval to check a service running status.

Default: 30

`--service_vm_config_json_path`

Path to the service_vm_config.json file.

Default: `/home/nutanix/data/stargate-storage/service_vm_config.json`

`--services_definition`

Zookeeper node where a service profile stored.

Default: `/appliance/logical/services_definition`

`--services_enablement_status`

Zookeeper node will only contain user intentfor services's state.

Default: `/appliance/logical/services_enablement_status`

`--set_dns_ntp_config_immutable`

Whether to make the DNS and NTP conf files immutable

Default: true

`--sg_raw_path`

Path to the sg_raw binary.

Default: `/usr/bin/sg_raw`

`--sg_ses_location`

Location of sg_ses utility.

Default: `/usr/local/nutanix/cluster/lib/sg3utils/bin/sg_ses`

`--single_ssd_repair_status_znode`

Location in Zookeeper where the status for node which is currently being repaired by Genesis is kept.

Default: `/appliance/logical/genesis/single_ssd_repair_status`

`--skip_handling_rf1_vms`

Whether to skip auto shutdown and start RF1 VMs during AOS and Host maintenance

Default: false

`--skip_service_fault_tolerant_check`

Override option to make upgrade proceed whenfault tolerance is not normal.

Default: false

`--small_pc_data_disk_size_gb`

The size of the data disk for a small PC.

Default: 500

`--small_pc_disk_max_size_gb`

The upper limit of disk size to check if PC is small PC.

Default: 2400

`--snmp_firewall_client`

SNMP client for opening/closing firewall ports

Default: `NutanixSnmpManager`

`--spdk_enabled`

If True, stargate SPDK related activities are performed in Hades.

Default: true

`--spdk_intent_path`

Path where intent is set by spdk_setup.sh.

Default: `/home/nutanix/data/stargate-storage/spdk_setup_intents`

`--spdk_setup_script_path`

Path to the SPDK setup script.

Default: `/usr/local/nutanix/bin/spdk_setup.sh`

`--ssd_threshold_to_enable_multiqueue_nic`

Number of SSDs needed to be on the node to enable features like multi-queue

Default: 4

`--ssh_key_dir`

Location of ssh keys and ssl certificates

Default: `/home/nutanix/ssh_keys`

`--ssh_security_level`

Each level is inclusive of all the levels below it. 0 is default, 1 disables password authentication, 2 disables external access entirely.

Default: 0

`--stargate_block_store_disable_constraint`

Flag to circumvent blockStore constraint check.

Default: false

`--stargate_block_store_enabled`

Flag to enable block store on the cluster.

Default: true

`--stargate_disk_config`

JSON file on stargate disk holding disk config

Default: `disk_config.json`

`--stargate_ha_term_conn_ack_timeout_msecs`

The stargate NfsTerminate ACK timeout value to be set. This is used only when there are hyper-v nodes in the cluster. This value is set on all stargates before upgrade and also set persistently in stargate.gflags

Default: 30000

`--stargate_health_znode`

Zookeeper node where each stargate creates an ephmeral node indicating it is currently available.

Default: `/appliance/logical/health-monitor/stargate`

`--stargate_max_system_mem_for_huge_pages_in_gb`

Maximum memory that could be reserved for rest of the system except Stargate, when Stargate is configured to use hugepages.

Default: 24

`--stargate_min_cvm_memsize_for_huge_pages_in_gb`

Minimum CVM memory size at which we enable hugepages support.

Default: 48

`--stargate_min_system_mem_for_huge_pages_in_gb`

Minimum memory that must be reserved for rest of the system except Stargate, when Stargate is configured to use hugepages.

Default: 19

`--stargate_num_huge_page_allocation_attempt`

Maximum number of attempts to allocate hugepage memory before Stargate decides to fatal.

Default: 10

`--stargate_port`

Port for Stargate.

Default: 2009

`--stargate_system_mem_for_huge_pages_increment_gb`

Amount of physical memory reserved per every 8G over stargate_min_cvm_memsize_for_huge_pages_in_gb for the rest of the system except Stargate, when Stargate is configured to use hugepages.

Default: 1

`--svm_backplane_netif`

Backplane network interface of SVM to communicate with other nodes within cluster.

Default: `eth2`

`--svm_ca_bundle_path`

Path to the CA certificate bundle file.

Default: `/home/nutanix/certs/ca.pem`

`--svm_ca_certs_dir`

Directory where CA certificates are cached on the filesystem.

Default: `/home/nutanix/certs/CA_certs`

`--svm_certs_dir`

Directory where certificates are cached on the filesystem.

Default: `/home/nutanix/certs`

`--svm_csr_path`

Path to local Svm's certificate signing request file.

Default: `/home/nutanix/certs/svm.csr`

`--svm_eth1_mtu_size`

eth1 MTU size configured in the SVMs.

Default: 1500

`--svm_external_netif`

External network interface of SVM to communicate with other SVMs.

Default: `eth0`

`--svm_internal_ip`

IP address of the SVM IP address on the internal Nutanix vSwitch.

Default: `192.168.5.2`

`--svm_internal_netif`

Internal network interface of the SVM to communicate with local hypervisor.

Default: `eth1`

`--svm_internal_sub_netif`

Internal network sub-interface of the SVM to communicate with local hypervisor.

Default: `eth1:1`

`--svm_mtu_size`

eth0 MTU size configured in the SVMs.

Default: 1500

`--svm_os_version`

The default version of SVM os

Default: `centos`

`--svm_private_key_path`

Path to local Svm's digital certificates private key.

Default: `/home/nutanix/certs/svm.key`

`--svm_self_signed_cert_path`

Path to local Svm's self signed certificate.

Default: `/home/nutanix/certs/svm.crt`

`--svmboot_iso_upgraded`

Name of the customized svmboot iso used during AOS upgrade

Default: `svmboot.iso.upgraded`

`--svmboot_xenserver_iso`

xenserver specific svmboot -- to workaround kexec issues

Default: `svmboot-415-xenserver.iso`

`--tentative_backplane_ips`

Location in Zookeeper where the node tentative ips used for network segmentation would be kept.

Default: `/appliance/logical/genesis/tentative_backplane_ips`

`--test_lcm_rpc_enable`

Flag that can be set to allow lcm to expose rpc call to automation test.

Default: false

`--tiny_pc_disk_max_size_gb`

The upper limit of disk size to check if PC is tiny PC.

Default: 480

`--tiny_pc_swap_disk_size_gb`

The size of the swap disk for a tiny PC.

Default: 8

`--tiny_pc_swap_space`

Flag to control the inclusion of swap space in a tiny PC.

Default: false

`--two_node_genesis_rpc_timeout_secs`

Timeout for each Genesis RPC in two node clusters.

Default: 180

`--two_node_zookeeper_host_port_list`

Host port pair list for zookeeper in two node cluster.

Default: `zk1:9876,zk2:9876`

`--upgrade_finish_flock_path`

Path to the flock in order that must be acquired before running the finish script.

Default: `/tmp/upgrade_finish_lock`

`--upgrade_flock_path`

Path to the flock in order that must be acquired before running the upgrade install script.

Default: `/tmp/upgrade_lock`

`--upgrade_info_firmware_znode`

Location in a zookeeper where we keep the upgrade information for firmware upgrades published by Prism.

Default: `/appliance/logical/upgrade_info/firmware`

`--upgrade_info_znode`

Location in a zookeeper where we keep the Upgrade node information.

Default: `/appliance/logical/upgrade_info/prism_central`

`--upgrade_info_znode_dir`

Parent Directory of location in a zookeeper where we keep the Upgrade node information.

Default: `/appliance/logical/upgrade_info`

`--upgrade_params_znode`

Zookeeper location to store upgrade paramteters.

Default: `/appliance/logical/upgrade_info/upgrade_params`

`--use_default_aws_cloud_disk_size`

If True, the default Amazon cloud disk size will be in use.

Default: true

`--vmd_led_tool`

A tool that helps manage NVMe LED when CVM has access to physical root port.

Default: `/usr/local/nutanix/cluster/bin/vmd_led`

`--vmxnet3_rx_queue_size`

RX buffer size for vmxnet3.

Default: 2048

`--vmxnet3_tx_queue_size`

TX buffer size for vmxnet3.

Default: 512

`--wait_secs_for_spdk_intents_to_clear_before_stargate_stop`

Time in seconds for which genesis will wait for ongoing SPDK setup to complete before stopping Stargate. Presence of intent files is used to detect if SPDK setup is in progress.

Default: 120

`--witness_config_cache_path`

Path to the witness configuration cache kept in sync by witness manager

Default: `/home/nutanix/config/witness_config_proto.dat`

`--witness_config_zknode`

Zookeeper node containing the witness config once it is created.

Default: `/appliance/witness_config`

`--xen_details_znode`

Location in Zookeeper where we keep the Xen details, like credentials, etc when adding new xen nodes to cluster.

Default: `/appliance/logical/genesis/xen_details`

`--xi_cluster_ssh_key`

Xi Cluster SSH key used to login using the cluster vipcreated by DCM

Default: `/home/nutanix/xi_ssh_key_dir/nutanix`

`--xi_ssh_key_dir`

Location of ssh keys specific to the cluster created by DCM

Default: `/home/nutanix/xi_ssh_key_dir`

`--zk_commit_timeout`

Max timeout until which genesis attemps to commit to zookeeper.

Default: 60

`--zk_pc_cmsp_failback_spec_path`

Zk path for backup and restore aware utils during failback

Default: `/appliance/logical/cmsp/failback_spec`

`--zk_pc_cmsp_failover_spec_path`

Zk path for backup and restore aware utils during failover

Default: `/appliance/logical/cmsp/failover_spec`

`--zookeeper_ca_certs_zkdir`

Path to the zookeeper directory node where CA certificates are stored.

Default: `/appliance/logical/certs/CA_certs`

`--zookeeper_certs_zkdir`

Path to the zookeeper directory node where Svm certificates are stored.

Default: `/appliance/logical/certs`

`--zookeeper_host_port_list`

Host port pair list for zookeeper.

Default: `zk1:9876,zk2:9876,zk3:9876`

`--zookeeper_migration_zknode`

Path to the zookeeper migration status zknode.

Default: `/appliance/logical/zookeeper_migration`

`--zookeeper_server_config_file`

zookeeper server config file.

Default: `/home/nutanix/data/zookeeper_monitor/zk_server_config_file`

cluster.client.container.docker.utils

`--default_multipath_config_path`

Location of default multipath config.

Default: `/usr/share/doc/device-mapper-multipath-0.4.9/multipath.conf`

`--default_volume_plugin_name`

Name of default docker nutanix volume plugin

Default: `pc/nvp`

`--default_volume_plugin_type`

Default docker volume plugin type

Default: `default`

`--docker_systemd_service`

Name of the systemd docker service

Default: `docker`

`--docker_volume_plugin_binary_path`

Path to volume plugin install script

Default: `/home/nutanix/bin/create_plugin_from_tar.sh`

`--docker_volume_plugin_image_path`

Path to docker volume plugin image

Default: `/usr/local/nutanix/volume-plugin/dvp.tar.gz`

`--volume_plugin_install_timeout_secs`

Timeout in secs for volume plugin installation

Default: 60

`--volume_plugin_version_znode`

Path to docker volume plugin image

Default: `/appliance/logical/genesis/volume_plugin`

cluster.client.esx_upgrade_helper

`--esx_vib_extraction_dir`

Directory where ESXi vibs are extracted on CVM before copying to host.

Default: `/home/nutanix/tmp/.esx_upgrade`

`--foundation_esx_vib_path`

Path in foundation package where ESX VIBs are stored.

Default: `/home/nutanix/foundation/lib/driver/esx/vibs`

`--poweroff_uvms`

Power off UVMs during hypervisor upgrade if Vmotion is not enabled or Vcenter is not configured for cluster.

Default: false

`--undefined_esx_bios_uuid`

Hex value of undefined bios uuid of ESX hypervisor.

Default: `00000000-0000-0000-0000-000000000000`

`--update_foundation_vibs`

Update VIBS which are present in foundation during ESX hypervisor upgrade.

Default: true

cluster.client.firewall.consts

`--all_ips_ipset_name`

Name of the Nutanix cluster ipset to be used in iptables rules containing both CVM and hypervisor IP addresses

Default: `ntnx-all-ips`

`--cluster_function_temp_file`

Path to temporary file which has the cluster function

Default: `/home/nutanix/tmp/cluster_function`

`--consider_salt_framework`

Whether to consider salt framework or not

Default: true

`--cvm_ips_ipset_name`

Name of the Nutanix cluster ipset to be used in iptables rules containing only CVM IP addresses

Default: `ntnx-cvm-ips`

`--execute_concurrent_salt_call`

Indicate if salt call should be executed concurrently.

Default: true

`--firewall_rules_znode`

Location in Zookeeper where cluster-wide firewall rules are kept.

Default: `/appliance/logical/genesis/firewall_rules`

`--handle_cerebro_firewall_rules`

Boolean to decide if Genesis handles Cerebro firewall rules

Default: false

`--host_ips_ipset_name`

Name of the Nutanix cluster ipset to be used in iptables rules containing only hypervisor IP addresses

Default: `ntnx-host-ips`

`--ip_set_based_firewall_support`

Whether to use ipset based firewall rules or subnet based for nutanix services

Default: false

`--iptables_command_timeout_secs`

Timeout in seconds for any iptables command.

Default: 10

`--iptables_salt_blacklist_path`

Path to salt config for iptables state.

Default: `/srv/pillar/blacklist.sls`

`--iptables_salt_config_path`

Path to salt config for iptables state.

Default: `/srv/pillar/iptables.sls`

`--iptables_temp`

Path to temporary iptables file

Default: `/home/nutanix/tmp/iptables.temp`

`--jumpbox_vm_deployments`

Comma separated list of deployments where Jumpbox rules are applied

Default: `512`

`--prism_client_key`

Client key value for prism

Default: `prism.nfs_whitelists`

`--salt_call_command_path`

Path to the salt-call command.

Default: `/usr/bin/salt-call`

`--salt_call_command_timeout_secs`

Timeout in seconds for the salt-call command.

Default: 60

`--salt_dynamic_blacklist`

Path to the dynamic template

Default: `/home/nutanix/config/salt_templates/dynamic_blacklist.json`

`--salt_dynamic_blacklist_temp`

Path to the temporary dynamic blacklist template

Default: `/home/nutanix/config/salt_templates/dynamic_blacklist_temp.json`

`--salt_dynamic_template`

Path to the dynamic template

Default: `/home/nutanix/config/salt_templates/dynamic_open_list.json`

`--salt_dynamic_template_temp`

Path to the temporary dynamic template

Default: `/home/nutanix/config/salt_templates/dynamic_open_list_temp.json`

`--salt_jinja_template`

Path to salt ipv4 jinja template.

Default: `/srv/salt/security/CVM/iptables/iptables4.jinja`

`--salt_pc_jinja_template`

Path to salt ipv4 jinja template on PC.

Default: `/srv/salt/security/PC/iptables/iptables4.jinja`

`--salt_states_templates_dir`

Path to dir holding the salt templates.

Default: `/home/nutanix/config/salt_templates`

`--unsupported_firewall_deployments`

Comma separated list of deployments where firewall will FATAL out

Default: `128,64,8`

`--valid_firewall_deployments`

Comma separated list of deployments where firewall uses salt

Default: `1,16,2,4,256,512,32,4096`

cluster.client.flow_upgrade_helper

`--flow_installation_path`

Location where FLOW is installed on a CVM.

Default: `/usr/local/nutanix/flow/`

`--flow_installation_root`

Root location where FLOW is installed on a CVM.

Default: `/usr/local/nutanix/`

`--flow_installer_path`

Location for uncompressing nutanix FLOW binaries.

Default: `/usr/local/nutanix/flow/installer/`

`--flow_retry_interval`

Interval after which the thread will retry.

Default: 10

`--flow_tmp_path`

Location for uncompressing nutanix FLOW binaries.

Default: `/home/nutanix/software_downloads/flow_tmp/`

`--flow_ui_install_path`

Location where Flow UI bits should be installed

Default: `/home/apache/www/console/`

`--flow_ui_path`

Location where Flow UI bits are located

Default: `/usr/local/nutanix/flow/ui/flowUINg/console/`

`--flow_version_path`

Location for flow_version installed on a CVM.

Default: `/home/nutanix/flow/flow_version.txt`

`--flow_version_znode`

Zookeeper node where we keep the current release version of FLOW.

Default: `/appliance/logical/genesis/flow_version`

cluster.client.genesis.cluster_manager_helper

`--agave_dir`

Identify if agave is running on cluster.

Default: `/home/nutanix/agave`

`--cloud_start_dynamic_ring_changer`

Whether to start Dynamic ring changer for cloud nodes.Dynamic Ring changer is required to run for some time to add new metadata disk. This may be required in customer escalations if existing metadata disk becomes full.

Default: false

`--cluster_state_znode`

Location in Zookeeper where we keep whether a node start or stop.

Default: `/appliance/logical/genesis/cluster_state`

`--cluster_upgrade_method`

Location in Zookeeper where we keep upgrade method.

Default: `/appliance/logical/genesis/cluster_upgrade_method`

`--cluster_versions_znode`

Location in Zookeeper where we keep the desired software versions map.

Default: `/appliance/logical/genesis/cluster_versions`

`--component_registration_endpoint`

Component registry endpoint for registration.

Default: `/api/v3/clients/apikeys/validate`

`--component_registration_max_retries`

Maximum number of retries for component registration.

Default: 60

`--component_registration_retry_max_delay_ms`

The maximum delay between retries in ms.

Default: 60000

`--component_registration_retry_slot_time_ms`

Delay between retries in ms.

Default: 1000

`--component_registration_status_znode`

Location in Zookeeper where we keep track of registration status with component registry.

Default: `/appliance/logical/orchestrator/component_registration_status`

`--component_registry_config_path`

Path to userdata used to connect to component registry.

Default: `/home/nutanix/component_registry_config.json`

`--cvm_reboot_wait`

Timeout for waiting for cvm reboot.

Default: 100

`--force_disable_blackbox`

File to disable blackbox mode completely.

Default: `/home/nutanix/.force_disable_blackbox`

`--node_shutdown_token_state_znode`

Location in Zookeeper where we keep state of which node has currently requested to go down for maintenance.

Default: `/appliance/logical/genesis/node_shutdown_token`

`--node_upgrade_status`

Location in Zookeeper where we store upgrade status of nodes.

Default: `/appliance/logical/genesis/node_upgrade_status`

`--orchestrator_config_filepath`

Path where userdata copied during deployment.

Default: `/home/nutanix/config/mcm_config/pc-environment.json`

`--orchestrator_metdadata_znode`

Location in Zookeeper where we keep orchetrator metadata.

Default: `/appliance/logical/orchestrator/metadata/userdata`

`--pc_dr_recovery_genesis_znode`

Recovery zk node for services to be created right at the beginning by genesis itself

Default: `/appliance/logical/prism/pcdr/recover`

`--pc_dr_recovery_progress_znode`

Zookeeper node of the Pc Recovery Progress

Default: `/appliance/logical/prism/pcdr/startup`

`--pc_release_version_znode`

Zookeeper node where we keep the current pc release version of the cluster.

Default: `/appliance/logical/genesis/pc_release_version`

`--planned_outage_monitor_interval`

Interval at which planned outage watcher thread checksif some remote node needs to be started

Default: 600

`--planned_outage_monitor_znode`

Location in Zookeeper to trigger remote monitoring of CVMs

Default: `/appliance/logical/genesis/planned_outage_monitor`

`--prism_user_repository_znode`

Zookeeper node of the Prism user repository

Default: `/appliance/physical/userrepository`

`--rdma_config_timeout`

Timeout for waiting for RDMA configuration to complete.

Default: 300

`--release_tag_znode`

Zookeeper node where we keep the current release tag of the cluster.

Default: `/appliance/logical/genesis/release_tag`

`--release_version_znode`

Zookeeper node where we keep the current release version of the cluster.

Default: `/appliance/logical/genesis/release_version`

`--shutdown_token_timeout`

Timeout for waiting for shutdown token.

Default: 60

`--uncompress_path`

Location for uncompressing nutanix binaries.

Default: `/home/nutanix/software_uncompressed/nos/`

cluster.client.genesis.compute_only.client

`--configured_marker_file`

Path to the marker file containing cluster id if node is part of a cluster

Default: `/root/configured`

cluster.client.genesis.compute_only.consts

`--factory_config_json_path_on_host`

Path to factory_config.json on the CO host

Default: `/root/factory_config.json`

`--hardware_config_json_path_on_host`

Path to hardware_config.json on the CO host

Default: `/root/hardware_config.json`

cluster.client.genesis.convert_cluster.utils

`--cluster_conversion_preserve_mac`

Preserve MAC addresses of VM NICs in conversion

Default: true

`--convert_cluster_blacklisted_vms`

List of VM UUIDs which won't be converted during cluster conversion

Default: `/appliance/logical/genesis/convert_cluster/blacklisted_vms`

`--convert_cluster_disable_marker`

Marker file to disable hypervisor conversion on node.

Default: `/home/nutanix/.convert_cluster_disable`

`--convert_cluster_node_ids`

List of node ids which will be converted to target hypervisor

Default: `/appliance/logical/genesis/convert_cluster/converting_node_ids`

`--converting_vm_info`

Path to zk node where the reg info of all VMs undergoing conversion is stored

Default: `/appliance/logical/genesis/convert_cluster/converting_vm`

`--default_vcenter_port`

Default port to register with Vcenter.

Default: 443

`--fail_vm_uuids_conversion`

Comma separated list of VM UUIDs which will fail vm conversion

`--fail_vm_uuids_power_off`

Comma separated list of VM UUIDs which will fail vm power off operation during conversion

`--fail_vm_uuids_power_on`

Comma separated list of VM UUIDs which will fail vm power on operation during conversion

`--ignore_reverse_conversion_checks`

Ignore checks performed in reverse conversion eg. presence of ESXi cluster name in Vcenter and existence of metadata zknode, etc.

Default: false

cluster.client.genesis.convert_cluster.vm_migration

`--disable_vm_migration`

Disable VM migration for the node. This is used for error injection and testing.

Default: false

cluster.client.genesis.expand_cluster.utils

`--nos_packages_file`

File containing packages present in the nos software

Default: `install/nutanix-packages.json`

`--num_retries_to_check_for_genesis_availability`

Number of retries to verify if the RPCs are reachable

Default: 30

cluster.client.genesis.hibernate_resume.hibernate_resume_utils

`--service_start_stop_timeout`

Max timeout until which genesis attempts to start or stop cluster services.

Default: 1200

cluster.client.genesis.networking.esx_dvs_helper

`--vm_nic_update_timeout_secs`

Timeout in seconds for vNIC updation operation on CVM

Default: 60

cluster.client.genesis.networking.network_config_utils

`--network_config_zknode`

Location where we create child zknode with the name <node-uuid> and store node specific network configuration.

Default: `/appliance/logical/genesis/network_configuration`

`--network_configuration_zk_connection_timeout_secs`

Timeout in seconds for the network_config_utils module.

Default: 10

cluster.client.genesis.node_manager_helper

`--auto_discovery_interval_secs`

Number of seconds to sleep when local node can't join any discovered cluster.

Default: 5

`--co_nodes_unconfigure_marker`

Path to marker file to indicate that node has to unconfigure CO nodes as part of unconfiguring itself. The contents of the marker file containsspace seperated IPs of the CO nodes to unconfigure

Default: `/home/nutanix/.co_nodes_unconfigure`

`--discovery_timeout_secs`

Timeout for discovery.

Default: 1

`--download_staging_area`

Directory where we will download directories from other SVMs.

Default: `/home/nutanix/tmp`

`--fault_tolerance_initialization_max_retries`

The maximum number of retries to commit modified proto to zeus. Retry indefinitely if set to -1.

Default: 30

`--firmware_disable_auto_upgrade_marker`

Path to marker file to indicate that automatic firmware upgrade should not be performed on this node.

Default: `/home/nutanix/.firmware_disable_auto_upgrade`

`--foundation_disable_auto_upgrade_marker`

Path to marker file to indicate that automatic foundation upgrade should not be performed on this node.

Default: `/home/nutanix/.foundation_disable_auto_upgrade`

`--genesis_restart_timeout`

Time we wait for the genesis to restart.

Default: 120

`--gevent_resolver_reinitialize_on_dns_change`

The flag determines if gevent ares dns resolver should be reinitialized on a DNS config change.

Default: true

`--gold_image_version_path`

Path to the file that contains the version of the gold image.

Default: `/etc/nutanix/svm-version`

`--hcl_znode_path`

Zookeeper node containing the hcl.

Default: `/appliance/physical/hcl`

`--mem_resolution_for_zk_proto_mb`

Minimum memory change of CVM to reflect in proto.

Default: 128

`--move_time_back`

The flag is set to allow time to be moved back by more than time_rollback_tolerance_secs.

Default: false

`--nagios_config_path`

Path to the nagios configuration file.

Default: `/home/nutanix/serviceability/config/nagios3/nutanix_nagios.cfg`

`--node_disable_auto_upgrade_marker`

Path to marker file to indicate that automatic software upgrade should not be performed on this node.

Default: `/home/nutanix/.node_disable_auto_upgrade`

`--node_ssh_key_dir`

Path for node specific ssh keys on local disk

Default: `/home/nutanix/ssh_keys/.blackbox`

`--node_unconfigure_marker`

Path to marker file to indicate that node is ready to be unconfigured.

Default: `/home/nutanix/.node_unconfigure`

`--num_parallel_services_stop`

Maximum number of services to stop in parallel.

Default: 32

`--rpm_genesis_log_file`

Path to rpm log file during Genesis self install.

Default: `/home/nutanix/data/logs/rpm.genesis.out`

`--safe_update_or_restart_retry_interval`

Retry interval to check whether genesis update or restart can be issued.

Default: 30

`--set_node_configured_time_timeout`

Timeout for setting the node_configured_time field inzeus config.

Default: 5

`--sshd_config_path`

Path to sshd config file.

Default: `/etc/ssh/sshd_config`

`--stop_service_zknode`

Start services upto but not including the service specified in this zk node.

Default: `/appliance/logical/genesis/stop_service`

`--stop_services_timeout`

Maximum time to wait for all services to stop.

Default: 120.0

`--svm_internal_ips`

Internal IP addresses on eth1. The first one is the primary IP address, and the remaining are aliases.

Default: `192.168.5.2,192.168.5.254`

`--svm_internal_netif_netmask`

The netmask for the internal IP addresses on eth1.

Default: `255.255.255.128`

`--svm_non_ha_internal_netmask`

The netmask for the non-data internal IP aliases on eth1.

Default: `255.255.255.0`

`--svm_upgrade_finish_zknode`

Zk path to keep track of the svm_upgrade script on two node clusters.

Default: `/appliance/logical/genesis/svm_upgrade_finish`

`--timezones_dir`

Directory where all of the valid timezones exist.

Default: `/usr/share/zoneinfo`

`--vlan_sniffer_log`

Path to vlan_sniffer log.

Default: `/home/nutanix/data/logs/vlan_sniffer.log`

`--vlan_sniffer_path`

Path to vlan sniffer proxy.

Default: `/home/nutanix/cluster/bin/vlan_sniffer_proxy`

`--zookeeper_start_timeout_secs`

Timeout for waiting on Zookeeper connection on startup.

Default: 120

cluster.client.genesis.ns.consts

`--client_for_kVolumes`

Client name for kVolumes, used to pass to Firewall

Default: `service_segmentation_kVolumes`

`--disable_wait_time`

Time to wait (in seconds), between removing network segmentation configuration in zeus and removing the interface configuration on cvms

Default: 5

`--hyperv_physical_backplane_segmentation_support`

Boolean to control whether physical backplane segmentation is supported or not on hyperv

Default: true

`--ip_pool_based_backplane_segmentation`

Boolean to control whether ip pool based backplane segmentation is suppported

Default: true

`--iser_supported_pci_model_ids`

PCI model ids of NICs on which iSER will be supported

Default: `15b3:1017`

`--max_interfaces_for_service_segmentation`

The maximum number of interfaces which can be created on the CVM for service segmentation

Default: 3

`--mixed_hypervisor_backplane_support`

Boolean to control whether backplane network support on mixed hypervisoris allowed or not

Default: true

`--ns_config_cache_path`

Path to the ns configuration cache

Default: `/home/nutanix/config/ns_configuration_proto.dat`

`--ns_state_machine_timeout`

The timeout for completion of network segmentation state machine.

Default: 600

`--ns_wait_for_master_ephemeral`

Ephemeral zknode used in network segmentation enable to ensure pre_reboot phase of rolling reboot is executed after genesis master

Default: `/appliance/logical/genesis/ns/wait_for_master`

`--ports_for_kVolumes`

Comma separated TCP ports needed by kVolumes

Default: `3205,3260`

`--rdma_supported_pci_model_ids`

PCI model ids of NICs on which RF2 RDMA will be supported

Default: `15b3:1013,15b3:1015,15b3:1017`

`--rdma_ztr_supported_pci_model_ids`

PCI model ids of NICs on which RF2 RDMA will be supported

Default: `15b3:1017`

`--retry_count_zk_map_publish`

Retry count for publishing new zk mapping.

Default: 3

`--revert_ns_config_on_failure`

Revert the network segmentation configuration in the case of a failure.

Default: true

`--service_segmentation_support_for_dr`

Boolean to control whether to enable/disable support for DR service segmentation

Default: true

`--zookeeper_ports_for_kVolumes`

Comma separated zookeeper TCP ports needed by kVolumes to be opened on CVM IPs only.

Default: `9876`

cluster.client.genesis.ns.dvs_helper

`--configure_vnics_on_new_node_timeout_secs`

Timeout in seconds to configure vNICS on new node

Default: 120

`--cvm_power_on_wait_time_secs`

Number of seconds we wait for CVM to startup after we power it on

Default: 10

`--cvm_shutdown_timeout_secs`

Timeout in seconds for CVM to shutdown during rolling reboot

Default: 600

`--vm_nic_create_timeout_secs`

Timeout in seconds for vNIC creation operation on CVM

Default: 180

cluster.client.genesis.ns.rdma_helper

`--ahv_iser_ip`

IP address of of the iSER interface on the AHV host

Default: `192.168.5.3`

`--check_rdma_switch_config_script`

Script to check RDMA interface and port config

Default: `/usr/local/nutanix/cluster/bin/check_rdma_switch_config`

`--cvm_iser_ip`

IP address of of the iSER interface on the CVM

Default: `192.168.5.4`

`--mellanox_tc_wrap`

Path to the Mellanox's tc_wrap.py script

Default: `/usr/local/nutanix/bin/tc_wrap.py`

`--rdma_nic_config_file`

Path to json containing the mac of the nic to be used for rdma

Default: `/etc/nutanix/nic_config.json`

cluster.client.genesis.ns.service_utils

`--wait_for_master_timeout_secs`

Timeout in seconds for pre_reboot phase to wait for the genesis master to finish its execution

Default: 600

cluster.client.genesis.pb_client.client

`--genesis_service_host`

The genesis service host.

Default: `127.0.0.1`

`--genesis_service_port`

The genesis service port.

Default: 2100

cluster.client.genesis.resource_management.rm_helper

`--common_pool_map`

Mapping of node with its common pool memory in kb

Default: `/appliance/logical/genesis/common_pool_map`

`--common_pool_mem_for_high_mem_nodes_deep_storage_gb`

Common pool memory reservation for nodes with cvm memory equal to 36gb

Default: 16

`--common_pool_mem_for_high_mem_nodes_gb`

Common pool memory reservation for nodes with cvm memory equal to 36gb

Default: 14

`--common_pool_mem_for_low_mem_nodes_gb`

Common pool memory reservation for nodes with cvm memory less than 20gb

Default: 8

`--common_pool_mem_for_very_high_mem_nodes_deep_storage_gb`

Common pool memory reservation for nodes with cvm memory greater than or equal to 40gb

Default: 20

`--common_pool_mem_for_very_high_mem_nodes_gb`

Common pool memory reservation for nodes with cvm memory greater than or equal to 40gb

Default: 17

`--deep_nodes_capacity_threshold_tb`

Memory threshold of nodes to consider for high common pool mem

Default: 60

`--default_common_pool_memory_in_gb`

Stargate default common pool memory reservation

Default: 12

`--memory_update_history`

File containing history of memory update on node

Default: `/home/nutanix/config/memory_update.history`

`--memory_update_resolution`

Minumum amount of memory difference for update

Default: 2097152

`--rolling_restart_memory_update_reason`

Reason set in rolling restart for memory update

Default: `cvm_memory_update`

`--target_memory_zknode`

CVM target memory map zk node

Default: `/appliance/logical/genesis/target_memory_map`

cluster.client.genesis.resource_management.rm_prechecks

`--cushion_memory_in_kb`

Cushion Memory required in nodes before update

Default: 2097152

`--delta_memory_for_nos_upgrades_kb`

Amount of CVM memory to be increased during NOS upgrade

Default: 4194304

`--host_memory_threshold_in_kb`

Min host memory for memory update , set to 62 Gb

Default: 65011712

`--max_cvm_memory_upgrade_kb`

Maximum allowed CVM memory for update during upgrade

Default: 31457280

cluster.client.genesis.resource_management.rm_tasks

`--cvm_reconfig_component`

Component for CVM reconfig

Default: `kGenesis`

`--cvm_reconfig_operation`

Component for CVM reconfig

Default: `kCvmreconfig`

cluster.client.genesis_utils

`--genesis_up_retries`

Number of retries for node genesis rpcs to be up after reboot

Default: 40

`--maintenance_mode_retry_count`

Number of times Genesis will attempt to put the host into maintenance mode.

Default: 3

`--min_nos_version_for_stargate_huge_page_support`

NOS version starting which huge pages are supported for Stargate

Default: `5.19`

`--orion_config_path`

Path to orion config

Default: `/appliance/logical/orion/config`

`--svm_default_login`

User name for logging into SVM.

Default: `nutanix`

`--sync_network_configuration`

Location in Zookeeper where a workflow can write to run an on demand network collection

Default: `/appliance/logical/genesis/network_configuration/sync_network_configuration`

`--timeout_HA_route_verification`

Timeout for setting HA route.

Default: 180

`--timeout_zk_operation`

Timeout for zk operation like write

Default: 120

`--upgrade_fail_marker`

Marker to indicate upgrade has failed.

Default: `/appliance/logical/genesis/upgrade_failed`

cluster.client.hades.flags

`--hades_import_timeout_secs`

Timeout for import Disk Manager using hades_proxy.

Default: 600

`--hades_jsonrpc_url`

URL of the JSON RPC handler on the Hades HTTP server.

Default: `/jsonrpc`

`--hades_port`

Port that Hades listens on.

Default: 2099

`--hades_rpc_timeout_secs`

Timeout for each Hades RPC.

Default: 30

`--lsiutil_location`

Location of lsiutil utility.

Default: `/home/nutanix/cluster/lib/lsi-sas/lsiutil`

cluster.client.host_upgrade_common

`--host_poweroff_uvm_file`

File containing list of UVM which are powered off for host upgrade.

Default: `/home/nutanix/config/.host_poweroff_vm_list`

`--upgrade_delay`

Seconds to wait before runninguprgade script in Esx Host.

Default: 0

cluster.client.host_upgrade_helper

`--host_disable_auto_upgrade_marker`

Path to marker file to indicate that automatic host upgrade should not be performed on this node.

Default: `/home/nutanix/.host_disable_auto_upgrade`

`--hypervisor_upgrade_history_file`

File path where hypervisor upgrade history is recorded.

Default: `/home/nutanix/config/hypervisor_upgrade.history`

`--hypervisor_upgrade_info_znode`

Location in a zookeeper where we keep the Hypervisor upgrade information.

Default: `/appliance/logical/upgrade_info/hypervisor`

cluster.client.hyperv_upgrade_helper

`--perpetual_container`

Container name to test if hypervisors have access to storage.

Default: `NutanixManagementShare`

`--upgrade_helper_rshell_timeout`

The default timeout for RemoteShell connection made by HyperVHostObject.

Default: 600

cluster.client.ipv4config

`--end_linklocal_ip`

End of the range of link local IP4 addresses.

Default: `169.254.254.255`

`--esx_cmd_timeout_secs`

Default timeout for running a remote command on an ESX host.

Default: 120

`--hyperv_cmd_timeout_secs`

Default timeout for running a remote command on an hyperv host.

Default: 120

`--ip_version_preference`

Version to give preference when the machine is dual stack

Default: 4

`--ipmi_apply_config_retries`

Number of times to try applying an IPMI IPv4 configuration before failing.

Default: 6

`--kvm_cmd_timeout_secs`

Default timeout for running a remote command on an KVM host.

Default: 120

`--kvm_external_network_interface`

Default name of the network device for KVM's external network.

Default: `br0`

`--linklocal_netmask`

End of the range of link local IP4 addresses.

Default: `255.255.0.0`

`--start_linklocal_ip`

Start of the range of link local IP4 addresses.

Default: `169.254.1.0`

`--xen_external_network_interface`

Default name of the network device for Xen's external network.

Default: `xapi1`

cluster.client.kvm_upgrade_helper

`--ahv_enter_maintenance_mode_check_max_retry_count`

Max retry count for checking enter maintenance mode

Default: 20

`--ahv_enter_maintenance_mode_retry_max_delay_secs`

Max delay time for exponential backoff retries to enter maintenance

Default: 720

`--ahv_enter_maintenance_mode_retry_slot_time_secs`

Slot time for exponential backoff retries to enter maintenance

Default: 120

`--ahv_enter_maintenance_mode_timeout_secs`

Seconds to wait before retrying enter maintenance mode operation on failures

Default: 7200

`--ahv_remove_unknown_packages_on_upgrade`

Remove RPMs not in AHV manifest file on upgrade

Default: false

`--kvm_reboot_delay`

Seconds to delay before reboot KVM host

Default: 30

cluster.client.lcm.lcm_genesis_utils

`--rpc_timeout`

Default timeout for RPC.

Default: `600.0`

cluster.client.license_config

`--license_config_file`

Zookeeper path where license configuration is stored.

Default: `/appliance/logical/license/configuration`

`--license_config_proto_file`

License configuration file shipped with NOS.

Default: `configuration.cfg`

`--license_dir`

License feature set files directory shipped with NOS.

Default: `/home/nutanix/serviceability/license`

`--license_public_key`

License public key string shipped with NOS.

Default: `/appliance/logical/license/public_key`

`--license_public_key_str`

License public key string shipped with NOS.

Default: `public_key.pub`

`--zookeeper_license_cbl_path`

Zookeeper path where cbl flag information is stored.

Default: `/appliance/logical/license/pc_cbl_status`

`--zookeeper_license_root_path`

Zookeeper path where license information is stored.

Default: `/appliance/logical/license`

`--zookeeper_license_trial_root_path`

Zookeeper path where license trial information is stored.

Default: `/appliance/logical/license/trial`

cluster.client.lite_upgrade.core.consts

`--cluster_sync_path`

Path to cluster_sync command on SVMs.

Default: `/home/nutanix/cluster/bin/cluster_sync`

`--hades_path`

Path to hades command on SVMs.

Default: `/home/nutanix/cluster/bin/hades`

`--lite_upgrade_running_zknode`

Zk node that tracks if lite upgrade is running.

Default: `/appliance/logical/genesis/lite_upgrade/active`

cluster.client.lite_upgrade.interfaces.genesis_interface

`--genesis_lu_intent_zknode`

Lite upgrade zk node, set with target version.

Default: `/appliance/logical/genesis/lite_upgrade/genesis_intent`

cluster.client.multihome_utils

`--multihome_zkpath`

Marker to indicate if any node of cluster is multihome.

Default: `/appliance/logical/genesis/multihome`

cluster.client.ncc_upgrade_helper

`--cluster_health_shutdown_max_retries`

Max number of retries to shutdown cluster health.

Default: 5

`--cluster_health_shutdown_threshold_ms`

Time threshold (ms) between cluster health shutdown retries.

Default: 2000

`--ncc_installation_path`

Location where NCC is installed on a CVM.

Default: `/home/nutanix/ncc`

`--ncc_num_nodes_to_upload`

Number of nodes to upload the NCC installer directory to.

Default: 2

`--ncc_uncompress_path`

Location for uncompressing nutanix NCC binaries.

Default: `/home/nutanix/data/ncc/installer`

`--ncc_upgrade_info_znode`

Location in a zookeeper where we keep the Upgrade node information.

Default: `/appliance/logical/upgrade_info/ncc`

`--ncc_upgrade_params_znode`

Zookeeper location to store NCC upgrade parameters.

Default: `/appliance/logical/upgrade_info/ncc_upgrade_params`

`--ncc_upgrade_status`

Location in Zookeeper where we store upgrade status of nodes.

Default: `/appliance/logical/genesis/ncc_upgrade_status`

`--ncc_upgrade_timeout_secs`

Timeout in seconds for the NCC upgrade module.

Default: 30

`--ncc_version_znode`

Zookeeper node where we keep the current release version of NCC.

Default: `/appliance/logical/genesis/ncc_version`

cluster.client.preupgrade_checks

`--arithmos_binary_path`

Path to the arithmos binary.

Default: `/home/nutanix/bin/arithmos`

`--check_flow_networking_enablement`

Flag to control if Flow Networking enablement is to be checked.

Default: true

`--check_pc_cmsp_compatibility`

Flag to control if CMSP enablement should be enforced.

Default: false

`--cluster_external_state`

Cluster external state zk path

Default: `/appliance/physical/clusterexternalstate`

`--connected_cluster_path`

Connected cluster zk path

Default: `/appliance/physical/zeusconfig`

`--disable_xi_build_to_build_upgrades`

Disable build to build upgrade support for Xi clusters

Default: False

`--flow_networking_ga_version`

For all PCs before this version, Flow Networking must be disabled before upgrading beyond it.

Default: `2021.7`

`--license_config_path`

The path to the license configuration.

Default: `/appliance/logical/license/configuration`

`--min_disk_space_for_upgrade`

Minimum space (KB) required on /home/nutanix for upgrade to proceed.

Default: 3600000

`--min_replication_factor`

Minimum replication factor required per container.

Default: 1

`--min_space_post_uncompress`

Minimum space (KB) required on /home/nutanix on the preupgrade driving node for upgrade to proceed.

Default: 1500000

`--minimum_memory_for_prism_pro_for_tiny_pc_in_gb`

Minimum memory needed for prism pro features in case of tiny pc.

Default: 8

`--minimum_memory_for_prism_pro_in_gb`

Minimum memory needed for prism pro features.

Default: 16

`--mountsfile`

Path to the mounts file in proc.

Default: `/proc/mounts`

`--prism_gateway_port`

The port on which prism gateway is running.

Default: 9440

`--role_mapping_path`

Role mapping zk path

Default: `/appliance/logical/prism/rolemapping`

`--signature_verification_available_version`

Release version in which gpg signature validation is supported.

Default: `5.8`

`--stargate_mount_prefix`

Mount point prefix for stargate disks

Default: `/home/nutanix/data/stargate-storage/disks/`

`--svm_rpc_timeout`

RPC timeout per SVM in the cluster.

Default: 10

cluster.client.preupgrade_checks_ncc_helper

`--ncc_esx_lockdown_cmd_timeout_secs`

Time limit imposed while querying for lockdown status on ESX hosts.

Default: 60

`--ncc_temp_location`

Location to extract NCC.

Default: `/home/nutanix/ncc_preupgrade`

cluster.client.service.curator_service

`--cluster_fault_tolerance_status`

Value that indicates that the service is fault tolerant to allow shutdown of a node at this time

Default: 1

`--curator_cgroup_memory_limit_mb`

If -1, creation of cgroup with memory limit is disabled for this component. If specified as a positive number then a cgroup is created for this component with this set as the memory limit.

Default: 3328.0

`--curator_config_json_file`

JSON file with curator configuration

Default: `curator_config.json`

`--curator_data_dir_size`

Curator data directory size in MB (80 GB).

Default: 81920

`--curator_data_dir_symlink`

Path to curator data directory symlink.

Default: `/home/nutanix/data/curator`

`--curator_data_disk_subdir`

Path to curator subdirectory in data disk.

Default: `curator`

`--curator_oom_score`

If -1, OOM is disabled for this component. If in [1, 1000], it is taken as the OOM score to be applied for this component.

Default: -1

`--curator_path`

Path to the curator binary.

Default: `/home/nutanix/bin/curator`

`--curator_rss_mem_limit`

Maximum amount of resident memory Curator may use on an Svm with 8GB memory configuration.

Default: 536870912

cluster.client.service.foundation_service

`--foundation_memory_limit_mb`

If -1, creation of cgroup with memory limit is disabled for this component. If specified as positive number then a cgroup is created for this component with this set as the memory limit.

Default: 160

`--foundation_path`

Path to the foundation service script.

Default: `/home/nutanix/foundation/bin/foundation`

`--foundation_rss_mem_limit`

Maximum amount of resident memory Foundation may use on a CVM with 8GB memory configuration.

Default: 167772160

cluster.client.service.kafka_gflags

`--kafka_disk_size_factor_wrt_data_disk`

Kafka disk size = data disk size / <this constant>

Default: 7

`--kafka_disks_base_dir`

Base directory of kafka data disks. Each disk is mounted with directory name being disk serial on this base dir. Applicable to host attached disks

Default: `/home/nutanix/data/kafka/disks/`

`--kafka_pc_deployed_disk_scsi_id`

When PC is deployed, kafka disk is attached at this scsi id

Default: 4

cluster.client.service.service_utils

`--auto_set_cloud_gflags`

If true, recommended gflags will be automatically set for cloud instances.

Default: true

`--cgroup_subsystems`

Default subsystems used for cgroup creation.

Default: `memory,freezer,net_cls`

`--containersvcmon_path`

Path to container service monitor binary.

Default: `/home/nutanix/bin/containersvcmon`

`--enable_service_monitor`

If true, C based service_monitor will re-spawn the service process upon exit, else Python based self_monitor will be used.

Default: true

`--memory_limits_base_size_kb`

Total memory size of the standard Svm based on which memory limits are derived.

Default: 8388608

`--path_to_cgclassify_binary`

Path to cgclassify binary, which is used to add a process into cgroup.

Default: `/bin/cgclassify`

`--path_to_cgcreate_binary`

Path to cgcreate binary, which is used to create a cgroup.

Default: `/bin/cgcreate`

`--path_to_cgset_binary`

Path to cgset binary, which is used to set parameter for cgroup.

Default: `/bin/cgset`

`--path_to_chrt_binary`

Path to chrt binary, which is used to set realtime priority for a process.

Default: `/usr/bin/chrt`

`--path_to_root_cgroup`

Path to root cgroup directory.

Default: `/sys/fs/cgroup`

`--path_to_taskset_binary`

Path to taskset binary, which is used to set cpu affinity for a process.

Default: `/bin/taskset`

`--service_cmdlines_dir`

Path to the directory where service command lines are stored.

Default: `/home/nutanix/data/cmdlines`

`--service_monitor_disable_utc_timestamps`

It True, use the local system time instead of UTC timestamps in the log messages emitted by the service monitor and service_proxy.

Default: false

`--service_monitor_path`

Path to the service monitor binary.

Default: `/home/nutanix/bin/service_monitor`

`--service_proxy_path`

Path to the service proxy binary.

Default: `/home/nutanix/bin/service_proxy`

`--service_start_as_container`

Flag to indicate service can start as containers

Default: true

`--service_start_wait_sec`

Wait time in sec for genesis service to start.

Default: 5

`--service_stop_wait`

Wait time for how long we should wait before we start issuing SIGKILL instead of SIGTERM.

Default: 5

cluster.client.sshkeys_helper

`--authorized_keys_file`

Path to file containing list of permitted RSA keys

Default: `/home/nutanix/.ssh/authorized_keys2`

`--authorized_keys_file_admin`

Path to file containing list of permitted RSA keys for admin.

Default: `/home/admin/.ssh/authorized_keys2`

`--external_keys_file`

Path to file containing list of external keys

Default: `/home/nutanix/.ssh/external_keys`

`--id_rsa_path`

Nutanix default SSH key used for logging into SVM.

Default: `/home/nutanix/.ssh/id_rsa`

`--id_rsa_public_pem_path`

PEM formatted public key for this node.

Default: `/home/nutanix/.ssh/id_rsa.pub.pem`

`--node_ip_in_authorized_keys`

Whether or not to have all the node ips in the from= section of the authorized_keys2 file to add additional security.

Default: false

`--ssh_client_configuration`

Location of ssh client configuration file.

Default: `/home/nutanix/.ssh/config`

`--ssh_config_server_alive_count_max`

Max SSH keepalive messages missed before declaring connection dead.

Default: 3

`--ssh_config_server_alive_interval`

SSH keepalive message interval.

Default: 10

`--ssh_fingerprint_file`

Location of ssh public key fingerprint file.

Default: `/home/nutanix/.ssh/fingerprints`

`--ssh_path`

Location of ssh folder for nutanix.

Default: `/home/nutanix/.ssh`

`--ssh_path_admin`

Location of ssh folder for admin.

Default: `/home/admin/.ssh`

cluster.client.time_manager.time_manager_utils

`--ntpdate_timeout_secs`

Timeout to wait for ntp server to return a valid time.

Default: 10

cluster.client.two_node.helper_utils

`--allow_replacement_of_separated_node`

If True keep genesis alive when there is only one node in the zeus config proto, even though peer monitoring cannot be done. This will allow the addition of a second node, after which the two node cluster manager will reinitialize.

Default: true

cluster.client.two_node.state_transitions

`--witness_state_history_size`

Maximum length of witness state history to keep in zeus configuration.

Default: 10

cluster.client.upgrade_helper

`--arithmos_rpc_timeout`

Timeout for arithmos RPCs retries

Default: 180

`--cluster_name_update_timeout`

Default timeout for updating the cluster name in zeus.

Default: 5

`--num_nodes_to_upload`

Number of nodes to upload the installer directory to.

Default: 2

`--nutanix_packages_json_basename`

Base file name of the JSON file that contains the list of packages to expect in the packages directory.

Default: `nutanix-packages.json`

`--uncompress_buffer_ratio`

Space Buffer ratio to uncompress a compressed file.

Default: 0.2

`--upgrade_genesis_restart`

Location in Zookeeper where we store if genesis restart is required or not.

Default: `/appliance/logical/upgrade_info/upgrade_genesis_restart`

cluster.client.utils.foundation_rest_client

`--foundation_ipv6_interface`

Ipv6 interface corresponding to eth0

Default: 2

cluster.client.utils.foundation_utils

`--foundation_root_dir`

Root directory for foundation in CVM

Default: `/home/nutanix`

cluster.client.utils.hypervisor_ha

`--ha_service_reforward_fix_enabled`

If True, return False if we failed to update the route.

Default: false

`--internal_nutanix_portgroup_name`

Name of the internal Nutanix portgroup configured on ESX.

Default: `vmk-svm-iscsi-pg`

`--terminate_connection_timeout_secs`

Timeout in seconds for the NfsTerminateConnection RPC issued to stargate on a failback.

Default: 15

cluster.client.utils.state_machine

`--sm_debug_mode`

State Machine runs in debug mode based on a config file

Default: false

`--sm_debug_mode_config_file`

Location of configuration json for debugging

Default: `/home/nutanix/tmp/sm_debug_config.json`

cluster.client.utils.volume_utils

`--hosting_pe_external_ip`

PE VM endpoint.

`--pe_zeus_config_path`

Zookeeper node path to PE config.

Default: `/appliance/physical/zeusconfig`

cluster.client.xen_upgrade_helper

`--xen_maintenance_mode_retries`

Seconds to delay before reboot Xen host

Default: 5

`--xen_reboot_delay`

Seconds to delay before reboot Xen host

Default: 30

`--xen_uvm_no_migration_counter`

Number of retries to wait for UVMs to migrate from xen host after it is put in maintenance mode

Default: 7

`--xen_webserver_port`

Port for webserver that will serve files required during xen upgrades

Default: 8999

cluster.cmsp.consts

`--airgap_url`

Airgap URL.

`--aws_docker_registry_endpoint`

Amazon Docker Registry endpoint

Default: `https://464585393164.dkr.ecr.us-west-2.amazonaws.com`

`--aws_ecr_api_endpoint`

Amazon ECR API endpoint

Default: `https://api.ecr.us-west-2.amazonaws.com`

`--aws_s3_endpoint`

Amazon S3 buckets endpoint

Default: `https://prod-us-west-2-starport-layer-bucket.s3.us-west-2.amazonaws.com`

`--cmsp_network_name`

CMSP network name

Default: `cmsp_network`

`--cmsp_precheck_retry_cnt_nodes_alive`

CMSP precheck to perform retries if SVMIPs does not match IDF server list.

Default: 30

`--cmsp_upgrade_init_marker`

Marker file to determine cmsp post upgrade initialisation

Default: `/home/nutanix/.cmsp_init_upgrade`

`--cmsp_upgrade_marker`

Marker file to determine cmsp upgrade

Default: `/home/nutanix/.cmsp_upgrade`

`--connection_time_out_secs`

Amount of time in seconds, the client waits to connect to the server

Default: 5

`--cred_helper_path`

Path to credential helper to encrypt and encode strings

Default: `/home/docker/msp_controller/bootstrap/msp_tools/credHelper`

`--default_aplos_api_port`

Default Aplos API port on PC.

Default: 9440

`--default_data_services_port`

Default data services port on PE.

Default: 3260

`--download_portal_url`

Download portal URL

Default: `http://download.nutanix.com`

`--enable_timeout`

Timeout for enable process.

Default: 7200

`--idf_health_znode`

Zookeeper node where each idf creates an ephemeral node indicating it is current availability.

Default: `/appliance/logical/health-monitor/insights_server`

`--inject_error_cmsp_state`

State for which cmsp state machine should FATAL

Default: 1000

`--lcm_config_node_path`

LCM Config ZKNode path

Default: `/appliance/logical/lcm/config`

`--local_registry_path`

Mount point of registry in msp_controller container

Default: `/home/msp_config/msp-registry.qcow2`

`--migration_timeout`

Timeout for migration process per cluster in seconds.

Default: 1200

`--min_ips_scale_out`

Minimum IPs needed in the block for scale-out PC.

Default: 7

`--min_ips_single_node`

Minimum IPs needed in the block for single node PC.

Default: 2

`--min_pc_memory_for_cmsp`

Minimum installed memory on PC needed to enable CMSP.

Default: 16

`--msp_api_url`

Base URL for MSP APIs.

Default: `http://localhost:2082/msp`

`--msp_bundle_absolute_path`

Absolute path of the MSP bundle in the LCM server

Default: `/var/www/html/release/builds/msp-builds`

`--msp_bundle_relative_path`

Relative path of the MSP bundle in the LCM server

Default: `/builds/msp-builds`

`--msp_controller_start_timeout`

Timeout for MSP controller service to come up.

Default: 1800

`--msp_controller_start_wait`

Wait time for MSP controller service to come up.

Default: 180

`--msp_volume_plugin_name`

Name of CMSP volume plugin

Default: `nutanix:latest`

`--mspctl_path`

Path to mspctl on PC.

Default: `/usr/local/nutanix/cluster/bin/mspctl`

`--pc_cmsp_services_zk_path`

ZK path for overriding list of PC services to be registered on C-MSP.

Default: `/appliance/logical/genesis/pc_services_cmsp_list`

`--pc_disk_component`

Component name for additional disk on PC.

Default: `msp`

`--pc_disk_path`

Path to additional persistent storage disk on PC.

Default: `/home/nutanix/data/sys-storage`

`--pc_disk_size_mb`

Size of additional persistent storage disk on PC.

Default: 102400

`--set_airgap_url`

Force Airgap URL using flag

Default: false

`--skip_migration_prechecks`

Skip prechecks for migration

Default: false

cluster.deployment.deployment

`--APP_DEPLOYMENT_TIMEOUT`

Timeout in seconds for "Application Deployment" taskto complete. Timeout for App deployment state machine execution

Default: 8400

`--CMSP_DEPLOYMENT_TIMEOUT`

Timeout in seconds for CMSP Deployment

Default: 7200

`--DEPLOYMENT_TASK_TIMEOUT`

Timeout in seconds for the PC deployment task(Prism Central Deployment). Timeout of PCD genesis part.

Default: 9000

`--GENESIS_BOOT_TIMEOUT`

Timeout in seconds for genesis to come up on PC.

Default: 3000

`--SHORT_POLL_FREQUENCY`

Frequency in seconds at which PE genesis polls PC genesis to check if it has come up or not.

Default: 5

`--TASK_TRACKING_TIMEOUT`

Time interval (in seconds) after which __check_pc_deployment thread checks for an incomplete PC deployment task and mark it failed if its stuck

Default: 300

`--enable_cmsp_over_ssh`

Enable cmsp over ssh

Default: true

`--min_total_multi_vdisk_size_mb`

The minimum total size of data disk in PC that qualifies for multi VDisks

Default: 2560000

`--multi_vdisk`

Set this to True to enable attaching multi VDisk for data disk size larger than 2.5 TB.

Default: false

`--multi_vdisk_pc_start_scsi_id`

The first scsi id of extra data disk to be added for PC deployment

Default: 5

`--num_cmsp_disks`

Total Number of CMSP disks

Default: 1

`--number_of_vdisk`

Number of VDisk to be added for PC deployment

Default: 4

`--prepare_pc_disks_script`

Path to the script that partitions PC data disks

Default: `/usr/local/nutanix/cluster/bin/prepare_pc_disks`

`--single_vdisk_pc_scsi_id`

The scsi id for single data disk to be added for PC deployment

Default: 3

cluster.deployment.deployment_utils

`--cmsp_api_timeout`

Timeout in seconds for CMSP apis.

Default: 300

`--cmsp_esx_gen_retry_count`

Retry count for generating esx spec during cmsp spec generation.

Default: 30

`--default_password_reset_timeout`

Timeout in seconds for the executing the password reset script on the PC VM.

Default: 480

cluster.disk_flags

`--clean_disk_log_path`

Path to the logs from the clean_disks script.

Default: `/home/nutanix/data/logs/clean_disks.log`

`--clean_disk_script_path`

Path to the clean_disks script.

Default: `/home/nutanix/cluster/bin/clean_disks`

`--disk_partition_margin`

Limit for the number of bytes we will allow to be unpartitioned on a disk.

Default: 2147483648

`--disk_size_threshold_percent`

Percentage of available disk space to be allocated to stargate

Default: 95

`--disk_size_threshold_size`

Size of the available disk space not to be allocated to stargate

Default: 107374182400

`--enable_all_ssds_for_oplog`

DEPRECATED: Use all ssds attached to this node for oplog storage.

Default: true

`--enable_fio_realtime_scheduling`

Use realtime scheduling policy for fusion io driver.

Default: false

`--fio_realtime_priority`

Priority for fusion io driver, when realtime scheduling policy is being used.

Default: 10

`--format_fusion_percent`

The percentage of total capacity of fusion-io drives that should be formatted as usable

Default: 60

`--max_ssds_for_oplog`

Maximum number of ssds used for oplog per node. If value is -1, use all ssds available. If only_select_nvme_disks_for_oplog gflag is true and NVMe disks are present, only NVMe disks are used for selecting oplog disks.

Default: 12

`--metadata_maxsize_GB`

Maximum size of metadata in GB

Default: 30

`--only_select_low_latency_disks_for_oplog`

If true and low latency disks are present, only use these disks for selecting oplog disks.

Default: false

`--only_select_nvme_disks_for_oplog`

If true and NVMe disks are present, only use NVMe disks for selecting oplog disks.

Default: true

`--path_to_setscheduler_binary`

Path to setscheduler binary, which is used to set realtime priority for fusion io driver.

Default: `/home/nutanix/bin/setscheduler`

`--skip_metadata_link_setup`

Skip creation of metadata links (Use rootfs for storing metadata

Default: false

`--skip_scsi_bus_rescan`

Skip rescanning scsi bus while running disk prep

Default: false

`--striped_models_csv`

A comma-separated list of disk models that should be striped together to form a single logical device.

Default: `SSD_910_200GB`

cluster.genesis.breakfix.host_bootdisk_breakfix

`--timeout_for_repaired_node_to_join_healthy_genesis_list`

Waiting period for repaired node to join healthy genesis list

Default: 1800

cluster.genesis.breakfix.host_bootdisk_graceful

`--clone_bootdisk_default_timeout`

The default timeout for completion of cloning of bootdisk.

Default: 28800

`--restore_bootdisk_default_timeout`

The default timeout for completion of restore of bootdisk.

Default: 14400

`--wait_for_phoenix_boot_timeout`

The maximum amount of time for which the state machine waits after cloning for the node, to be booted in phoenix environment.

Default: 36000

cluster.genesis.breakfix.host_bootdisk_utils

`--host_boot_timeout`

The maximum amount of time for which the state machine waits for host to be up.

Default: 36000

cluster.genesis.breakfix.ssd_breakfix

`--ssd_repair_copy_svmrescue_timeout`

Timeout for copying svmrescue.iso from CVM to host

Default: 900

cluster.genesis.breakfix.ssd_breakfix_esx_helper

`--svm_regex`

Regular expression used to find the SVM vmx name.

Default: `ServiceVM`

cluster.genesis.convert_cluster.cluster_conversion

`--relax_cluster_ahv_version_check`

Relax cluster's AHV version check which makes sure all nodes(including storage-only nodes) will have same AHV version after conversion. With the flag we allow a maximum of 2 count of AHV versions in the cluster.

Default: false

cluster.genesis.expand_cluster.expand_cluster

`--average_latency`

Average latency threshold

Default: 2

`--expand_ssh_retry_timeout`

Time till which the retries by the ssh_client will be performed.

Default: 300

`--network_bandwidth_checks_enabled`

Enable / Disable for network bandwidth checks

Default: true

`--node_up_retries`

Number of retries for node genesis rpcs to be up after reboot

Default: 40

`--packet_loss`

Packet loss %

Default: 0

`--pending_restart_retry_interval`

Interval after which the node should retry volunteering for cluster expansion

Default: 300

`--volunteer_check_retry_interval`

Interval after which the node should retry volunteering for cluster expansion

Default: 60

cluster.genesis.expand_cluster.pre_expand_cluster_checks

`--cassandra_metadata_failure_check_threshold`

Threshold for the cassandra metadata failure check.

Default: 3

cluster.genesis.hibernate_resume.hibernate_cluster

`--hibernate_cluster_timeout`

Hibernate cluster statemachine timeout.

Default: 8640000

cluster.genesis.hibernate_resume.resume_cluster

`--resume_cluster_timeout`

Resume cluster statemachine timeout.

Default: 8640000

cluster.genesis.network_configuration_keeper.network_configuration_keeper

`--network_collector_interval`

Interval at which network collector thread collects network related data.

Default: 1800

cluster.genesis.service_management.service_mgmt_utils

`--core_services_managed`

This Flag will be used to force service mgmt to enable/disable core services.

Default: false

cluster.hades.disk_diagnostics

`--disk_health_monitor_interval_secs`

Time interval in seconds for monitoring disk health.

Default: 300

`--disk_health_monitor_mark_disks_bad`

Allow disk health monitor to mark disks bad if the health check fails.

Default: true

`--disk_health_monitor_smart_checks_count`

Number of health checks to run before confirming a disk is bad.

Default: 2

`--enable_disk_health_monitor`

Enable periodic disk health monitoring.

Default: true

`--hades_retry_count`

Default retry count.

Default: 5

`--max_disk_offline_count`

Maximum error count for disk after which disk is to removed.

Default: 3

`--max_disk_offline_timeout`

Maximum time value where disk offline events areignored.

Default: 3600

cluster.hades.disk_error_parser.consts

`--disk_error_parse_interval_secs`

Periodicity with which kernel logs need to be checkedto detect disk errors.

Default: 300

`--enable_disk_error_parser`

Enable disk error parser.

Default: true

cluster.hades.disk_manager

`--aws_cores_partition`

Partition in which core files are stored on AWS.

Default: `/dev/xvdb1`

`--clean_disks_pe_pool_size`

How many disks to clean in parallel.

Default: 12

`--default_max_sectors_kb`

Default value for max_sectors_kb size for disks is set to512, unless SSD block_params in hcl.json overrides it

Default: `512`

`--device_mapper_name`

A name of the device mapper that is to be created in case striped devices are discovered.

Default: `dm0`

`--disk_unmount_retry_count`

Number of times to retry unmounting the disk.

Default: 60

`--experimental_delay_disk_add_udev_event_max_secs`

Maximum time in seconds to delay Hades handling of disk add udev events.

Default: 0

`--experimental_delay_disk_remove_udev_event_max_secs`

Maximum time in seconds to delay Hades handling of disk remove udev events.

Default: 0

`--firmware_upgrade_default_wait`

Default wait time after issuing firmware upgrade.

Default: 10

`--force_disable_xtrim_disk_models`

List of disk models to disable xtrim. The setting in hcl.json is ignored for these disks.

`--force_enable_xtrim_disk_models`

List of disk models to enable xtrim. The setting in hcl.json is ignored for these disks.

`--hdparm_path`

Location of hdparm.

Default: `/sbin/hdparm`

`--populate_motherboard_info`

Hades populates motherboard information in config proto. This is cloud substrate dependent, this flag can be used to skip populating motherboard information and continuing Hades configuration correctly.

Default: true

`--run_wipefs`

Clear all filesystem signatures from disk before creating filesystem

Default: true

`--sas3flash_location`

Location of sas3flash utility.

Default: `/home/nutanix/cluster/lib/lsi-sas/sas3flash`

`--sas3flash_target_runtime`

Approximate time taken by sas3flash for upgrading HBA

Default: `300`

`--sas_disk_firmware_upgrade_retry_count`

Numter of times to retry SAS disk firmware upgrade.

Default: 60

`--sata_disk_firmware_upgrade_retry_count`

Default retry count for firmware upgrade.

Default: 12

`--sg_start_path`

Location of sg_start binary.

Default: `/usr/local/nutanix/cluster/lib/sg3utils/bin/sg_start`

`--sg_write_buffer_path`

Location of sg_write_buffer. The version of the current binary used is 1.18 20141107.

Default: `/usr/local/nutanix/cluster/lib/sg3utils/bin/sg_write_buffer`

`--skip_disk_remove_reboot`

Skip reboot for bad or removed disk.

Default: false

`--skip_disk_remove_reboot_marker`

Skip reboot marker for bad or removed disk.

Default: `/home/nutanix/.skip_disk_remove_reboot_marker`

`--ssd_default_rules_path`

Location of the SSD udevadm rules file.

Default: `/etc/udev/rules.d/999-defaultssd.rules`

`--ssd_read_ahead_size_kb`

Read ahead size in KB for SSDs. Min: 0 Max: 256 KB

Default: 16

`--ssd_rules_path`

Location of the SSD udevadm rules file.

Default: `/etc/udev/rules.d/99-ssd.rules`

`--stargate_directory`

Directory where stargate disks should be mounted.

Default: `/home/nutanix/data/stargate-storage`

`--unmount_default_wait_sec`

Default wait time for unmount during disk remove.

Default: 5

`--unmount_timeout`

Default wait time for unmounting to succeed.

Default: 120

cluster.hades.raid_utils

`--min_raid_sync_speed`

Minimum raid sync speed

Default: 50000

cluster.hyperv_upgrade

`--vmms_setting_info`

Location where the vmms settings are stored

Default: `/appliance/logical/genesis/vmms_setting_info`

cluster.rsyslog_helper

`--ahv_rsyslog_queue_memory_size`

Size of rsyslog remote logging action queue in bytes.

Default: 104857600

`--audit_log_dir`

Default path for prism audit log files.

Default: `/home/nutanix/data/prism`

`--lock_dir`

Default path for nutanix lock files.

Default: `/home/nutanix/data/locks/`

`--log_dir`

Default path for nutanix log files.

Default: `/home/nutanix/data/logs`

`--module_level`

Level of syslog used for sending module logs

Default: `local0`

`--read_timeout_in_sec`

Time in seconds after which the logs will be forced to syslog server, if no new message is seen.

Default: 10

`--rsyslog_conf_dir`

Default directory to store rsyslog related configs.

Default: `/etc/rsyslog.d`

`--rsyslog_conf_file`

Default Configuration file for Rsyslog service.

Default: `/etc/rsyslog.d/rsyslog-nutanix.conf`

`--rsyslog_configure_queue`

Boolean indicating whether to configure action queue for rsyslog.

Default: true

`--rsyslog_queue_memory_size`

Size of rsyslog remote logging action queue in bytes.

Default: 104857600

`--rsyslog_rule_header`

Nutanix specified rsyslog rules are appended only below this marker.

Default: `# Nutanix remote server rules`

`--rsyslog_rule_header_end`

Nutanix specified rsyslog rules are added above this marker.

Default: `# Nutanix remote server rules end`

`--rsyslog_work_dir`

Default path for syslog state files. This stores thestate of rsyslog across restarts. We Will create this folder if path is in /home/nutanix

Default: `/home/nutanix/rsyslog`

`--syslog_module`

Dummy module name to denote non nutanix default syslog.

Default: `syslog_module`

cluster.service.acropolis_service

`--acropolis_cgroup_memory_limit_mb`

If -1, creation of cgroup with memory limit is disabled for this component. If specified as positive number then a cgroup is created for this component with this set as the memory limit in megabytes.

Default: 3640

`--acropolis_path`

Path to the acropolis binary.

Default: `/home/nutanix/bin/acropolis`

`--host_connection_retry_count`

Number of retries if host agent connection fails.

Default: 2

cluster.service.cluster_config_service

`--cluster_config_memory_limit_mb`

If -1, creation of cgroup with memory limit is disabled for this component. If specified as positive number then a cgroup is created for this component with this set as the memory limit.

Default: 320

`--cluster_config_path`

Path to the Cluster Config binary.

Default: `/home/nutanix/bin/cluster_config`

`--cluster_config_server_rss_mem_limit`

Maximum amount of resident memory Cluster Config may use on an Svm with 8GB memory configuration.

Default: 268435456

`--cluster_config_use_backplane_ip`

Whether to use the backplane IP for this service.

Default: true

`--cluster_config_use_service_proxy`

Whether to use the service proxy to run this service.

Default: true

cluster.service.ha_service

`--def_stargate_stable_interval`

Default number of seconds a stargate has to be alive to be considered asstable and healthy.

Default: 30

`--hyperv_internal_switch_health_timeout`

Timeout for how long we should wait before waking up the thread that monitors internal switch health on HyperV.

Default: 30

`--num_worker_threads`

The number of worker threads to use for running tasks.

Default: 8

`--old_stop_ha_zk_node`

When this node is created the old ha should not take any actions on the cluster.

Default: `/appliance/logical/genesis/ha_stop`

`--stargate_aggressive_monitoring_secs`

Default number of seconds a stargate is aggressively monitored after it is down.

Default: 3

`--stargate_aggressive_monitoring_two_node_multiplier`

Multiplier to the --stargate_aggressive_monitoring_secsfor the two node cluster.

Default: 4

`--stargate_exit_handler_aggressive_timeout_secs`

Aggressive timeout for accessing the Stargate exit handler page during an unplanned failover.

Default: 3

`--stargate_exit_handler_timeout_secs`

Default timeout for accessing the Stargate exit handler page.

Default: 10

`--stargate_health_watch_timeout`

Timeout for how long we should wait before waking up the thread that monitors stargate health.

Default: 30

`--stargate_initialization_secs`

Number of seconds to wait for stargate to initialize.

Default: 30

`--stop_ha_zk_node`

When this node is created ha 2.0 should not take any actions on the cluster.

Default: `/appliance/logical/genesis/ha_2_stop`

`--tcpkill_el6_path`

Path to el6 tcpkill binary.

Default: `/home/nutanix/cluster/lib/el6/tcpkill`

`--tcpkill_log`

Path to tcpkill binary.

Default: `/home/nutanix/data/logs/tcpkill.log`

`--tcpkill_path`

Path to el7 tcpkill binary.

Default: `/home/nutanix/cluster/lib/el7/tcpkill`

cluster.service.hyperint_service

`--hyperint_esx_connect_max_retries`

Maximum number of retries for connecting to esx hosts.Value of 0 indicates unlimited retries.

Default: 0

`--hyperint_esx_connect_retry_time_in_secs`

Retry delay for connecting to esx hosts.

Default: 10

`--hyperint_esx_one_time_password_script`

Path to helper script for obtaining one time password on ESX

Default: `/home/nutanix/cluster/lib/esx5/get_one_time_password.py`

`--hyperint_memory_limit_mb`

If -1, creation of cgroup with memory limit is disabled for this component. If specified as positive number then a cgroup is created for this component with this set as the memory limit.

Default: 500

`--hyperint_monitor_path`

Path to the Hyperint monitor binary.

Default: `/home/nutanix/bin/hyperint_monitor`

`--hyperint_rss_mem_limit`

Maximum amount of resident memory Hyperint may use on an Svm with 8GB memory configuration.

Default: 67108864

cluster.service.kafka_service

`--kafka_bootstrap_binary`

Path to binary which will start kafka on PC in docker container

Default: `/usr/local/nutanix/bin/bootstrap_kafka`

`--kafka_data_volume_mode`

Volume mode to provide data volume to kafka. If host attached, vmdisk is attached to PC VM. Else, docker volume plugin is used to conenct to remote PE cluster iscsi endpoint

Default: `host_attached`

`--kafka_rss_mem_limit`

Maximum amount of resident memory Kafka may use on an svm with 8 GB memory configuration.

Default: 268435456

cluster.service.snmp_service

`--snmp_manager_path`

Path to the snmp_manager binary.

Default: `/home/nutanix/cluster/bin/snmp_manager`

cluster.two_node.cluster_manager

`--block_transitions_if_cluster_stopped`

If true, don't transition to standalone mode if cluster is in stopped state

Default: true

`--ensure_set_two_node_leader_conditions`

If true, ensure conditions are met before manually setting two node leader

Default: true

`--minimum_pass_pings_to_allow_fails`

Default minimum number of pings to peer node that must pass before we allow some pings to fail (controlled by number_of_ping_fails_to_reset_pass_counter flag) and not reset pass counter.

Default: 10

`--minimum_uptime_mins_before_starting_pings`

Minimum system uptime before starting pinging the peer node.

Default: 4

`--minutes_of_successful_pings_for_transition`

Default number of minutes for successful pings beforeconsidering peer node stable and moving to kSwitchToTwoNode mode.

Default: 7

`--node_health_check_ping_interval_secs`

Default time (in seconds) for which health thread sleeps before checking on another node.

Default: 2

`--non_leader_get_cluster_op_mode_from_zk_retries`

Number of retries a non-leader node will do get cluster operation mode from Zookeeper after initiating as non-leader from witness.

Default: 3

`--number_of_ping_fails_to_reset_pass_counter`

Default number of pings to peer node that should failto reset the pass counter.

Default: 4

`--retry_count_for_kSwitchToTwoNode`

Default number of retries for successful initialization in kSwitchToTwoNode state.

Default: 0

`--rpc_retry_timeout_seconds`

Default time till when to keep retrying RPCs to peer node before claiming failure.

Default: 600

`--seconds_of_successful_pings_during_planned_outage`

Default number of seconds to wait before initiating transition for planned outages.

Default: 8

`--seconds_to_wait_before_checking_ha_forwarding`

Default number of seconds to wait before checking again if host is forwarding storage traffic.

Default: 30

`--seconds_to_wait_before_pinging_upgrading_node`

Default number of seconds to wait before pinging a node that separated due to upgrade.

Default: 5

`--seconds_to_wait_before_retry_for_kSwitchToTwoNode`

Default number of seconds to wait before retrying for successful initializaton in kSwitchToTwoNode state.

Default: 2

`--threshold_number_of_consecutive_failed_pings`

Default threshold for number of times a ping to a nodecan fail before it is assumed down.

Default: 4

`--transition_to_standalone_due_to_degraded_cassandra_on_peer`

If True, allow transitions to kStandAlone if peer cassandra is degraded. transition_to_standalone_due_to_degraded_service_on_peer also needs to be true to support this.

Default: true

`--transition_to_standalone_due_to_degraded_service_on_peer`

If True, allow transitions to kStandAlone if peer services are degraded. This gflag is only used to control global support for transition due to any degraded services.

Default: true

`--transition_to_standalone_due_to_degraded_stargate_on_peer`

If True, allow transitions to kStandAlone if peer stargate is degraded. transition_to_standalone_due_to_degraded_service_on_peer also needs to be true to support this.

Default: true

`--unit_test_mode`

If True, running in unit test mode

Default: false

cluster.two_node.witness_manager

`--enable_sync_up_logical_ts_between_cluster_and_witness`

If True, will sync up logical timestamp between witness vm and cluster automatically once out of sync be detected in ping_witness

Default: true

`--sync_up_logical_ts_retrieval_delay_msecs`

Delay between sync up logical timestamp between witness VM and cluster retrieval attempts.

Default: 1000

`--witness_logical_timestamp_sync_up_retries`

Number of retry times for the cluster logical timestamp sync up between witness VM and cluster

Default: 5

`--witness_ping_interval_seconds`

Frequency (in seconds) to check the availability of the witness VM.

Default: 60

`--witness_refresh_cluster_name`

If True, will check whether to update the witness with the current cluster info.

Default: true

cluster.utils.device_mapper_utils

`--stripe_size_sectors`

Stripe chunk size in number of 512-byte sectors.

Default: 128

cluster.utils.hyperv_ha_utils

`--default_internal_switch_monitor_interval`

Default polling period for monitoring internal switch health

Default: 30

`--default_ping_success_percent`

Default percentage of success which is used to determine switch health

Default: 100

`--default_total_ping_count`

Default number of pings sent to determine switch health

Default: 10

cluster.utils.new_node_nos_upgrade

`--stand_alone_upgrade_log`

Log file for stand-alone upgrade.

Default: `/home/nutanix/data/logs/stand_alone_upgrade.out`

cluster_config.client.flags

`--auto_enable_cmsp`

Flag to control if CMSP should be auto enabled.

Default: false

`--cassandra_disk_size_mb_large_pc`

Large PC Disk size in MB for Cassandra

Default: 2560000

`--cassandra_disk_size_mb_small_pc`

Small PC Disk size in MB for Cassandra

Default: 512000

`--cassandra_disks_base_dir`

Base directory for the Cassandra disks mount path

Default: `/home/nutanix/data/stargate-storage/disks`

`--cluser_ssp_config_migrate_workers`

Number of cluster ssp config execution threads

Default: 2

`--cluster_config_download_recovery_timeout_in_sec`

Max seconds since download task creation to resume the task.

Default: 86400

`--cluster_config_ha_leadership_timeout_secs`

Timeout to wait for leadership to occur.

Default: 30

`--cluster_config_ha_quorum_time_secs`

Time to sleep for cluser config to form quorum.

Default: 10

`--cluster_config_ha_zeus_retry_count`

Number of times to retry getting zeus leader/contenders

Default: 6

`--cluster_config_image_poll_timeout`

Timeout in seconds for image to be migrated

Default: 300

`--cluster_config_list_max_length`

The maximum length for list.

Default: 300

`--cluster_config_poll_download_task_timeout_in_sec`

The default maximum poll download task timeout in seconds.

Default: 120

`--cluster_config_poll_prism_central_deployment_task_timeout_in_sec`

The default maximum poll deployment task timeout in seconds for prism central deployment.

Default: 3600

`--cluster_config_poll_prism_central_download_task_timeout_in_sec`

The default maximum poll download task timeout in seconds for prism centraldeployment.

Default: 7200

`--cluster_config_poll_remote_pc_deployment_task_timeout_in_sec`

The default maximum poll task timeout in seconds for remote prism central deployment.

Default: 12000

`--cluster_config_poll_task_timeout_in_sec`

The default maximum poll task timeout in seconds.

Default: 3600

`--cluster_config_query_upgrade_task_timeout_in_sec`

The default maximum poll download task timeout in seconds.

Default: 3600

`--cluster_config_retry_max_delay_ms`

Maximum delay in milliseconds.

Default: 2000

`--cluster_config_retry_max_retries`

Maximum number of retries, or None for infinite retries

Default: 3

`--cluster_config_retry_slot_time_ms`

Delay slot time in milliseconds.

Default: 500

`--cluster_download_workers`

Number of cluster download execution threads

Default: 5

`--cmsp_auto_enable_gateway_address`

Gateway address for use by CMSP auto enablement.

Default: `192.168.5.1`

`--cmsp_auto_enable_ip_block_end`

End IP for use by CMSP auto enablement.

Default: `192.168.5.64`

`--cmsp_auto_enable_ip_block_start`

Start IP for use by CMSP auto enablement.

Default: `192.168.5.2`

`--cmsp_auto_enable_pc_domain_name`

PC domain name for use by CMSP auto enablement.

Default: `prism-central.cluster.local`

`--cmsp_auto_enable_retry_delay_secs`

Retry delay in seconds for auto enabling CMSP.

Default: 300

`--cmsp_auto_enable_subnet_mask`

Subet mask for use by CMSP auto enablement.

Default: `255.255.255.0`

`--cmsp_auto_enablement_timeout_secs`

Timeout for use by CMSP auto enablement.

Default: 3600

`--enable_pc_multivdisk_deployment`

Enable multivdisk deploy workflow on PC

Default: true

`--enable_pc_multivdisk_deployment_on_large_pc`

Enable multivdisk deploy workflow on large PC

Default: true

`--enable_pc_multivdisk_deployment_on_small_pc`

Enable multivdisk deploy workflow on small PC

Default: false

`--general_service_workers`

Number of execution threads for general use

Default: 5

`--largest_schedule_gap_in_days`

The largest number of days that start time can be scheduled ahead in the future.

Default: 30

`--least_schedule_gap_in_mins`

The Least number of minutes that start time can be scheduled ahead in the future.

Default: 10

`--least_time_out_in_hours`

The least number of hours that schedule time out could be.

Default: 1

`--max_cassandra_data_disks`

Maximum number of Cassandra data disks

Default: 4

`--mcc_server_host`

Default server host.

Default: `127.0.0.1`

`--mcc_server_port`

Default server port.

Default: 9461

`--multicluster_config_actions_workers`

Number of multicluster config actions including cancel, pause execution threads

Default: 3

`--multicluster_config_workers`

Number of multicluster config execution threads

Default: 6

`--nucalm_container_name`

Name of the container which will be created to install volume plugin for enabling Nucalm.

Default: `NutanixManagementShare`

`--num_create_retries`

Number of retries to create the disks

Default: 3

`--num_disks_to_create`

Number of additional disks to be created

Default: 3

`--pc_deploy_workers`

Number of prism central deployment execution threads

Default: 3

`--resize_vm_task_timeout_secs`

Timeout in seconds for resizing all PC VMs.

Default: 300

`--storage_container_uuid`

Container UUID on which the disk will be created

`--wait_for_pe_pc_sync_secs`

Time to wait for the PE-PC data sync

Default: 300

`--watch_for_pe_registration`

Watch for host PE to be registered to PC, to initiate the disk addition

Default: true

util.cluster.consts

`--authn_ca_bundle`

File name of the CA bundle that should be used for authentication.

Default: `ca.pem`

`--authn_certs_root_path`

Base location of the certificates used in service authentication.

Default: `/home/certs`

`--authn_ica`

Name of the service authn intermediate ca file.

Default: `ica.crt`

`--authn_ica_key`

Name of the service authn intermediate ca private key.

Default: `ica.key`

`--authn_ica_req`

Name of the service authn intermediate ca file CSR.

Default: `ica.csr`

`--authn_root_ca`

Name of the service authn root ca file.

Default: `root.crt`

`--authn_root_ca_key`

Name of the service authn root ca private key.

Default: `root.key`

`--authn_zk_base_path`

Zookeeper base path for authn certificates.

Default: `/appliance/logical/auth/`

`--authorized_certs_file`

Path to file containing list of permitted SSL certs.

Default: `/home/nutanix/ssh_keys/AuthorizedCerts.txt`

`--azure_cert_dir`

Directory in which Azure certificates are stored.

Default: `/home/cloud/azure`

`--cloud_credentials_zkpath`

Zookeeper node path to the cloud credentials node.

Default: `/appliance/logical/cloud_credentials`

`--default_host_password`

Default password for the hypervisor.

Default: `nutanix/4u`

`--default_remote_shell_receive_timeout_secs`

The default timeout for idle connections after which the connection is terminated if there is no activity.

Default: 3600

`--default_remote_shell_socket_timeout_secs`

Timeout for the socket connecting to NutanixHostAgent.

Default: 300

`--default_remote_shell_timeout_secs`

The default timeout for completion of a powershell command made by calling the execute method.

Default: 3600

`--dell_ptagent_password`

Password for dell ptagent user 'ptuser'

Default: `Dellam123`

`--dell_ptagent_port`

Port of dell PTagent REST service

Default: `8086`

`--dell_ptagent_user_name`

User name for dell ptagent

Default: `ptuser`

`--disable_hades_marker_path`

Path of the marker to disable Hades.

Default: `/home/nutanix/.disable_hades`

`--disable_ssh_client_multiplexing`

Disable SSHClient multiplexing for the service.

Default: false

`--disk_location_json_path`

Path to the disk_location.json file.

Default: `/etc/nutanix/disk_location.json`

`--factory_config_json_path`

Path to the factory_config.json file.

Default: `/etc/nutanix/factory_config.json`

`--host_ssh_key`

Location of ssh key for accessing the CVM's host.

Default: `/home/nutanix/ssh_keys/host`

`--host_ssl_cert`

Location of ssl cer for accessing the CVM's Hyperv host.

Default: `/home/nutanix/ssh_keys/host.cer`

`--hyperv_hypervisor_username`

The username to use when logging into the local Hyper-V node.

Default: `Administrator`

`--hyperv_ipmicfg_relative_path`

Path of the ipmicfg utility in a Hyper-V node relative to the Nutanix directory.

Default: `\ipmicfg\IPMICFG-Win.exe`

`--hyperv_nutanix_path`

The path of the Nutanix directory on the Hyper-V host.

Default: `${env:ProgramFiles}\Nutanix`

`--hypervisor_ahv_nonroot_support`

Flag to enable non root user(nutanix) for cvmto AHV ssh command execution.

Default: false

`--hypervisor_esxi_smartctl_path`

Path to the smartctl binary on ESX.

Default: `/smartctl_cmd`

`--hypervisor_esxi_smartctl_wrapper_path`

Path to the smartctl wrapper script on ESX.

Default: `/smartctl`

`--hypervisor_internal_ip`

Internal IP address of the hypervisor.

Default: `192.168.5.1`

`--hypervisor_ipmicfg_menu_path`

Path to the ipmicfg Menu.dat on hypervisor.

Default: `/Menu.dat`

`--hypervisor_ipmicfg_path`

Path to the ipmicfg binary on hypervisor.

Default: `/ipmicfg`

`--hypervisor_username`

The username to use when logging into the local hypervisor.

Default: `root`

`--ipmi_password`

The password to use when logging into the local IPMI device.

Default: `ADMIN`

`--ipmi_username`

The username to use when logging into the local IPMI device.

Default: `ADMIN`

`--ledctl_location`

Location of ledctl utility.

Default: `/usr/sbin/ledctl`

`--local_esxi_smartctl_path`

Local path to the smartctl binary for ESX.

Default: `/usr/local/nutanix/bootstrap/lib/smartctl.esx`

`--local_esxi_smartctl_wrapper_path`

Local path to the smartctl wrapper for ESX.

Default: `/usr/local/nutanix/bootstrap/lib/smartctl.esx.sh`

`--local_ipmicfg_menu_path`

Path to the ipmicfg Menu.dat on the local host.

Default: `/usr/local/nutanix/cluster/lib/esx5/Menu.dat`

`--local_ipmicfg_path`

Path to the ipmicfg binary on the local host.

Default: `/usr/local/nutanix/cluster/lib/esx5/IPMICFG-Linux.x86_64`

`--maintenance_mode_history_file`

The path of maintenance mode history file which records the time when CVM entered maintenance mode.

Default: `/home/nutanix/config/maintenance_mode.history`

`--megacli64_location`

Location of MegaCli64 utility.

Default: `/usr/local/nutanix/cluster/lib/MegaCli/MegaCli64`

`--nested_esx_marker_path`

Is this nested ESX?

Default: `/home/nutanix/.nested_esx_marker`

`--node_ssh_key`

Location of ssh key for accessing remote CVMs and hosts.

Default: `/home/nutanix/.ssh/id_rsa`

`--node_ssl_cert`

Location of ssl cer for accessing remote Hyperv hosts.

Default: `/home/nutanix/.ssh/id_rsa.cer`

`--nutanix_host_agent_port`

The port on which the NutanixHostAgent service listens on the Hyper-V host.

Default: 3071

`--nutanix_log_dir`

Directory containing logs.

Default: `/home/nutanix/data/logs`

`--prism_monitor_port`

Default port used by Prism Monitor.

Default: 2019

`--sas3ircu_location`

Location of sas3ircu utility.

Default: `/home/nutanix/cluster/lib/lsi-sas/sas3ircu`

`--ssh_client_controlpersist`

Specifies the duration for which the master ssh connection would remain open after the last commandexecution.

Default: `15m`

`--storcli64_location`

Path to storcli utility.

Default: `/home/nutanix/cluster/lib/storcli/storcli64`

`--svm_non_ha_internal_ip`

Internal IP address of SVM that is not redirected for HA.

Default: `192.168.5.254`

`--vcenter_info`

zk node of host to vcenter info

Default: `/appliance/logical/genesis/vcenter_info`

`--zeus_config_cache_path`

Path to the zeus configuration cache kept in sync by Genesis.

Default: `/home/nutanix/config/configuration_proto.dat`

`--zkmigration_wal_path`

Path to the zookeeper migration local WAL.

Default: `/home/nutanix/data/zookeeper_monitor/zookeeper_migration.wal`

`--zkquorum_change_wal_path`

Path to the zookeeper quorum change local WAL.

Default: `/home/nutanix/data/zookeeper_monitor/quorum_change.wal`

`--zookeeper_timeout_secs`

Zookeeper connection timeout in seconds.

Default: 60

util.ndb.infrastructure.cluster

`--infra_service_vm_config_json_path`

Path to the service_vm_config.json file with the svm id.

Default: `/home/nutanix/data/stargate-storage/service_vm_config.json`

genesis

Usage

```
Usage: /usr/local/nutanix/cluster/bin/genesis start|stop [all|<service1> [<service2> ...]]|
restart|status
```

/usr/local/nutanix/cluster/bin/genesis

`--force`

Flag to indicate if services shouldbe forcefully stopped

Default: true

`--foreground`

Run Genesis in foreground.

Default: false

`--genesis_debug_stack`

Flag to indicate whether signal handler need to be registered for debugging greenlet stacks.

Default: true

`--genesis_self_monitoring`

Genesis to do self monitoring.

Default: true

`--genesis_upgrade`

Flag to indicate that genesis restarted because it is upgrading itself.

Default: false

`--help`

show this help

Default: 0

`--helpshort`

show usage only for this module

Default: 0

`--helpxml`

like --help, but generates XML output

Default: false

`--version`

Print version info and exit

Default: None

cluster.client.genesis.node_manager_helper

`--auto_discovery_interval_secs`

Number of seconds to sleep when local node can't join any discovered cluster.

Default: 5

`--co_nodes_unconfigure_marker`

Path to marker file to indicate that node has to unconfigure CO nodes as part of unconfiguring itself. The contents of the marker file containsspace seperated IPs of the CO nodes to unconfigure

Default: `/home/nutanix/.co_nodes_unconfigure`

`--discovery_timeout_secs`

Timeout for discovery.

Default: 1

`--download_staging_area`

Directory where we will download directories from other SVMs.

Default: `/home/nutanix/tmp`

`--fault_tolerance_initialization_max_retries`

The maximum number of retries to commit modified proto to zeus. Retry indefinitely if set to -1.

Default: 30

`--firmware_disable_auto_upgrade_marker`

Path to marker file to indicate that automatic firmware upgrade should not be performed on this node.

Default: `/home/nutanix/.firmware_disable_auto_upgrade`

`--foundation_disable_auto_upgrade_marker`

Path to marker file to indicate that automatic foundation upgrade should not be performed on this node.

Default: `/home/nutanix/.foundation_disable_auto_upgrade`

`--genesis_restart_timeout`

Time we wait for the genesis to restart.

Default: 120

`--gevent_resolver_reinitialize_on_dns_change`

The flag determines if gevent ares dns resolver should be reinitialized on a DNS config change.

Default: true

`--gold_image_version_path`

Path to the file that contains the version of the gold image.

Default: `/etc/nutanix/svm-version`

`--hcl_znode_path`

Zookeeper node containing the hcl.

Default: `/appliance/physical/hcl`

`--mem_resolution_for_zk_proto_mb`

Minimum memory change of CVM to reflect in proto.

Default: 128

`--move_time_back`

The flag is set to allow time to be moved back by more than time_rollback_tolerance_secs.

Default: false

`--nagios_config_path`

Path to the nagios configuration file.

Default: `/home/nutanix/serviceability/config/nagios3/nutanix_nagios.cfg`

`--node_disable_auto_upgrade_marker`

Path to marker file to indicate that automatic software upgrade should not be performed on this node.

Default: `/home/nutanix/.node_disable_auto_upgrade`

`--node_ssh_key_dir`

Path for node specific ssh keys on local disk

Default: `/home/nutanix/ssh_keys/.blackbox`

`--node_unconfigure_marker`

Path to marker file to indicate that node is ready to be unconfigured.

Default: `/home/nutanix/.node_unconfigure`

`--num_parallel_services_stop`

Maximum number of services to stop in parallel.

Default: 32

`--rpm_genesis_log_file`

Path to rpm log file during Genesis self install.

Default: `/home/nutanix/data/logs/rpm.genesis.out`

`--safe_update_or_restart_retry_interval`

Retry interval to check whether genesis update or restart can be issued.

Default: 30

`--set_node_configured_time_timeout`

Timeout for setting the node_configured_time field inzeus config.

Default: 5

`--sshd_config_path`

Path to sshd config file.

Default: `/etc/ssh/sshd_config`

`--stop_service_zknode`

Start services upto but not including the service specified in this zk node.

Default: `/appliance/logical/genesis/stop_service`

`--stop_services_timeout`

Maximum time to wait for all services to stop.

Default: 120.0

`--svm_internal_ips`

Internal IP addresses on eth1. The first one is the primary IP address, and the remaining are aliases.

Default: `192.168.5.2,192.168.5.254`

`--svm_internal_netif_netmask`

The netmask for the internal IP addresses on eth1.

Default: `255.255.255.128`

`--svm_non_ha_internal_netmask`

The netmask for the non-data internal IP aliases on eth1.

Default: `255.255.255.0`

`--svm_upgrade_finish_zknode`

Zk path to keep track of the svm_upgrade script on two node clusters.

Default: `/appliance/logical/genesis/svm_upgrade_finish`

`--timezones_dir`

Directory where all of the valid timezones exist.

Default: `/usr/share/zoneinfo`

`--vlan_sniffer_log`

Path to vlan_sniffer log.

Default: `/home/nutanix/data/logs/vlan_sniffer.log`

`--vlan_sniffer_path`

Path to vlan sniffer proxy.

Default: `/home/nutanix/cluster/bin/vlan_sniffer_proxy`

`--zookeeper_start_timeout_secs`

Timeout for waiting on Zookeeper connection on startup.

Default: 120

cluster.client.genesis.ns.consts

`--client_for_kVolumes`

Client name for kVolumes, used to pass to Firewall

Default: `service_segmentation_kVolumes`

`--disable_wait_time`

Time to wait (in seconds), between removing network segmentation configuration in zeus and removing the interface configuration on cvms

Default: 5

`--hyperv_physical_backplane_segmentation_support`

Boolean to control whether physical backplane segmentation is supported or not on hyperv

Default: true

`--ip_pool_based_backplane_segmentation`

Boolean to control whether ip pool based backplane segmentation is suppported

Default: true

`--iser_supported_pci_model_ids`

PCI model ids of NICs on which iSER will be supported

Default: `15b3:1017`

`--max_interfaces_for_service_segmentation`

The maximum number of interfaces which can be created on the CVM for service segmentation

Default: 3

`--mixed_hypervisor_backplane_support`

Boolean to control whether backplane network support on mixed hypervisoris allowed or not

Default: true

`--ns_config_cache_path`

Path to the ns configuration cache

Default: `/home/nutanix/config/ns_configuration_proto.dat`

`--ns_state_machine_timeout`

The timeout for completion of network segmentation state machine.

Default: 600

`--ns_wait_for_master_ephemeral`

Ephemeral zknode used in network segmentation enable to ensure pre_reboot phase of rolling reboot is executed after genesis master

Default: `/appliance/logical/genesis/ns/wait_for_master`

`--ports_for_kVolumes`

Comma separated TCP ports needed by kVolumes

Default: `3205,3260`

`--rdma_supported_pci_model_ids`

PCI model ids of NICs on which RF2 RDMA will be supported

Default: `15b3:1013,15b3:1015,15b3:1017`

`--rdma_ztr_supported_pci_model_ids`

PCI model ids of NICs on which RF2 RDMA will be supported

Default: `15b3:1017`

`--retry_count_zk_map_publish`

Retry count for publishing new zk mapping.

Default: 3

`--revert_ns_config_on_failure`

Revert the network segmentation configuration in the case of a failure.

Default: true

`--service_segmentation_support_for_dr`

Boolean to control whether to enable/disable support for DR service segmentation

Default: true

`--zookeeper_ports_for_kVolumes`

Comma separated zookeeper TCP ports needed by kVolumes to be opened on CVM IPs only.

Default: `9876`

cluster.client.genesis.pb_client.client

`--genesis_service_host`

The genesis service host.

Default: `127.0.0.1`

`--genesis_service_port`

The genesis service port.

Default: 2100

cluster.client.genesis_utils

`--genesis_up_retries`

Number of retries for node genesis rpcs to be up after reboot

Default: 40

`--maintenance_mode_retry_count`

Number of times Genesis will attempt to put the host into maintenance mode.

Default: 3

`--min_nos_version_for_stargate_huge_page_support`

NOS version starting which huge pages are supported for Stargate

Default: `5.19`

`--orion_config_path`

Path to orion config

Default: `/appliance/logical/orion/config`

`--svm_default_login`

User name for logging into SVM.

Default: `nutanix`

`--sync_network_configuration`

Location in Zookeeper where a workflow can write to run an on demand network collection

Default: `/appliance/logical/genesis/network_configuration/sync_network_configuration`

`--timeout_HA_route_verification`

Timeout for setting HA route.

Default: 180

`--timeout_zk_operation`

Timeout for zk operation like write

Default: 120

`--upgrade_fail_marker`

Marker to indicate upgrade has failed.

Default: `/appliance/logical/genesis/upgrade_failed`

ncc

Usage

```
nutanix@cvm$ /home/nutanix/ncc/bin/ncc [flags]
```

/home/nutanix/ncc/bin/ncc

`--health_checks_and_log_collector`

Flag to run health_checks and log_collector together.

Default: false

`--help`

show this help

Default: 0

`--helpshort`

show usage only for this module

Default: 0

`--helpxml`

like --help, but generates XML output

Default: false

`--ncc_logging_dir`

Directory where script log files are stored.

Default: `/home/nutanix/data/logs`

`--ncc_version`

Show script version.

Default: false

`--show_version`

Show detailed NCC version.

Default: false

`--version`

Show script version.

Default: false

ncc.cluster_checker

`--auto_log_coll`

If true, ncc runs the periodic log collector and panacea.

Default: false

`--auto_log_coll_output_location`

The location where periodic log collector stores the log files.

Default: `/home/nutanix/data/periodic_log_collector`

`--ignore_frequency_check`

If set, NCC frequency check is ignored.

Default: false

`--max_ncc_lock_checks`

Maximum number of times to check if ncc lock is acquired.

Default: 5

`--ncc_email_subject`

Subject of the email to be sent.

Default: `NCC Email Digest`

`--ncc_failed_plugins_file`

Zookeeper path where the list of last failed plugins is stored.

Default: `/appliance/logical/serviceability/last_failed_plugins`

`--ncc_lock_sleep`

Sleep for these manys seconds if ncc lock is not acquired.

Default: 5

`--ncc_send_email`

If True, ncc tries to sends an email from Zookeeper leader if configured time constraints are met.

Default: false

`--ncc_task_timeout_secs`

The timeout for the ncc task.

Default: 600

`--send_email_force`

If set, NCC always sends an email.

Default: false

`--send_email_local`

If set, NCC sends emails from local node.

Default: false

`--zk_auto_log_coll_config`

Zk node having configuration for log collector and panacea runs.

Default: `/appliance/logical/analytics/log_collector_panacea`

`--zk_ncc_email`

Zk node having NCC email configuration.

Default: `/appliance/logical/analytics/ncc_email`

`--zk_ncc_lock`

Ephemeral node for acquiring NCC execution lock.

Default: `/appliance/logical/analytics/ncc_lock`

ncc.forker.forker

`--modify_fd_and_exec_binary_path`

location of the modify and exec binary.

Default: `modify_fd_and_exec`

ncc.ncc_utils.cluster_utils

`--hcl_file_path`

The path of file hcl.json on CVM.

Default: `/etc/nutanix/hcl.json`

`--hyperv_ncc_cmd_timeout_secs`

Timeout seconds for some commands run on HyperV

Default: 120

`--min_storage_for_high_density_nodes`

Min storage in TB for high density nodes.The value should be comma separated list of <AOS version>:<threshold in terabytes>

Default: `5.11:140,6.1:212`

`--prism_secure_key_zk_path`

Znode containing the PrismSecureKey.

Default: `/appliance/logical/prismsecurekey`

ncc.ncc_utils.command_executor

`--ncc_command_executor`

Flag that specifies which command executor needs to be used.

Default: false

ncc.ncc_utils.config_utils

`--ncc_config_dir`

Directory path where ncc config files are kept

Default: `/home/nutanix/ncc/config`

`--ncc_node_config_file_name`

Config file name for keeping node specific info

Default: `node_config.json`

ncc.ncc_utils.esx_utils

`--fix_high_perf_policy`

Fix high performance policy.

Default: False

`--ncc_resource_pool_min_size_in_mb`

Min amount of memory to be reserved for resource pool

Default: 256

`--ncc_resource_pool_on_esx`

Default resource pool name for running ncc cmds

Default: `host/vim/vmvisor/ntnx_ncc`

`--ncc_resource_pool_size_in_mb`

Max amount of memory to be used by resource pool

Default: 800

`--temp_cvm_vmx`

Temporary path to store the CVM vmx file.

Default: `/tmp/ServiceVM.vmx`

ncc.ncc_utils.hyperv_utils

`--ipmiutil_lock_file`

ipmiutil.exe does not behave correctly when multuple parallel instances are running at the same time. This file will be used for flock around ipmiutil.exe invocations to protect against concurrent NCC and cluster_health instances.

Default: `/tmp/.ipmiutil_lock`

`--ipmiutil_max_num_retries`

Max number of times ipmiutil is run to get error free output

Default: 10

`--ipmiutil_retry_interval`

Time in seconds after which ipmiutil is run again after an unsuccessful run

Default: 10

`--ncc_server_port`

Port that the HTTP server started by NCC listens on.

Default: 2101

ncc.ncc_utils.hypervisor_logs

`--copy_from_host_timeout`

Timeout used while copying log files from host.

Default: 300

ncc.ncc_utils.hypervisor_utils

`--nic_link_down_job_state`

Path to the job state for nic status.

Default: `/home/nutanix/data/serviceability/check_cvm_health_job_state.json`

`--nic_link_down_timeout`

Nic link down timeout. It means that if a nic link's down time is longer than the timeout time, this nic is regarded as disconnected and will be removed from the nic status list saved in nic status file, which means this nic will no longer be checked until its status becomes Up that makes it reenter the nic status list. The default timeout time is one day.If timeout time is 0, it means the nic will always be in the nic status list, thus always be checked, just like before.

Default: 86400

`--raw_sel_log_file_path`

Path to file storing raw sel log information.

Default: `/tmp/raw_sel.log`

ncc.ncc_utils.ipmi_stats

`--ipmi_lock_file`

ipmitool/ipmiutil does not behave correctly when multuple parallel instances are running at the same time. This file will be used for flock around ipmitool/ipmiutil invocations to protect against concurrent NCC and cluster_health instances.

Default: `/tmp/.ipmi_lock`

`--ipmi_max_num_retries`

Max number of times ipmitool/ipmiutil is run to get error free output

Default: 10

`--ipmi_retry_interval`

Time in seconds after which ipmitool/ipmiutil is run again after an unsuccessful run

Default: 10

ncc.ncc_utils.network_utils

`--ping_command_timeout_secs`

Ping command timeout secs for checking reachable hosts

Default: 10

ncc.plugins.log_collector.log_utils.hyperv_log

`--hyperv_cluster_logs`

Collect HyperV cluster logs

Default: false

util.ncc.config_module.config

`--auto_resolve_time_multiple`

The multiple of the auto resolve check time interval, for the wait time used to auto resolve an alert

Default: 6

`--template_checks_max_id`

The last ID to be used for template checks.

Default: 220000

`--template_checks_min_id`

The first ID to be used for template checks.

Default: 210000

util.ncc.data_access.insights_data_access

`--batch_entity_cnt`

Batch cnt of entities read and written to Entity DB

Default: 50

`--batch_entity_cnt_long_term`

Batch cnt of entities read and written to Entity DBfor long term data

Default: 10

`--num_intervals_for_latest_data`

The number of interval for latest data from IDF if not given time range.

Default: 12

`--overwrite_metric_schema`

If set we force overwrite the metric schema for capacity metrics.

Default: true

`--resample_hourly_data`

If we want to resample daily data using hourly values.

Default: true

`--updated_gap_in_hours`

The gap for updating the non-overwritten-historical data in hours.

Default: 0

util.ncc.ncc_async_task_service.async_task_interface.async_task_base_manager

`--use_ergon_interface`

Flag to specify if the code should use ergon as the task interface.

Default: true

util.ncc.ncc_logger

`--enable_plugin_wise_logging`

Enabling this flag will add plugin name to log record during plugin run

Default: true

util.ncc.ncc_utils.arithmos_interface.arithmos_interface_new

`--arithmos_update_interval`

Time in seconds between updates to arithmos.

Default: 30

`--cos_update_interval`

Interval measured in arithmos cycles during which check.overall_score is updated.

Default: 2

`--score_update_paused`

Score updates will be kept in memory until this flag is set to False again. Useful for testing.

Default: false

util.ncc.ncc_utils.arithmos_utils

`--cluster_cpu_usage_sampling_interval_sec`

Sampling interval for CPU usage for a cluster.

Default: 300

`--cluster_memory_usage_sampling_interval_sec`

Sampling interval for Memory usage for a cluster.

Default: 300

`--cpu_capacity_data_path`

The file containing the cpu capacity of a cluster in hz.

Default: `/home/nutanix/data/ncc/cpu_capacity.json`

`--cpu_usage_processed_data_path`

The file containing the cpu usage of a cluster in hz.

Default: `/home/nutanix/data/ncc/cpu_usage.json`

`--cpu_usage_raw_data_path`

The file containing the cpu usage of a cluster in cycles along with timestamp.

Default: `/home/nutanix/data/ncc/cpu_usage_raw.json`

`--is_compute_only_node_supported`

Turn on to support compute only node specific checks.

Default: true

`--memory_capacity_data_path`

The file containing the memory capacity of a cluster in bytes.

Default: `/home/nutanix/data/ncc/memory_capacity.json`

`--memory_usage_processed_data_path`

The file containing the Memory usage of a cluster in bytes.

Default: `/home/nutanix/data/ncc/memory_usage.json`

`--memory_usage_raw_data_path`

The file containing the memory usage of a cluster in bytes along with timestamp.

Default: `/home/nutanix/data/ncc/memory_usage_raw.json`

util.ncc.ncc_utils.checkpoint

`--ignore_checkpoints`

Whether to force execution even when previous checkpoints exists.

Default: true

util.ncc.ncc_utils.cluster_utils

`--cmd_timeout_secs`

Timeout seconds for commands run on ESX and CVMs

Default: 60

`--copy_timeout_secs`

Timeout seconds for file copy operations.

Default: 600

`--cvm_timeout_cmd`

Timeout for any command executed on cvms via ssh

Default: `/usr/bin/timeout -s 9 %d`

`--cvm_uname`

Username used to authenticate with cluster CVMs

Default: `nutanix`

`--dummy_target_for_local_ip`

Dummy ip to which socket is connected in order to get local ip address

Default: `1.1.1.1`

`--ncc_max_rss_mem_mb`

Maximum memory used by ncc for heap or data segment.

Default: 256

`--ping_cmd_timeout_secs`

Timeout secs for ping command to check alive nodes

Default: 10

`--zkserver_config_proto_path`

Path to the zookeeper server config proto in Zookeeper.

Default: `/appliance/logical/zkserver_config`

util.ncc.ncc_utils.gflags_definition

`--anonymize_output`

Flag to specify if the output of log collector should be anonymized.

Default: false

`--auto_log_coll_frequency_hour`

If set log collector and panacea will run periodically. E.g Run every 1hr: <--auto_log_coll_frequency_hour=1>

Default: 0

`--case_number`

The case number for which logs are collected. If specified, logs are stored in this directory on FTP server.

`--collect_all_hypervisor_logs`

Flag to specify if all /var/log/* files should be collected.

Default: true

`--collect_boot_log`

Flag to specify if the /var/log/boot.log should be collected.

Default: true

`--collect_cvm_service_log`

Flag to specify if ServiceVM_Centos.*.out from cvm vm should be collected.

Default: true

`--collect_cvm_vmware_log`

Flag to specify if vmware.log from cvm vm should be collected.

Default: true

`--collect_cvm_vmx_log`

Flag to specify if ServiceVM_Centos.vmx from cvm vm should be collected.

Default: true

`--collect_dbserver_log`

Flag to specify if the /var/log/openvswitch/ovsdb-server.log should be collected.

Default: true

`--collect_dmesg`

Flag to specify if the /var/log/dmesg should be collected.

Default: true

`--collect_fdm`

Flag to specify if the fdm log should be collected.

Default: true

`--collect_firstboot_log`

Flag to specify if first boot log of hypervior should be collected.

Default: true

`--collect_host_upgrade`

Flag to specify if the host_upgrade logs should be collected.

Default: true

`--collect_hostd`

Flag to specify if the hostd log should be collected.

Default: true

`--collect_installer_log`

Flag to specify if the /var/log/installer.log should be collected.

Default: true

`--collect_isci_redirector_log`

Flag to specify if /var/log/isci_redirector files should be collected.

Default: true

`--collect_kernel_log`

Flag to specify if the kernel log should be collected.

Default: true

`--collect_libvirtd_log`

Flag to specify if /var/log/libvirt/libvirtd.log should be collected

Default: true

`--collect_messages`

Flag to specify if the /var/log/messages should be collected.

Default: true

`--collect_microseg_logs`

If the flag is set to true, then only micro-segmentation related logs are collected

Default: false

`--collect_openvswitch_log`

Flag to specify if the /var/log/openvswitch/ovs-vswitchd.log should be collected.

Default: true

`--collect_qemu_files`

Flag to specify if /var/log/libvirt/qemu/* files should be collected.

Default: true

`--collect_root_install_log`

Flag to specify if /root/install.log* files should be collected.

Default: true

`--collect_root_post_install_log`

Flag to specify if /root/post-install*.log* files should be collected.

Default: true

`--collect_shell_log`

Flag to specify if the shell log should be collected.

Default: true

`--collect_vmkernel_psod_dumps`

Flag to specify if the vmkernel psod logs should be collected.

Default: true

`--collect_vmksummary`

Flag to specify if the vmksummary log should be collected.

Default: true

`--collect_vmkwarning`

Flag to specify if the vmkwarning log should be collected.

Default: true

`--collect_vobd`

Flag to specify if the vobd log should be collected.

Default: true

`--collect_vpxa`

Flag to specify if the vpxa log should be collected.

Default: true

`--collect_yum_log`

Flag to specify if /var/log/yum.log files should be collected.

Default: true

`--collector_plugin_timeout`

Flag to specify timeout in seconds for the collection of logs.

Default: `86400`

`--cvm_list`

List of CVMs on which the plugin should be run. E.g. if the cvm ips are 10.4.45.54 and 10.4.45.55: <--cvm_list="10.4.45.54,10.4.45.54">

`--delete_auto_log_coll_config`

If true, delete the periodic debug config

Default: false

`--delete_email_config`

If True, ncc email configuration is deleted.

Default: false

`--enable_log_bundling`

If set to false the node level log bundle files are preserved at node level

Default: true

`--end_time`

End time till when logs should be collected. Time format is : %Y/%m/%d-%H:%M:%S E.g. <--end_time=2022/01/27-13:56:39>

`--enhanced_log_collector`

Flag to specify if the files needs to be collected by enhanced log collector mode.

Default: false

`--force`

Boolean to specify if the logs should be written even if there is more than 90% disk space used.

Default: 0

`--force_run`

The flag to run the server even when cluster is down.

Default: true

`--force_run_auto_log_coll`

If True, auto log collector is triggered on this node immediately without any checks.

Default: false

`--ftp_password`

Password of the server where files are to be uploaded. E.g. if password is 123: <--ftp_password=123>

Default: `nutanix/4u`

`--ftp_server`

Nutanix FTP server where the collected logs should be uploaded. E.g.: if server name is my_server: <--ftp_server=my_server>

Default: `ftp.nutanix.com`

`--ftp_target_location`

Target location where the files are uploaded, the location path is relative to home dir of the sftp/ftp user, absolute paths are not allowed e.g. --ftp-target_location='abc/def', then the files are uploaded to ~/abc/def/LogCollector path

`--ftp_timeout`

The timeout for ftp plugin

Default: 5400

`--ftp_username`

Username of the server where files are to be uploaded. E.g. if username is 123: <--ftp_username=123>

Default: `inbound`

`--hyperv_log_level`

List of log levels to be collected on Hyper-v E.g. collect all CRITICAL and ERROR: <--hyperv_log_level=CRITICAL,ERROR>. Valid values:CRITICAL, ERROR, WARNING, INFO, VERBOSE.

Default: CRITICAL,ERROR,WARNING,INFO,VERBOSE

`--last_no_of_days`

Number of days for which logs should be collected. E.g. collect 3 days of logs: <--last_no_of_days=3>

Default: 0

`--last_no_of_hours`

Number of hours for which logs should be collected. By default, 4 hrs of logs are collected. E.g. collect 3 hours of logs: <--last_no_of_hours=3>

Default: 4

`--min_collector_plugin_timeout`

Flag to specify timeout in seconds for the collection of logs.

Default: 3600

`--ncc_output_printer`

Set the printer for the NCC run.

Default: `plugin`

`--not_unzip_file`

Flag to specify if the .gz files will be decompressed.

Default: true

`--parallel`

The flag specifies the number of plugins to be run concurrently. The value should be at most 4.

Default: 1

`--plugin_list`

List of plugins under the same module to be run. E.g. <--plugin_list=p1,p2,p3>

`--qemu_logs_timeout`

Timeout in secs for copying the qemu logs.

Default: 300

`--rerun_failing_plugins`

If set, run the plugins from the previous run which had status among ERROR, FAIL, WARN

Default: false

`--secure_file_transfer`

Boolean flag to specify if the file should be transfered using Secure File Transfer Protocol (SFTP); True/False.

Default: false

`--segregate_msgtype`

Flag to specify if the log files needs to be collected/segregated based on log message type.

Default: true

`--set_email_frequency`

Frequency in hours at which NCC should periodically run. E.g Run ncc every 24 hours: <--set_email_frequency=24>

Default: 0

`--show_auto_log_coll_config`

If True, display periodic debug config.

Default: false

`--show_email_config`

If set, display NCC email configuration

Default: false

`--start_time`

Start time from when logs should be collected. Time format is : %Y/%m/%d-%H:%M:%S E.g. <--start_time=2022/01/27-13:56:39>

`--upload_file`

Flag to specify if the collected logs should be uploaded on the remote server; True/False. E.g. enable upload: <--upload_file=True>

Default: false

`--use_hardlinks`

Flag to specify if the logs should be collected using hardlinks

Default: true

`--use_npyscreen`

If true, uses npyscreen for output.

Default: false

`--use_start_time_filter`

Flag to specify if the files will be filtered based on start time from file name regex and end time from last modified time.

Default: true

util.ncc.ncc_utils.globals

`--insights_rpc_server_ip`

The IP address of Insights RPC server.

Default: `127.0.0.1`

`--insights_rpc_server_port`

The port where the Insights RPC server listens.

Default: 2027

util.ncc.ncc_utils.heart_filter

`--ncc_query_unresolved_alerts`

Flag to identify if HeartFilter should query for unresolved alerts from IDF

Default: true

util.ncc.ncc_utils.progress_monitor

`--progress_sleep_duration_sec`

Time duration for which progress monitor should sleep between retries.

Default: 4

`--rsync_max_retry_count`

Number of retries when rsync fails to fetch remote progress file.

Default: 3

util.ncc.ncc_utils.rogue_plugin_utils

`--cpu_intensive_retries_threshold`

The number of retries after which plugin is marked as rogue if it was cpu intensive.

Default: 5

`--greenlet_switch_time`

If the greenlet executes continously for more then the given seconds then it is a CPU intensive greenlet.

Default: 18

`--rogue_cpu_time_refresh_interval`

Interval at which in memory value of CPU usage time is refreshed

Default: 2

`--rogue_plugin_file`

This file stores the rogue plugins before they are stored in insights.

Default: `/tmp/rogue_plugin.txt`

util.ncc.ncc_utils.upgrade_utils

`--cluster_upgrading_timeout`

Timeout for checking whether cluster is upgrading or not

Default: 600

util.ncc.plugins.base_plugin

`--accumulate_pagination_results`

If we want to collect results from the pagination calls.

Default: false

`--cluster_container_batch_entity_cnt`

Batch cnt of entities read and written to Entity DB

Default: 5

`--cluster_node_batch_entity_cnt`

Batch cnt of entities read and written to Entity DB

Default: 5

`--cluster_vm_batch_entity_cnt`

Batch cnt of entities read and written to Entity DB

Default: 1

`--default_cluster_uuid`

The flags specifies default_cluster_uuid

Default: `Unknown`

`--enable_cbl_for_neuron_plugins`

Switch to enable cluster based licensing for neuron plugins

Default: true

`--is_run_all`

Is this a run_all run ?

Default: false

`--multiple_entity_types_batch_entity_cnt`

Batch cnt of entities read and written to Entity DB

Default: 10

`--ncc_plugin_para_file_dir`

Directory path where ncc plugin parameter files are kept

Default: `/home/nutanix/ncc/plugin_config/plugin_param`

`--node_vm_batch_entity_cnt`

Batch cnt of entities read and written to Entity DB

Default: 2

`--scale_out_pc_write_entity_cnt`

Write batch size on scale out pc

Default: 40

`--waiting_plugin_sleep_time`

The flags specifies the amount of time the waiting plugins should sleep before checking for idle threads

Default: 0.05

util.ncc.plugins.consts

`--HDD_latency_threshold_ms`

HDD await threshold (ms/command) to determine Disk issues.

Default: 500

`--SSD_latency_threshold_ms`

SSD await threshold (ms/command) to determine Disk issues.

Default: 50

`--ahv_crash_file_age_policy_days`

The age of crash files that are to be alerted.

Default: 7

`--ahv_enable_audit_disk_usage_ncc_check`

Enable/Disable audit disk usage NCC check for AHV host

Default: false

`--ahv_enable_file_integrity_ncc_check`

Enable/Disable file integrity NCC check for AHV host

Default: false

`--ahv_enable_remote_log_forwarding_ncc_check`

Enable/Disable remote log forwarding NCC check for AHV host

Default: false

`--ahv_read_only_fs_check_exclude_list`

The list of filesystems that are expected to be read-only.

Default: /sys/fs/cgroup

`--audit_volume_usage_threshold_pct_critical`

Critical threshold for high disk usage on /home/log/audit.

Default: 99

`--audit_volume_usage_threshold_pct_warning`

Warn threshold for high disk usage on /home/log/audit.

Default: 75

`--autocomplete`

If true, recreate autocomplete bash script.

Default: false

`--autocomplete_dir`

Directory to store the auto-generated completion file.

Default: `/home/nutanix/ncc`

`--avg_io_latency_threshold_ms`

Average I/O Latency threshold (ms) for a VM.

Default: 5000

`--cas_failures_threshold`

Threshold count for cas failure log warnings.

Default: 5

`--cassandra_column_families`

Comma separated list of column families to check against size thresholds.

Default: `historical_stats`

`--cassandra_crashes_threshold`

Threshold for number of cassandra crashes.

Default: 5

`--cassandra_gc_activity_threshold`

Threshold for garbage collection related messages.

Default: 5

`--cassandra_high_heap_usage_threshold`

Threshold for number of heap usage log messages.

Default: 5

`--cassandra_init_json_file_path`

Path to find cassandra_init.json file.

Default: `/home/nutanix/data/cassandra/conf/cassandra_init.json`

`--cassandra_load_threshold`

Threshold for cassandra load-related messages.

Default: 5

`--cassandra_min_token_size`

Minimum size cassandra token

Default: 60

`--cassandra_progress_timeout_threshold`

Threshold for number of cassandra progress timeout log messages.

Default: 5

`--cassandra_restarts_threshold`

Threshold for cassandra restarts.

Default: 5

`--cassandra_retries_threshold`

Threshold count for cassandra retry log warnings.

Default: 5

`--cassandra_ring_cluster_version`

Cluster version of the nodetool_ring_file.

Default: `None`

`--cassandra_stage_counters_threshold`

Threshold for number of missing stage counter messages.

Default: 5

`--cassandra_token_precision`

Maximum chars in the cassandra token that are expected to be unique.

Default: 8

`--cassandra_unavailability_threshold`

Threshold count for cassandra unavailability log warnings.

Default: 5

`--check_cloud_cvm_services`

Comma-separated list of services which are checked on cloud CVM.

Default: cassandra,pithos,hera,stargate,insights_server,ergon,cerebro,chronos,curator,alert_manag

`--check_max_failure_score`

Health score to indicate check failure.

Default: 24

`--check_max_info_score`

Health score to indicate check info.

Default: 98

`--check_max_warning_score`

Health score to indicate check warning.

Default: 74

`--cluster_zeus_config_base_path`

Path where we keep zeus config for each cluster.

Default: `/appliance/physical/zeusconfig`

`--config_file_dir`

The base directory contains all plugin configs

Default: `/home/nutanix/ncc/plugin_config`

`--ctr_ec_ideal_delay_threshold`

Current ideal EC delay value setting for better Erasure coding benefits.

Default: 604800

`--cvm_avg_mem_util_threshold`

CVM memory average usage threshold (%) to determine memory contention.

Default: 90

`--cvm_commit_memory_threshold_pct`

CVM Memory Commit Threshold(%).

Default: 90

`--cvm_enable_file_integrity_ncc_check`

Whether to enable aide_check.

Default: false

`--cvm_load_average_threshold_critical`

CVM Load level.

Default: 100

`--cvm_load_factor`

CVM Load factor used with number of vCPUs to compute CVM Load threshold.

Default: 3.0

`--cvm_mem_min_threshold`

Minimum threshold for CVM memory for NOS > 2.6.4

Default: 15728640

`--cvm_peak_cpu_util_threshold_pct`

Peak utilization threshold (%) to determine CVM CPU contention.

Default: 90

`--default_hardware_type`

Default hardware block type for Svms without the rackable unit id field.

Default: `kNX2000`

`--detect_rogue_plugin`

If True, cpu intensive plugins will be detected and marked rogue. If False, cpu intensive plugins will not be detected but plugins may still be marked rogue for other offences.

Default: true

`--disk_abort_rate_threshold`

Disk read/write abort threshold. (aborts/sec)

Default: 0

`--disk_corruptions_threshold`

Threshold count for disk corruption log warnings.

Default: 5

`--disk_read_latency_threshold`

Latency threshold (ms/command) to determine storage issues.

Default: 10

`--disk_read_write_errors_threshold`

Threshold count for disk read/write log warnings.

Default: 5

`--disk_usage_threshold_pct_critical`

Threshold for disk usage (percentage) from being labeled as critical.

Default: 90

`--disk_usage_threshold_pct_warning`

Threshold for disk usage (percentage) from being labeled as a warning.

Default: 75

`--disk_write_latency_threshold`

Latency threshold (ms/command) to determine storage issues.

Default: 10

`--display_details_for_num_moves`

Displays the effect on the ring by applying the suggested node moves. The default is not to display any, the value of 0 is to display all until max_number_of_node_moves.

Default: -1

`--email_from_address`

Default senders email addresses.

Default: `cluster@nutanix.com`

`--email_human_readable_format`

Whether emails are to be sent in text or binary format.

Default: true

`--enable_kernel_mitigations_check`

Whether to enable kernel mitigations NCC check.

Default: false

`--enable_rsyslog_forwarding_check`

Whether to enable rsyslog forwarding NCC check.

Default: false

`--end_time_secs`

End point of time range.

Default: 1643291796

`--entity_id`

The id of the entity.

Default: `None`

`--entity_type`

The type of entity.

Default: `cluster`

`--error_id_json_path`

The file containing the error ids

Default: `/home/nutanix/ncc/plugin_config/error_ids.json`

`--ext4_journal_sequence_threshold`

Threshold for high ext4 journal sequence id as it approaches 2^31.

Default: 1879048192

`--field`

Field of data to be trained. E.g. CPU usage, disk usage.

Default: `hypervisor_cpu_usage_ppm`

`--file_path`

Full path of file to copy.

`--forced_reschedule_interval_sec`

Reschedule interval if the check execution falls in maintenance window.

Default: 300

`--future_end_time_secs`

End point of a future time range.

Default: 1643896596

`--future_start_time_secs`

Start Point of a future time range.

Default: 1643291796

`--garbage_egroups_size_critical_threshold_pct`

Default threshold for reporting size (percentage) occupied by garbage egroups as critical event.

Default: 25

`--garbage_egroups_size_warning_threshold_pct`

Default threshold for reporting size (percentage) occupied by garbage egroups as warning event.

Default: 15

`--generate_node_add_order`

Generate the order for node adds specified in the operations file, which results in the minimal token skew.

Default: false

`--gx_cpuprofile_frequency`

Number of interrupts per second the pprof cpu-profiler uses to sample.

Default: 100

`--gx_mysql_db`

MySQL database that is used to store the profiles.

Default: `profiles`

`--gx_mysql_host`

MySQL server's IP address.

Default: `10.1.56.13`

`--gx_mysql_hostname`

MySQL server's host name.

Default: `gxprof.eng.nutanix.com`

`--gx_mysql_passwd`

MySQL password for the given 'gx_mysql_user'.

Default: `password`

`--gx_mysql_user`

MySQL user name that was granted access to 'gx_mysql_db'.

Default: `nutanix`

`--gx_output_file`

Output filename without extension where gx write profiling data.

Default: `None`

`--gx_output_type`

Specify whether the output file is in `html` or `text` format.

Default: `html`

`--gx_pprof_file`

Specify the pprof file that is to be translated to gxprof format.

Default: `None`

`--gx_profile_description`

A text string of other important notes about the profile.

Default: `None`

`--gx_profile_detail_json`

A dictionary json of additional details about the profile.

Default: `None`

`--gx_profile_id`

Numeric identifier of the profile in the database.

Default: None

`--gx_profile_name`

Name of the profile in the database.

Default: `no name`

`--gx_profile_recorder`

Name of the person who run the profiler.

Default: `unknown`

`--gx_program`

Specify the program that is being run.

Default: `None`

`--gx_program_build_date`

Version timestamp of the program that was being profiled.

Default: `Build date is unspecified`

`--gx_program_build_id`

Version number of the program that was being profiled.

Default: `Build version is unspecified`

`--gx_regression_id`

Numeric identifier of a regression in the database.

Default: None

`--gx_run_duration`

The duration it takes to complete a run of the program or test. The unit of the duration is dependent on which program is being run. It could be second, mili-second, micro-second or others.

Default: 0

`--gx_run_status`

Current status of the execution being profiled.

Default: `Running pprof.`

`--gx_sample_excluded_symbols`

Comma-separated list of symbols to be excluded from profiling data.

Default: __restore_rt,std::_Function_handler::_M_invoke,std::_Bind::__call,std::function::operator

`--gx_total_seconds`

The pprof sampling duration in seconds.

Default: None

`--gx_verbose`

Log the progress for all gx commands to help trouble-shooting.

Default: false

`--health_disk_latency_threshold_ms`

Latency threshold (ms/command) to determine Disk issues.

Default: 50

`--helpshort`

show usage only for this module

Default: 0

`--high_disk_usage_threshold`

Threshold count for disk usage log warnings.

Default: 5

`--home_nutanix_usage_threshold_pct_critical`

Critical threshold for high disk usage on /home/nutanix as well as other disks that are not in Arithmos (percentage).

Default: 90

`--home_nutanix_usage_threshold_pct_warning`

Warn threshold for high disk usage on /home/nutanix as well as other disks that are not in Arithmos (percentage).

Default: 75

`--host_avg_cpu_util_threshold_pct`

Host utilization threshold (%) to determine host CPU contention.

Default: 75

`--host_balloon_threshold`

Balloon threshold to determine host memory bottlenecks.

Default: 0

`--host_disk_usage_threshold_pct_critical`

Threshold for host disk usage (percentage) being labeled as critical.

Default: 95

`--host_disk_usage_threshold_pct_warning`

Threshold for host disk usage (percentage) being labeled as warning.

Default: 90

`--host_peak_cpu_util_threshold_pct`

Host utilization threshold (%) to determine host CPU contention.

Default: 90

`--host_preupgrade_log_path`

The log path of hypervisor preupgrade checks.

Default: `/home/nutanix/data/logs/host_preupgrade.out`

`--host_rcv_drop_threshold`

Receive packets drop threshold at host to determine network issues.

Default: 0

`--host_swap_in_threshold`

Swap in threshold (mbytes/sec) to determine host memory bottlenecks.

Default: 0

`--host_swap_out_threshold`

Swap out threshold (mbytes/sec) to determine host memory bottlenecks.

Default: 0

`--host_swap_threshold_mbps`

Swap rate threshold (mbytes/sec) to determine host memory bottlenecks.

Default: 0

`--host_swap_used_threshold_bytes`

Swap used threshold (bytes) to determine host memory bottlenecks.

Default: 0

`--host_tx_drop_threshold`

Transmitted packets drop threshold at host to determine network issues.

Default: 0

`--hwclock_time_difference_days`

Number of days of time difference that is allowed between the SVM and the hypervisor

Default: 2

`--hyperv_external_switch_rename_enabled`

Feature enable / disable flag for External Switch rename support on HyperV.

Default: `False`

`--hyperv_host_external_vswitch_name`

Default name of the network switch for Hyper-V's external network.

Default: `ExternalSwitch`

`--hyperv_host_internal_ip_address`

IP address of internal interface on the Hyper-V host

Default: `192.168.5.1`

`--hyperv_user_name`

User name to use to connect to the Hyper-V host

Default: `administrator`

`--hyperv_user_password`

Password to use to connect to the Hyper-V host

Default: `nutanix/4u`

`--hypervisor_pre_upgrade_check_enable`

The flag to specify if the current ncc run is for hypervisor pre-upgrade

Default: false

`--hypervisor_pre_upgrade_check_list_path`

The flag to specify the path of all hypervisor pre_upgrade checks

Default: `/home/nutanix/ncc/config/hypervisor_pre_upgrade_check_list.json`

`--hypervisor_sample_period`

Hypervisor monitoring sampling interval.

Default: 300

`--ignore_negative_slope`

Whether ignore the negative slope or not.

Default: true

`--inode_usage_threshold_pct_critical`

Threshold for inode usage (percentage) being labeled as critical.

Default: 90

`--inode_usage_threshold_pct_warning`

Threshold for inode usage (percentage) being labeled as a warning.

Default: 75

`--insights_max_memory_usage_MB`

Maximum amount of memory in MB that the insights collector can use.

Default: 256

`--intel_ssd_num_bytes_written_critical_limit`

Critical markers for number of bytes written into the Intel SSDs.

Default: 7e+15

`--intel_ssd_num_bytes_written_warning_limit`

Warning markers for number of bytes written into the Intel SSDs.

Default: 6.5e+15

`--intel_ssd_warning_marker_position_below_max_celcius`

Temperature warning marker position for the Intel 910 SSDs. Since maximum temperature is retrieved from the device automatically, this value is actually the number of celcuis below the maximum temperature allowed storagepool

Default: 7

`--inter_cvm_bw_mb_sec_threshold`

Minimum bandwidth between CVMs (MB/s).

Default: 800

`--ip_list`

List of IPs to copy the file to.

`--iperf_server_timeout_secs`

Timeout (in seconds) for the iperf server per client.

Default: 60

`--is_periodic_run`

Flag to specify if the current ncc task is async orperiodic. True if the current task is periodic else False.

Default: false

`--list_modules`

List all available ncc modules.

Default: false

`--local_http_proxy_port`

Port of the local service proxy (ikat_proxy).

Default: 0

`--log_collection_duration`

Time duration for which logs are parsed.

Default: 300

`--log_plugin_output`

Logs the protobug generated by each plugin execution - used mainly for debugging.

Default: false

`--log_types`

Log types to include within the compression.

Default: `None`

`--log_util_ip_list`

List of IPs to gather logs from.

Default: `None`

`--max_cassandra_restart_time_secs`

Maximum time for cassandra daemon to be killed and restarted.

Default: 15

`--max_health_score`

Maximum health score for any plugin.

Default: 100

`--max_number_of_node_moves`

The number of iterations to find minimum token skew. Default is to use 3 node moves. The time taken is exponential in the number of node moves and the number of nodes in cluster.

Default: 3

`--max_retry_on_not_master`

Number of retries when a rpc to cerebro returns kNotMaster

Default: 3

`--max_ring_load_balance_ratio`

Maximum ratio between least and greatest loaded node in cassandra ring.

Default: 3

`--max_ring_token_balance_ratio`

Maximum token skew allowed in cassandra ring.

Default: 2

`--minimum_container_rf`

Minimum replication factor for containers.

Default: 2

`--monitor_log_file_name`

file name of the server monitor log.

Default: `cluster_health_monitor.log`

`--ncc_autocomplete`

If true, recreate autocomplete bash script.

Default: false

`--ncc_autocomplete_dir`

Directory to store the auto-generated completion file.

Default: `/home/nutanix/ncc`

`--ncc_canvas_json_dump_file`

File to which canvas json of latest NCC run will be serialized.

`--ncc_enable_intrusive_plugins`

If true, run plugins with intrusive impacts also.

Default: false

`--ncc_execution_token`

Execution token for the task.

Default: `None`

`--ncc_factory_config_path`

Path to factory config file.

Default: `/etc/nutanix/factory_config.json`

`--ncc_global`

If true, any local plugins invoked will be run across all nodes in the cluster.

Default: true

`--ncc_hardware_config_path`

Path to hardware config file.

Default: `/etc/nutanix/hardware_config.json`

`--ncc_init_time`

Approx. upper limit of time taken by ncc client binary to initialise.

Default: 60

`--ncc_interactive`

If true, the plugins will be run in an interactive mode, where the result will be available as soon as its ready.

Default: true

`--ncc_master_ip`

IP of the node from where the command is run.

Default: `None`

`--ncc_monitor_progress`

Monitor progress for long running tasks.

Default: true

`--ncc_plugin_dir`

Directory from where plugin should be loaded.

Default: `/home/nutanix/ncc/bin/plugins`

`--ncc_plugin_output_file`

File to which raw output of latest NCC run will be written (use '' to disable).

Default: `/home/nutanix/data/logs/ncc-output-latest.log`

`--ncc_plugin_output_history_file`

File to which raw output of all NCC runs will be written.

Default: `/home/nutanix/data/logs/ncc-output.log`

`--ncc_pre_upgrade_init_time`

The number of seconds for NCC initialization during pre-upgrade. This will be added to the total timeout of the plugins.

Default: 60

`--ncc_preupgrade_output_log_path`

The log path that contains the output of the latest ncc preupgrade run.

Default: `/home/nutanix/data/logs/ncc_preupgrade_output_latest.out`

`--ncc_print_config`

Print the plugin config and exit without running.

Default: false

`--ncc_show_hidden`

If true, display hidden plugins/modules.

Default: false

`--ncc_show_tree`

If true, display all the plugins in a tree format.

Default: false

`--ncc_telemetry_anonymizer_expiry_time`

Amount of time after which anonymizer needs to be refreshed to consider new entities.

Default: 3600

`--ncc_telemetry_check_scores_flush_frequency`

Flushing frequency for check score table.

Default: 900

`--ncc_telemetry_check_scores_queue_maxsize`

Maximum size of queue which processes and sends check scores to Pulse.

Default: 1000

`--ncc_telemetry_counts_flush_frequency`

Flushing frequency for stats table.

Default: 28800

`--ncc_telemetry_enabled`

Send ncc stats to Pulse. Heartbeat sent by health_rpc_service won't be affected by this flag.

Default: true

`--ncc_via_rpc`

Indicates if NCC is running via RPC.

Default: false

`--ncli_password`

Use a non-default ncli password (will be prompted for password input).

Default: true

`--neuron_check_licensing_plugins_list`

The flags specifies the path of the plugin list for cbl.

Default: `/home/nutanix/neuron/plugin_config/licensing_plugins.json`

`--neuron_execution_token`

Execution token for the task.

Default: `None`

`--neuron_plugin_dir`

Directory from where plugin should be loaded.

Default: `/home/nutanix/neuron/bin/plugins`

`--neuron_plugin_output_file`

File to which raw output of latest NCC run will be written (use '' to disable).

Default: `/home/nutanix/data/logs/neuron-output-latest.log`

`--neuron_port`

The base server directory

Default: 2800

`--nfs_abort_rate_threshold`

NFS read/write abort threshold (aborts/sec).

Default: 0

`--nfs_command_latency_threshold`

Latency threshold (ms/command) to determine storage issues.

Default: 10

`--nfs_file_count_threshold`

Upper threshold on the number of NFS files and directories per datastore.

Default: 20000

`--nic_error_check_period_secs`

Period over which to monitor NIC errors (seconds).

Default: 3600

`--nic_error_threshold_info`

Maximum errors permissible for any NIC.

Default: 5

`--nic_error_threshold_warning`

Maximum errors permissible for any NIC before warning is issued.

Default: 100

`--nodetool_ring_file`

Path to the 'nodetool' ring output file. Default is to use the nodetool command on the cluster.

Default: `None`

`--non_shell_vdisk_count_safe_threshold`

Safe threshold for max number of regular vdisks in versions > 5.5

Default: 600000

`--non_shell_vdisk_count_threshold`

Maximum number of non-shell vdisks in cluster.

Default: 200000

`--nutanix_model_config_file_dir`

The base directory contains all nutanix models

Default: `/home/nutanix/ncc/config`

`--nvm_peak_cpu_util_threshold_pct`

Peak utilization threshold (%) to determine FSVM CPU contention.

Default: 90

`--nwviz_disable_switch_collector`

Disables switch_collector

Default: false

`--nwviz_disable_switch_interface_collector`

Disables switch_interface_collector

Default: false

`--nwviz_ignore_auto_discovery`

Disable auto discovery from host NICs using LLDP

Default: true

`--operations_file`

Path to node operations add/remove/no-op file.

Default: `None`

`--oplog_episode_count_threshold`

Threshold for the number of episodes per vdisk.

Default: 12000

`--oplog_episode_count_threshold_pre_4_6`

Threshold for the number of episodes for a vdisk in a cluster which has a version lower than 4.6.

Default: 100

`--orphan_snapshot_prefixes`

A commas separated list of prefixes of snapshot names for skip reporting orphan snapshots.

Default: `XDSNAP_`

`--override_existing_plugin_config`

Over-ride the plugin configuration stored in Zookeeper with the default NCC plugin configuration.

Default: false

`--plugin_output_history_file`

File to which raw output of all NCC runs will be written.

Default: `/home/nutanix/data/logs/ncc-output.log`

`--plugins_to_run`

The flag to specify the plugins that needs to be part of a single run e.g. --plugins_to_run=host_disk_usage_check,metadata_mounted_check

`--pre_upgrade_check_enable`

The flag to specify if the current ncc run is for pre-upgrade

Default: false

`--pre_upgrade_check_list_path`

The flag to specify the path of all pre_upgrade checks

Default: `/home/nutanix/ncc/config/pre_upgrade_check_list.json`

`--pre_upgrade_check_slave_run`

The flag to specify if the current ncc run is that of a pre-upgrade slave run

Default: false

`--preupgrade_log_path`

The log path of all preupgrade checks.

Default: `/home/nutanix/data/logs/preupgrade.out`

`--progress_timeout`

Timeout for getting progress updates from plugins by progress monitor.

Default: 2500

`--remote_site_ping_warn_threshold_ms`

Threshold for ping latency to remote beyond which the remote will be labeled as warning.

Default: 300

`--remote_site_socket_timeout_secs`

Timeout for creating a socket to Stargate running on the remote site.

Default: 5

`--remote_site_sync_critical_threshold_sec`

Threshold for remote cluster time (in seconds) being out of sync with source cluster to be labeled as critical.

Default: 300

`--remote_site_sync_warn_threshold_sec`

Threshold for remote cluster time (in seconds) being out of sync with source cluster to be labeled as warning.

Default: 180

`--replication_lag_critical_threshold`

Maximum number of pending replications beyond which a protection domain will be labeled as critical.

Default: 5

`--replication_lag_warn_threshold`

Maximum number of pending replications beyond which a protection domain will be labeled as warning.

Default: 2

`--root_nutanix_usage_threshold_pct_critical`

Critical threshold for high disk usage on /(percentage).

Default: 90

`--root_nutanix_usage_threshold_pct_warning`

Warn threshold for high disk usage on /(percentage).

Default: 80

`--rpc_timeouts_threshold`

Threshold count for rpc timeout log warnings.

Default: 5

`--run_using_gevent`

Run scheduler using gevent

Default: true

`--sampling_interval_secs`

The length of sampling interval.

Default: 600

`--scheduler_type_list`

List of scheduler type.

`--server_execution_token`

Execution token for the task.

Default: `None`

`--server_global`

If true, any local plugins invoked will be run across all nodes in the cluster.

Default: true

`--server_global_install`

Install server on all nodes.

Default: false

`--server_log_file_name`

file name of the server log.

Default: `health_server.log`

`--server_logging_dir`

Directory where script log files are stored.

Default: `/home/nutanix/data/logs`

`--server_name`

name of the server.

Default: `ncc`

`--server_plugin_dir`

Directory from where plugin should be loaded.

Default: `/home/nutanix/ncc/bin/plugins`

`--server_plugin_output_file`

File to which raw output of latest NCC run will be written (use '' to disable).

Default: `/home/nutanix/data/logs/ncc-output-latest.log`

`--server_port`

The base server directory

Default: 2700

`--server_watch_client_name`

name of the server.

Default: `ClusterHealth`

`--shell_vdisk_count_threshold`

Maximum number of shell vdisks in cluster.

Default: 4000000

`--smart_alert_zk_node_path`

flag for smart alert zk node to check if enabled

Default: `/appliance/logical/analytics/smart_alert_enabled`

`--smtp_tunnel_port`

Local port number for the SMTP tunnel on the zookeeper leader.

Default: 2525

`--snapshot_chain_height_threshold`

Default snapshot chain height threshold to determine if snapshot chains are not getting severed.

Default: 25

`--snapshot_usage_bytes_stat`

Generic stat name to get the space used by snapshots of a protection domain

Default: `dr.exclusive_snapshot_usage_bytes`

`--snapshot_usage_crit_threshold_pct`

Default threshold for reporting size (percentage) occupied by snapshots as critical event

Default: 40

`--snapshot_usage_warn_threshold_pct`

Default threshold for reporting size (percentage) occupied by snapshots as warning event

Default: 30

`--sstable_size_thresholds_bytes`

Comma separated list of SSTable file size threshold.

Default: `2147483648`

`--stargate_retries_threshold`

Threshold count for stargate retry log warnings.

Default: 5

`--stargate_vars`

List of stargate variables to store in arithmos (empty list returns all). Stats will only be stored if they are supported by the relevant plugin.

`--stargate_vars_url`

Url where stargate stats can be queried.

Default: `http://127.0.0.1:2009/h/vars`

`--start_time_secs`

Start Point of time range.

Default: 0

`--stats_collection_interval`

Default stats collection gathering interval.

Default: 300

`--stats_default_email_addresses`

Default target email addresses for sending stats.

Default: `stats@nutanix.com`

`--switch_interface_name_map`

Switch interface name map

`--test_data_config_file`

The path to the file containing the test data config

Default: `None`

`--total_usage_discrepancy_critical_threshold_pct`

Default threshold for reporting discrepancy of usage (percentage) between total container utilization and total transformed usage as critical event. The discrepancy should be at-least higher than --total_usage_discrepancy_min_size_gib for the event to be triggered.

Default: 25

`--total_usage_discrepancy_min_size_gib`

Minimum size required for reporting a discrepancy of usage (gib) between total container utilization and total transformed usage as any event.

Default: 10

`--total_usage_discrepancy_warning_threshold_pct`

Default threshold for reporting discrepancy of usage (percentage) between total container utilization and total transformed usage as warning event. The discrepancy should be at-least higher than --total_usage_discrepancy_min_size_gib for the event to be triggered.

Default: 15

`--transport_errors_threshold`

Threshold count for transport error log warnings.

Default: 5

`--update_plugin_config`

Update the plugin configuration stored in Zookeeper with the default NCC plugin configuration.

Default: true

`--use_esxtop`

Use esxtop to get stats. Otherwise use Arithmos

Default: false

`--vdisk_count_threshold`

Maximum number of vdisks in cluster.

Default: 80000

`--vm_avg_cpu_util_threshold`

Average utilization threshold (%) to determine VM CPU contention.

Default: 75

`--vm_avg_mem_util_threshold_pct`

VM memory average usage threshold (%) to determine memory contention.

Default: 80

`--vm_balloon_reclaim_threshold_mb`

Threshold to determine memory pressurefor a VM.

Default: 0

`--vm_balloon_threshold`

Balloon threshold to determine memory bottlenecks.

Default: 0

`--vm_cpu_util_threshold`

Utilization threshold (%) to determine VM CPU contention.

Default: 75

`--vm_peak_cpu_util_threshold_pct`

Peak utilization threshold (%) to determine VM CPU contention.

Default: 90

`--vm_peak_mem_util_threshold`

Peak utilization threshold (%) to determine VM memory bottlenecks.

Default: 90

`--vm_rcv_drop_threshold`

Receive packets drop threshold at a VM to determine network issues.

Default: 4096

`--vm_swap_in_threshold`

Swap in threshold (mbytes/sec) to determine memory bottlenecks.

Default: 0

`--vm_swap_out_threshold`

Swap out threshold (mbytes/sec) to determine memory bottlenecks.

Default: 0

`--vm_swap_threshold_mbps`

Swap out threshold (mbytes/sec) to determine memory bottlenecks.

Default: 0

`--vm_tx_drop_threshold`

Transmitted packets drop threshold at a VM to determine network issues.

Default: 0

`--vm_vcpu_ready_time_threshold`

Vcpu ready time threshold (%) to determine CPU contention.

Default: 10

`--zeus_config_file`

Path to the zeus config file. Default is to use the zeus config printer on the cluster.

Default: `None`

`--zeus_configuration_path`

Zookeeper path where zeus configuration is stored.

Default: `/appliance/physical/configuration`

`--zk_leader_path`

The path to leader of the zk node.

Default: `health_scheduler_master`

util.ncc.plugins.firstimport

`--ncc_base_dir`

The base NCC directory

Default: `/home/nutanix/ncc`

`--neuron_base_dir`

The base NEURON directory

Default: `/home/nutanix/neuron`

`--server_base_dir`

The base server directory

Default: `/home/nutanix/ncc`

`--use_infra_service_monitor`

Use infra's service monitor.

Default: true

util.ncc.plugins.log_collector.flags

`--append_ftp_cluster_metadata`

If true, cluster metadata is appended to the ftp_target_location. The dir structure in this case is : FLAGS.ftp_target_location/<cluster-uuid>/timestamp/<node-id>

Default: false

`--append_ftp_location`

If true, 'LogCollector' is appended to the target ftp location before uploadingthe log bundle. In case it is False, the log bundle is directly uploaded to FLAGS.ftp_target_location.

Default: true

`--authenticate_using_keys`

Boolean flag to specify if the authentication should happen using keys. Keys authentication is only supported via SFTP.

Default: false

`--auto_log_coll_slave`

If true, ncc is running the periodic log collector on slave

Default: false

`--cache_reset_limit_mb`

The disk space cache reset limit.

Default: 512

`--curr_logfile_timestamp`

The timestamp for the current tarball generated

Default: 0

`--gz_extension`

The extension of compressed tarfile.

Default: `.tar.gz`

`--http_timeout`

Timeout for http request.

Default: 60

`--hypervisor_log_input_dir`

If this flag is mentioned the logs on the hypervisor will be collected from the given directory, otherwise they will be collected from the default configured directory.

`--iam_svc_acc_zkpath`

IAM service account Zk Path.

Default: `/appliance/logical/iam/mgmt_plane/service_account_config`

`--log_input_dir`

If this flag is mentioned, logs will be collected from the given directory, otherwise they will be collected from the default configured directory.

`--max_log_file_mb`

The maximum log file size to be copied.

Default: 500

`--max_sftp_upload_retry_count`

Maximum number of sftp upload retry count for a file

Default: 3

`--max_used_percent_hard_limit`

The maximum used percentage after which we do not write any logs

Default: 95

`--max_used_percent_soft_limit`

The maximum used percentage after which we write the logs only if --force option is set byt the user.

Default: 90

`--output_log_location`

The location where log collector stores the log files.

Default: `/home/nutanix/data/log_collector`

`--sftp_authentication_http_server`

Server to which http request is sent for authentication.

Default: `https://10.4.244.65/nusights/debug_server/v1/putSftpKeys`

`--sftp_key_file`

File where generated private sftp key is stored.

Default: `/home/nutanix/tmp/log_collector_key`

`--sftp_upload_retry_interval_secs`

The gap between consecutive sftp file upload retry in seconds

Default: 5

`--skip_timezone`

The flag specifies whether the timezone info should not be written to the log file

Default: false

`--tar_extension`

The extension of the output tarfile.

Default: `.tar`

`--use_iam`

Boolean to specify if IAM token based authentication has to be enabled.

Default: false

setup_hyperv.py

Usage

```
Usage: setup_hyperv.py [FLAGS] [command]

commands:
 register_shares
 setup_metro
 setup_scvmm
 teardown_metro
```

/usr/local/nutanix/bin/setup_hyperv.py

`--configure_library_share`

Whether a library share should be configured

Default: None

`--default_host_group_path`

The default SCVMM host group

Default: `All Hosts`

`--help`

Print detailed help

Default: false

`--library_share_name`

The name of the container that will be registered as a library share in SCVMM

`--metro_smb_account_password`

Password for the new metro cluster pair fqdn.

`--metro_smb_name`

This is the name that identifies a unique pair of AOS clusters. This name must be used for the name of the SMB server when provisioning virtual machines and virtual hard disks on any metro container stretched between the pair of AOS clusters.

`--ncli_password`

Password to be used when running ncli

`--nutanix_management_share`

The storage container for nutanix cluster management

Default: `NutanixManagementShare`

`--password`

Domain account password for the host

`--scvmm_host_group_path`

Host Group to which this cluster should be added

`--scvmm_password`

SCVMM account password - defaults to <password>

`--scvmm_server_name`

Name of the server running SCVMM

`--scvmm_username`

SCVMM account username (with the FQDN) - defaults to <host_fqdn> \<username>

`--setup_metro_no_ad`

Skips adding AD and DNS, but verifies it is added before proceeding further. This does not require the domain username and password and requires manual AD and DNS setup prior to running the script.

Default: false

`--storage_pool_name`

The name of the storage pool created if one doesn't already exist

Default: `NTNX-SP-DEFAULT`

`--username`

Domain account username (without domain name) which has/will have administrative rights on the host

`--verbose`

Print verbose output

Default: false

`--zknode`

Zookeeper node for Secure Keys

Default: `/appliance/logical/prismsecurekey`

COPYRIGHT

Copyright 2022 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

License

The provision of this software to you does not grant any licenses or other rights under any Microsoft patents with respect to anything other than the file server implementation portion of the binaries for this software, including no licenses or any other rights in any hardware or any devices or software that are used to communicate with or in connection with this software.

Conventions

Convention	Description
`variable_value`	The action depends on a value that is unique to your environment.
`ncli> command`	The commands are executed in the Nutanix nCLI.
`user@host$ command`	The commands are executed as a non-privileged user (such as nutanix) in the system shell.
`root@host# command`	The commands are executed as the root user in the vSphere or Acropolis host shell.
`> command`	The commands are executed in the Hyper-V host shell.
`output`	The information is displayed as output from a command or in a log file.

Version

Last modified: February 24, 2022 (2022-02-24T21:07:35+05:30)

NX-1065-G8 SPECIFICATION

MODEL	Nutanix : Per Node (per Block) NX-1065-G8 (Configure to Order)
DEPLOYMENT MODEL	Factory Installed Software
USE CASE(S)	Remote Office/Branch Office, Test and Development, Analytics and Big Data
SERVER COMPUTE **	Dual Intel Ice Lake: Intel Xeon® Silver 4309Y [8 cores / 2.80 GHz] Intel Xeon® Silver 4310 [12 cores / 2.10 GHz] Intel Xeon® Silver 4310T [10 cores / 2.30 GHz] Intel Xeon® Silver 4314 [16 cores / 2.40 GHz] Intel Xeon® Silver 4316 [20 cores / 2.30 GHz] Intel Xeon® Gold 5315Y [8 cores / 3.20 GHz] Intel Xeon® Gold 5317 [12 cores / 3.00 GHz] Intel Xeon® Gold 5320T [20 cores / 2.30 GHz]
Boot	N/A
Dual M.2 RAID1	[2] x 512GB M.2 Boot Device
STORAGE CAPACITY	
SSD+HDD	[1] x SSD: [1.92 TB, 3.84 TB, 7.68 TB], [2] x HDD: [6.0 TB, 8.0 TB, 12.0 TB, 18.0 TB]
SSD+HDD SED	[1] x SSD: [1.92 TB, 3.84 TB], [2] x HDD: [6.0 TB, 8.0 TB, 12.0 TB]
MEMORY	128 GB, 256 GB, 384 GB, 512 GB, 768 GB, 1024 GB
NETWORK CONNECTIONS	
Server-board	[1] 1GbE Dedicated IPMI[2] 10GbE SFP+[2] 10GBaseT (IPMI Failover)
Network Card	[0, 1, 2] x 10GBaseT 2P NIC [0, 1, 2] x 10GBaseT 4P NIC [0, 1, 2] x 10GbE 4P NIC [0, 1, 2] x 25GbE 2P NIC
GPU	N/A
CERTIFICATIONS	Energy Star, CSAus, FCC, CSA, UL, cUL, ICES, CE, UKCA, KCC, RCM, VCCI-A, BSMI, EAC, SABS, S-MARK, UKRSEPRO, BIS, SII, RoHS, Reach, WEEE
DIMENSIONS (PER BLOCK)	Height: 3.47" (88.10 mm) Width: 17.68" (449 mm) Depth: 30.11" (764.75 mm) Rack Units: 2U4N
WEIGHT (PER BLOCK)	Standalone: 97.00 lbs.(44.0 kg) Package: 123.46 lbs.(56.0 kg)

MODEL	Nutanix : Per Node (per Block) NX-1065-G8 (Configure to Order)
SYSTEM COOLING	4x 80 mm heavy duty fans with PWM fan speed controls
OPERATING ENVIRONMENT (PER BLOCK)	Op Temp Range: 50° to 95°F(10° to 35°C) Non-Op Temp Range: -49° to 158°F(-45° to 70°C) Op Humidity Range (non-condensing): 68° to 194°F(20° to 90°C) Non-Op Humidity Range: 41° to 203°F(5° to 95°C)
POWER CONSUMPTION	Typical: 1680W Maximum: 2250W
POWER SUPPLY (DUAL SUPPLY / BLOCK)	1.2kW Output @100-127V, 11.0-14.0A, 50-60Hz 2.1kW Output @180-220V, 11.0-14.0A, 50-60Hz 2.2kW Output @220-240V, 11.0-12.0A, 50-60Hz Energy Star 80+ Titanium
THERMAL DISSIPATION	Typical: 5733.0 BTU/hr Maximum: 7644.0 BTU/hr
OPERATING REQUIREMENTS (PER BLOCK)	Input Voltage: 100-240V AC auto-range, Input Frequency: 50-60Hz

** Number of cores per CPU socket.

4316, 5315Y, 5317, 5320T have a maximum operating temperature of 30C.

Maximum measurements are shown. Width is from ear to ear; depth is from ears to pull rings.

4316, 5315Y, 5317, 5320T have a maximum operating temperature of 30C.

NUTANIX
YOUR ENTERPRISE CLOUD

T. 855.NUTANIX (855.688.2649) | F. 408.916.4039
info@nutanix.com | www.nutanix.com | @nutanix

© 2019 Nutanix, Inc. All rights reserved. Nutanix, the Nutanix logo and all product and service names mentioned herein are registered trademarks or trademarks of Nutanix, Inc. in the United States and other countries. All other brand names mentioned herein are for identi cation purposes only and my be the trademarks of their respective holder(s).

NX-8035-G8 SPECIFICATION

MODEL	Nutanix : Per Node (per Block) NX-8035-G8 (Configure to Order)
DEPLOYMENT MODEL	Factory Installed Software
USE CASE(S)	Database and Business Critical Apps, Analytics and Big Data
SERVER COMPUTE **	Dual Intel Ice Lake: Intel Xeon® Silver 4309Y [8 cores / 2.80 GHz] Intel Xeon® Silver 4310 [12 cores / 2.10 GHz] Intel Xeon® Silver 4310T [10 cores / 2.30 GHz] Intel Xeon® Silver 4314 [16 cores / 2.40 GHz] Intel Xeon® Silver 4316 [20 cores / 2.30 GHz] Intel Xeon® Gold 5315Y [8 cores / 3.20 GHz] Intel Xeon® Gold 5317 [12 cores / 3.00 GHz] Intel Xeon® Gold 5318Y [24 cores / 2.10 GHz] Intel Xeon® Gold 5320 [26 cores / 2.20 GHz] Intel Xeon® Gold 5320T [20 cores / 2.30 GHz] Intel Xeon® Gold 6326 [16 cores / 2.90 GHz] Intel Xeon® Gold 6330 [28 cores / 2.00 GHz] Intel Xeon® Gold 6334 [8 cores / 3.60 GHz] Intel Xeon® Gold 6336Y [24 cores / 2.40 GHz] Intel Xeon® Gold 6338 [32 cores / 2.00 GHz] Intel Xeon® Gold 6342 [24 cores / 2.80 GHz] Intel Xeon® Gold 6346 [16 cores / 3.10 GHz] Intel Xeon® Gold 6348 [28 cores / 2.60 GHz] Intel Xeon® Gold 6354 [18 cores / 3.00 GHz]
Boot	N/A
Dual M.2 RAID1	[2] x 512GB M.2 Boot Device
STORAGE CAPACITY	
All SSD SED	[2, 4, 6] x SSD: [1.92 TB, 3.84 TB]
NVMe+SSD	[2] x NVMe: [1.92 TB, 3.84 TB, 7.68 TB], [4] x SSD: [1.92 TB, 3.84 TB, 7.68 TB]
All SSD	[2, 4, 6] x SSD: [1.92 TB, 3.84 TB, 7.68 TB]
SSD+HDD	[2] x SSD: [1.92 TB, 3.84 TB, 7.68 TB], [4] x HDD: [6.0 TB, 8.0 TB, 12.0 TB, 18.0 TB]
SSD+HDD SED	[2] x SSD: [1.92 TB, 3.84 TB], [4] x HDD: [6.0 TB, 8.0 TB, 12.0 TB]
MEMORY	128 GB, 256 GB, 384 GB, 512 GB, 768 GB, 1024 GB, 1536 GB, 2048 GB
NETWORK CONNECTIONS	
Server-board	[1] 1GbE Dedicated IPMI[2] 10GbE SFP+[2] 10GBaseT (IPMI Failover)
Network Card	[0, 1, 2, 3] x 10GBaseT 4P NIC [0, 1, 2, 3] x 10GbE 4P NIC [0, 1, 2, 3] x 25GbE 2P NIC

MODEL	Nutanix : Per Node (per Block) NX-8035-G8 (Configure to Order)
GPU	N/A
CERTIFICATIONS	Energy Star, CSAus, FCC, CSA, UL, cUL, ICES, CE, UKCA, KCC, RCM, VCCI-A, BSMI, EAC, SABS, S-MARK, UKRSEPRO, BIS, SII, RoHS, Reach, WEEE
DIMENSIONS (PER BLOCK)	Height: 3.47" (88.10 mm) Width: 17.68" (449 mm) Depth: 30.11" (764.75 mm) Rack Units: 2U2N
WEIGHT (PER BLOCK)	Standalone: 80.69 lbs.(36.6 kg) Package: 108.03 lbs.(49.0 kg)
SYSTEM COOLING	4x 80 mm heavy duty fans with PWM fan speed controls
OPERATING ENVIRONMENT (PER BLOCK)	Op Temp Range: 50° to 95°F(10° to 35°C) Non-Op Temp Range: -49° to 158°F(-45° to 70°C) Op Humidity Range (non-condensing): 68° to 194°F(20° to 90°C) Non-Op Humidity Range: 41° to 203°F(5° to 95°C)
POWER CONSUMPTION	Typical: 1500W Maximum: 2000W
POWER SUPPLY (DUAL SUPPLY / BLOCK)	1.2kW Output @100-127V, 11.0-14.0A, 50-60Hz 2.1kW Output @180-220V, 11.0-14.0A, 50-60Hz 2.2kW Output @220-240V, 11.0-12.0A, 50-60Hz Energy Star 80+ Titanium
THERMAL DISSIPATION	Typical: 5118.0 BTU/hr Maximum: 6824.0 BTU/hr
OPERATING REQUIREMENTS (PER BLOCK)	Input Voltage: 100-240V AC auto-range, Input Frequency: 50-60Hz

** Number of cores per CPU socket.

When using all-SSD storage you must use only two, four, or six drives. Configurations of three or five drives are not supported.

NUTANIX
YOUR ENTERPRISE CLOUD

T. 855.NUTANIX (855.688.2649) | F. 408.916.4039
info@nutanix.com | www.nutanix.com | @nutanix

© 2019 Nutanix, Inc. All rights reserved. Nutanix, the Nutanix logo and all product and service names mentioned herein are registered trademarks or trademarks of Nutanix, Inc. in the United States and other countries. All other brand names mentioned herein are for identi cation purposes only and my be the trademarks of their respective holder(s).

SOLUTION BRIEF

Nutanix Flow

VISUALIZE | SECURE | MONITOR

FUNCTIONS

VM Microsegmentation

Application Visualization

Service Chaining

Policy-based Control

Identity Based Policy

Traffic Monitoring and Threat Detection

Compliance Audit and Reporting

The network is the foundation for the connected Enterprise. With increasing application complexity, driven by distributed services and rapid growth in SaaS and cloud-based offerings, the need to easily visualize, analyze and govern network communications becomes critical.

Nutanix Flow delivers advanced network security, providing visibility into the virtual network, application-centric protection from network threats, malware, and ransomware and security and compliance monitoring.

Flow allows organizations to deploy software-defined virtual network security without the complexity of installing and managing additional products that have separate management and independent software maintenance requirements.

Fully integrated into the Nutanix private cloud solution and Nutanix AHV virtualization, Nutanix Flow differs from traditional perimeter firewalls by allowing network policy to be attached to VMs and applications, rather than specific network segments (e.g., VLANS) or identifiers (IP addresses). Through centralized management from Prism, policies are auto-updated throughout the VM lifecycle, removing the burdens of change management.

Application and Network Visualization

Creating the right network policies requires a complete understanding of workload behavior. Nutanix Flow provides a detailed visualization of communications between VMs along with assistance in categorizing and grouping workloads, making it simple and straight-forward to set the right policies for the environment.

©2021 Nutanix, Inc. All Rights Reserved

KEY BENEFITS

- Application centric network policies for VMs
- Works with any network topology or architecture
- Application and network visibility
- Policy change management tied to VM lifecycle
- Prevent the propagation of network security threats
- Adds Identity based security to EUC/VDI deployments
- Security posture visibility and tracking dashboard
- Compliance audit and reporting regulations like HIPAA, PCI, NIST, and more
- Expand functionality via third-party network inspection and policy tools

Network Isolation and Microsegmentation

Microsegmentation provides granular control and governance of all traffic into and out of a virtual machine (VM), or groups of VMs. It ensures that only permitted traffic between application tiers or other logical boundaries is allowed and protects against advanced threats propagating within the virtual environment.

Identity-Based Network Policy

Identity Firewall in Flow allows EUC policy to incorporate the additional context of a user's identity. Group and role information pulled from directory services combine to create a more granular yet dynamic policy model.

Service Insertion and Chaining

Nutanix Flow functionality can be extended to leverage virtualized network functions from third-party software. These services are inserted in-line or in tap-mode with VM traffic, and can be easily enabled for all traffic, or deployed only for specific network traffic. Common network functions include virtual firewalls, load balancers, threat detection, and application performance monitoring.

Monitoring, Compliance Audit, and Reporting

Flow's Security Central provides a security operations portal for your Nutanix infrastructure adding visibility and control of network traffic and Nutanix security configurations along with out of the box audit and reporting for common security and compliance regulations such as HIPAA, PCI, and NIST. You can also create custom audit checks to align with your specific security compliance needs.

RESOURCES

- nutanix.com/products/flow
- nutanix.com/products/ahv
- nutanix.com/security

PRICING AND AVAILABILITY

Nutanix Flow is built into the Nutanix software and works seamlessly with AHV virtualization. Capabilities are delivered with AHV and are licensed separately from AOS Editions.

	AHV	Flow License
Simple Service Insertion and Chaining	✓	
Network Automation	✓	
Application Flow Visualization		✓
Policy-Based Service Insertion and Chaining		✓
Application Security with Microsegmentation		✓
Flow Security Central: Compliance Audit and Reporting, Security Posture and Status Dashboards		✓

NUTANIX™
YOUR ENTERPRISE CLOUD

T. 855.NUTANIX (855.688.2649) | F. 408.916.4039
info@nutanix.com | www.nutanix.com | @nutanix

©2021 Nutanix, Inc. All rights reserved. Nutanix, the Nutanix logo and all product and service names mentioned herein are registered trademarks or trademarks of Nutanix, Inc. in the United States and other countries. All other brand names mentioned herein are for identification purposes only and may be the trademarks of their respective holder(s).

Files 3.0

File Analytics Guide

March 21, 2022

Contents

File Analytics **4**
- Deployment Requirements 5
  - Network Requirements 5
- Limitations 5

Administration **7**
- Deploying File Analytics 7
  - Enabling File Analytics 9
  - Disabling File Analytics 11
- Launching File Analytics 11
- File Analytics VM Management 12
  - Removing File Analytics VMs 12
- Updating Credentials 12
- Managing Deleted Share/Export Audits 13
- Changing an FAVM Password 13
- Upgrades 14
  - Upgrade File Analytics at a Dark Site 14

Dashboard **16**
- Health 21
- Data Age 21
  - Configuring Data Heat Levels 22

Anomalies **23**
- Configuring Anomaly Detection 26
- Configuring an SMTP Server 27

Audit Trails **30**
- View Audit Trails 30
- Audit Trails - Users 30
- Audit Trails - Folders 32
- Audit Trails - Files 33
- Audit Trails - Client IP 35

Ransomware Protection **38**
- Ransomware Protection Features 38
- Enabling Ransomware Protection 46
- Configuring Ransomware Protection 47
- Enabling SSR 48

Reports **50**
- Reports Dashboard 50
- Creating a Custom Report 55

Create a Pre-Canned Report....55

File Analytics Options....57

Updating Data Retention....57
Scanning the File System....57
Blacklisting....58
Managing File Categories....58
Data Protection....59
Configuring Disaster Recovery....59
Activating Disaster Recovery....59
Deploying File Analytics on a Remote Site (AHV)....59
Deploying File Analytics on a Remote Site (ESXi)....62

Copyright....66

FILE ANALYTICS

File Analytics provides data and statistics on the operations and contents of a file server.

Once deployed, Nutanix Files adds a File Analytics VM (FAVM) to the Files cluster. A single File Analytics VM supports all file servers in the cluster; however, you must enable File Analytics separately for each file server. File Analytics protects data on the FAVM, which is kept in a separate volume group.

Once you deploy File Analytics, a new File Analytics link appears on the file server actions bar. Use the link to access File Analytics on any file server that has File Analytics enabled.

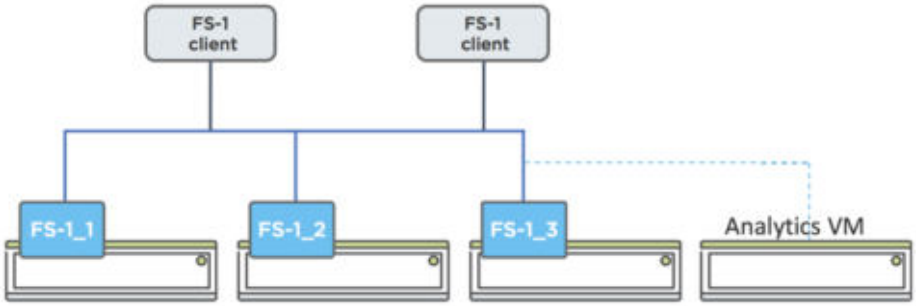


Figure 1: File Analytics VM

Display Features

The File Analytics web console consists of display features:

Main menu bar: The main menu bar appears at the top of every page of the File Analytics web console. The main menu bar includes the following display features:

- **Dashboard tab**: View widgets that present data on file trends, distribution, and operations, see Dashboard on page 16.
- **Audit Trails tab**: Search for a specific user or file and view various widgets to audit activity, see Audit Trails on page 30.
- **Anomalies tab**: Create anomaly policies and view anomaly trends, see Anomalies on page 23.
- **Ransomware tab**: Configure ransomware protection and self-service restore (SSR) snapshots, see Ransomware Protection on page 38.

  > **Warning:** Ransomware protection helps detect potential ransomware. Nutanix does not recommend using the File Analytics ransomware feature as an all-encompassing ransomware solution.

- **Reports tab**: Create custom reports or use pre-canned report templates, see Reports on page 50.
- **Status icon**: Check the file system scan status.
- **File server drop-down**: View the name of the file server for which data is displayed.
- **Settings drop-down**: Manage File Analytics and configure settings, see Administration on page 7 and File Analytics Options on page 57.
- **Health icon**: Check the health of File Analytics, see Health on page 21.
- **Admin dropdown**: Collect logs and view the current File Analytics version.

Deployment Requirements

Meet the following requirements prior to deploying File Analytics.

Ensure that you have performed the following tasks and your Files deployment meets the following specifications.

- Assign the file server administrator role to an Active Directory (AD) user, see *Managing Roles* in the *Nutanix Files Guide*.
- Log on as the Prism `admin` user to deploy the File Analytics server.
- Configure a VLAN with one dedicated IP address for File Analytics, or you can use an IP address from an existing Files external network. This IP address must have connectivity to AD, the control VM (CVM), and Files. See "Configuring a Virtual Network For guest VM Interfaces" in the Prism Web Console Guide.

  > **Note:** Do not install File Analytics on the Files internal network.

- (optional) Assign the file server administrator role to an LDAP user, see *Managing Roles* in the *Nutanix Files Guide*.
- Ensure that all software components meet the supported configurations and system limits, see the *File Analytics Release Notes*.

Network Requirements

Open the required ports, and ensure that your firewall allows bi-directional Internet Control Message Protocol (ICMP) traffic between the FAVM and CVMs.

The Port Reference provides detailed port information for Nutanix products and services, including port sources and destinations, service descriptions, directionality, and protocol requirements.

In addition to meeting the File Analytics network requirements, ensure to meet Nutanix Files port requirements as described in the *Port Reference*.

Limitations

File Analytics has the following limitations.

> **Note:** Depending on data set size, file count, and workload type, enabling File Analytics can affect the performance of Nutanix Files. High latency is more common with heavy file-metadata operations (directory and file creation, deletion, permission changes, and so on). To minimize the impact on performance, ensure that the host has enough CPU and memory resources to handle the File Analytics VM (FAVM), file servers, and guest VMs (if any).

- Only Prism admin can deploy File Analytics.
- File Analytics analyzes data from daily up to 1 year based on the configuration. File Analytics automatically deletes data beyond the defined configuration.

  > **Note:** After surpassing the audit event threshold, as specified in *File Analytics Release Notes*, Analytics archives the oldest events. Archived audit events do not appear in the Analytics UI.

- You cannot deploy or decommission File Analytics when a file server has high-availability (HA) mode enabled.
- You cannot use network segmentation for Nutanix Volumes with File Analytics.
- If file server DNS or IP changes, File Analytics does not automatically reconfigure.
- File Analytics does not collect metadata for files on Kerberos authenticated NFS v4.0 shares.
- File Analytics does not support hard links.
- You cannot enable File Analytics on a file server clone.

- You cannot move File Analytics to another storage container.
- File Analytics creates an unprotected Prism and an unprotected file server user for integration purposes. Do not delete these users.
- File Analytics does not support the following operations for graceful shutdown:
  - AHV: power cycle, power off
  - ESXi: power off, reset

ADMINISTRATION

Overview of administrative processes for File Analytics.

As an admin, you have the required permissions for performing File Analytics administrative tasks. To add a file server admin user, see *Managing Roles* in the *Nutanix Files Guide*. The topics in this chapter describe the basics for administering your File Analytics environment. For advanced administrative options, refer to the File Analytics Options on page 57 chapter.

Deploying File Analytics

Follow this procedure to deploy the File Analytics server.

Before you begin

Ensure that your environment meets all requirements prior to deployment, see Deployment Requirements on page 5.

Procedure

Deploying the File Analytics server.

1. Go to **Support Portal** > **Downloads** > **File Analytics**.
2. Download the File Analytics QCOW2 and JSON files.
3. Log on to Prism with the user name and password of the Prism administrator.

   Note: An Active Directory (AD) user or an AD user mapped to a Prism admin role cannot deploy File Analytics.

4. In Prism, go to the **File Server** view and click the **Deploy File Analytics** action link.

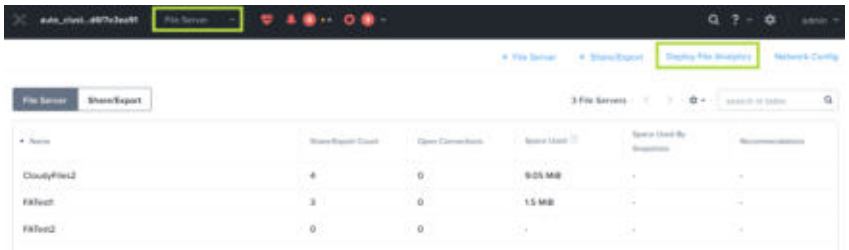


Figure 2: File Analytics

5. Review the File Analytics requirements and best practices in the **Pre-Check** dialog box.
6. In the **Deploy File Analytics Server** dialog box, do the following in the **Image** tab.
   - » Under **Available versions**, select one of the available File Analytics versions. (continue to step 8.).
   - » Install by uploading installation binary files (continue to next step).

7. Upload installation files.

   a. In the **Upload binary** section, click **upload the File Analytics binary** to upload the File Analytics JSON and QCOW files.

Figure 3: Upload Binary Link

   b. Under **File Analytics Metadata File (.Json)**, click **Choose File** to choose the downloaded JSON file.

   c. Under **File Analytics Installation Binary** (.Qcow2) click **Choose File** to choose the downloaded QCOW file.

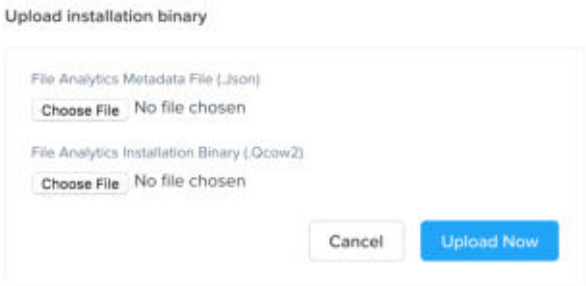


Figure 4: Upload Binary Files

   d. Click **Upload Now** after choosing the files.

8. Click **Next**.

9. In the **VM Configuration** tab, do the following in the indicated fields:

   a. **Name**: Enter a name for the File Analytics VM (FAVM).

   b. **Server Size**: Select either the **small** or **large** configuration. Large file servers require larger configurations for the FAVM. By default File Analytics selects the large configuration.

   c. **Storage Container:** select a storage container from the drop-down.

      The drop-down only displays file server storage containers.

   d. **Network List**: Select a VLAN.

> **Note:** If the selected network is **unmanaged**, enter more network details in the **Subnet Mask**, **Default Gateway IP**, and **IP Address** fields as indicated.

> **Note:** The FAVM must use the client-side network.

10. Click **Deploy**.

    In the main menu drop-down, select the **Tasks** view to monitor the deployment progress.

Results

Once deployment is complete, File Analytics creates an FAVM, CVM, and a new Files user to make REST API calls. Do not delete the CVM, FAVM, or the REST API user.

Enabling File Analytics

Steps for enabling File Analytics after deployment or disablement.

About this task

> **Attention:** Nutanix recommends enabling File Analytics during off-peak hours.

Follow these steps to enable File Analytics after disabling the application.

> **Note:** File Analytics saves all previous configurations.

Procedure

1. In the **File Server** view in **Prism**, select the target file server.
2. (skip to step 3 if you are re-enabling a file server) click **Manage roles** to add a file server admin user, see *Managing Roles* in the *Nutanix Files Guide*.
3. In the **File Server** view, select the target file server and click **File Analytics** in the tabs bar.
4. (Skip to step 5 if you are not re-enabling a disabled instance of File Analytics) to re-enable File Analytics, click **Enable File Analytics** in the message bar.

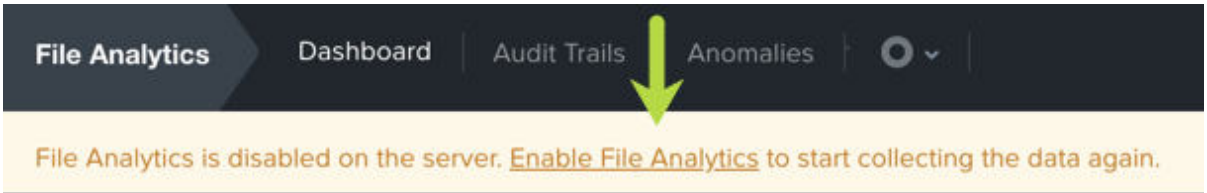


Figure 5: Enabling File Analytics Link

The **Enable File Analytics** dialog-box appears. Skip the remaining steps.

5. In the **Data Retention** field, select a data retention period. The data retention period refers to the length of time File Analytics retains audit events.

6. In the **Authentication** section, enter the credentials as indicated:

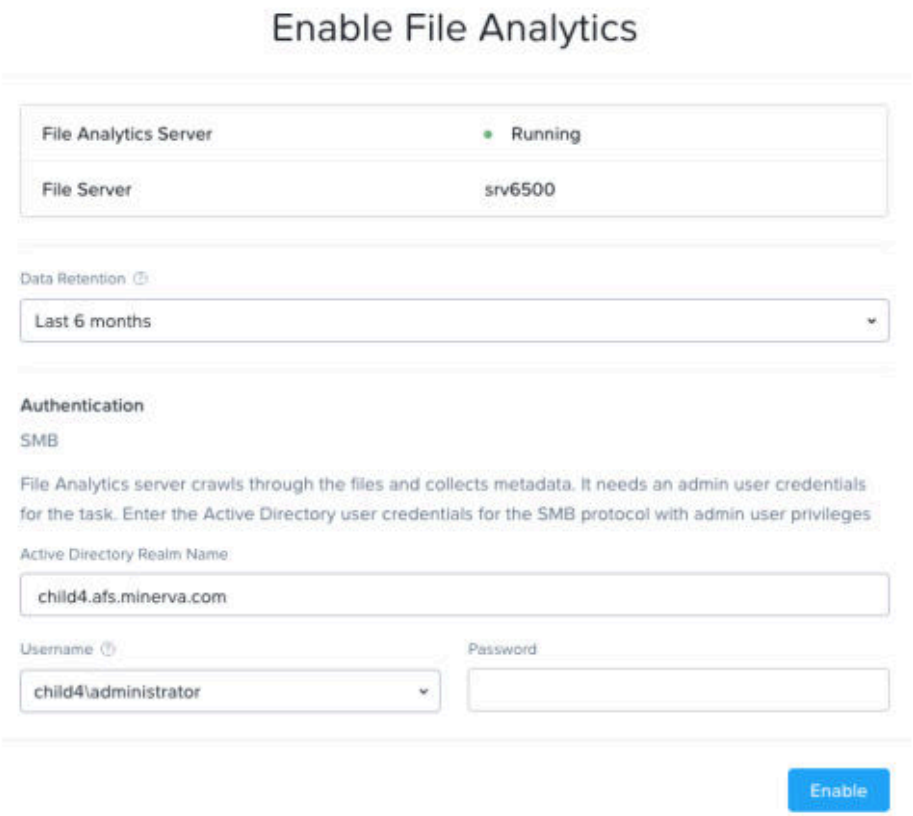
> **Note:** AD passwords for the file server admin cannot contain the following special characters: comma (,), single quote ('), double quote ("). Using the special characters in passwords prevents File Analytics from performing file system scans.

a. (For SMB users only) In the **SMB** section, do the following in the indicated fields to provide SMB authentication details:

- **Active Directory Realm Name**: Confirm the AD realm name for the file server.
- **Username**: Enter the AD username for the file server administrator, see **File Analytics Prerequisites**.
- **Password**: Enter the AD user password for the file server administrator.

b. (For NFS users only) In the **NFS Authentication** section, do the following in the indicated fields to provide NFS authentication details:

- **LDAP Server URI**: Enter the URI of the LDAP server.
- **Base DN**: Enter the base DN for the LDAP server.
- **Password**: Enter the LDAP user password for the file server administrator.



7. Click **Enable** .

Results

After enablement, File Analytics performs a one-time file system scan to pull metadata information. The duration of the scan varies depending on the protocol of the share. There is no system downtime during the scan.

Example

Scanning 3–4 million NFS files or 1 million SMB files takes about 1 hour.

Disabling File Analytics

About this task

Follow the steps as indicated to disable File Analytics.

Procedure

1. In **File Analytics** click the **gear icon** > **Disable File Analytics** .
2. In the dialog-box, click **Disable**.
   Disabling File Analytics disables data collection. The following message banner appears.

```
File Analytics is disabled on the server. Enable File Analytics to start collecting data
again or Delete File Analytics Data.
```

What to do next

To delete data, click the **Delete File Analytics Data** link in the banner described in Step 2.

Launching File Analytics

About this task

Do the following to launch File Analytics.

Procedure

1. From the Prism views drop-down, select the **File Server** view.
2. Select the target file server from the entity tab.
3. Click the **File Analytics** action button below the entity table.

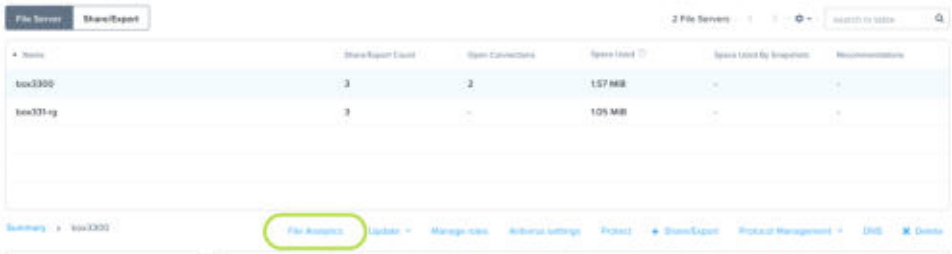


Figure 6: Launch File Analytics

File Analytics VM Management

To update a File Analytics VM (FAVM), refer to the sizing guidelines in the File Analytics release notes and follow the steps in the *VM Management* topic of the *Prism Web Console Guide*.

Removing File Analytics VMs

Remove a File Analytics VM (FAVM) by disabling it and deleting it from the cluster in Prism.

About this task

Follow the steps as indicated to remove an FAVM.

> **Note:** Do not delete an FAVM using the CLI, as this operation does not decommission the FAVM.

Procedure

1. Disable File Analytics on all file servers in the cluster, see Disabling File Analytics on page 11.
2. In the **File Server** view in Prism Element, do the following:
   a. In the top actions bar, click **Manage File Analytics**.
   b. Click Delete to remove the FAVM.
   When you delete an FAVM, you also delete all of your File Analytics configurations and audit data stored on the FAVM.

Updating Credentials

About this task

Follow the steps as indicated to update authentication credentials for LDAP or Active Directory.

Procedure

1. Click **gear icon** > **Update AD/LDAP Configuration**.
2. To update Active Directory credentials, do the following in the indicated fields (otherwise move on to the next step).

   > **Note:** AD passwords for the file server admin cannot contain the following special characters: comma (,), single quote ('), double quote ("). Using the special characters in passwords prevents File Analytics from performing file system scans.

   a. **Active Directory Realm Name:** confirm or replace the realm name.
   b. **Username:** confirm or replace the username.
   c. **Password:** type in the new password.
3. To update NFS configuration, do the following (otherwise move on to the next step).
   a. **LDAP Server URI:** confirm or replace the server URI.
   b. **Base DN:** confirm or replace the base distinguished name (DN).
   c. **Bind DN (Optional):** confirm or replace the bind distinguished name (DN).
   d. **Password:** type in the new password.

4. Click **Save**.

Managing Deleted Share/Export Audits

Manage the audit data of delete shares and exports.

About this task

By default, File Analytics retains deleted share and export data. The dashboard widgets do not account for data of deleted shares and exports. The **deleted** marker appears next to deleted shares and exports in audit trails. The **Manage Share/Export Audit data** window displays a list of deleted shares and exports.

Follow the directions as indicated to delete audit data for the deleted share or export.

> **Note:** You cannot restore the deleted audit data of a deleted share or export.

Procedure

1. Click the **gear icon** > **Manage Deleted Share/Export Audit**.
2. Check the box next to the share or export name.
3. Click **Delete**.
4. In the confirmation window, click **Delete** to confirm the deletion of data.
   In the **Manage Deleted Share/Export Audit**, a progress bar displays the progress of the deletion process next to the share name. File Analytics considers data deletion of a deleted share a low-priority task, which can take several hours to finish.

Changing an FAVM Password

Steps for updating the password of a File Analytics VM (FAVM).

About this task

Context for the current task

Procedure

1. Log on to an FAVM with SSH.
2. Change the nutanix password.

   ```
   nutanix@fsvm$ sudo passwd nutanix
   ```

3. Respond to the prompts, providing the current and new nutanix user password.

   ```
   Changing password for user nutanix.
   Old Password:
   New password:
   Retype new password:
   ```

```
passwd: all authentication tokens updated successfully.
```

> **Note:**
>
> The password must meet the following complexity requirements:
>
> - At least 8 characters long
> - At least 1 lowercase letter
> - At least 1 uppercase letter
> - At least 1 number
> - At least 1 special character
> - At least 4 characters difference from the old password
> - Should not be among the last 10 passwords

Upgrades

Perform File Analytics upgrades using the Life Cycle Manager feature in Prism Element.

Before you upgrade File Analytics, ensure that you are running a compatible version of AOS and Files. Refer to File Analytics release notes for compatibility details. You can upgrade both AOS and Files through Prism Element, see *AOS Upgrade* in the *Prism Web Console Guide*.

To upgrade File Analytics, perform inventory and updates using the Life-Cycle Manager (LCM), see the *Life Cycle Manager Guide* for instructions on performing inventory and updates. LCM cannot upgrade File Analytics when the protection domain (PD) for the File Analytics VM (FAVM) includes any other entities.

> **Note:** The File Analytics UI is not accessible during upgrades.

During the upgrade process, File Analytics takes a snapshot of the volume group (VG) that contains File Analytics data. If issues occur during an upgrade, File Analytics restores the FAVM to the pre-upgrade state. If the volume group is protected and is part a protection domain, the File Analytics creates a snapshot and sets the expiry time to 30 days. If the volume group is not protected, File Analytics creates a snapshot and deletes the snapshot after completing the upgrade successfully. If any errors occur, the system keeps the snapshot for 30 days to troubleshoot the issue.

Upgrade File Analytics at a Dark Site

Upgrade File Analytics at a dark site using the Life-Cycle Manager (LCM).

About this task

Before you begin

You need a local web server reachable by your Nutanix clusters to host the LCM repository.

Procedure

1. From a device that has public Internet access, go to **Nutanix Portal** > **Downloads** > **Tools & Firmware**.
   a. Download the tar file `lcm_dark_site_version.tar.gz`.
   b. Transfer `lcm_dark_site_version.tar.gz` to your local web server and untar into the `release` directory.

2. From a device that has public Internet access, go to the Nutanix portal and select **Downloads** > **File Analytics**.
   a. Download the following files.
      - `file_analytics_dark_site_version.tar.gz`
      - `nutanix_compatibility.tgz`
      - `nutanix_compatibility.tgz.sign`
   b. Transfer `file_analytics_dark_site_version.tar.gz` to your local web server and untar into the `release` directory.
   c. Transfer the `nutanix_compatibility.tgz` and `nutanix_compatibility.tgz.sign` files to your local web server (overwrite existing files as needed).
3. Log on to Prism Element.
4. Click **Home** > **LCM** > > **Settings**.
   a. In the **Fetch updates from** field, enter the path to the directory where you extracted the tar file on your local server. Use the format `http://webserver_IP_address/release`.
   b. Click **Save**.
      You return to the Life Cycle Manager.
   c. In the LCM sidebar, click **Inventory** > **Perform Inventory**.
   d. Update the LCM framework before trying to update any other component.
      The LCM sidebar shows the LCM framework with the same version as the file you downloaded.

DASHBOARD

The Dashboard tab displays data on the operational trends of a file server.

Dashboard View

The **Dashboard** tab is the opening screen that appears after launching File Analytics from Prism. The dashboard displays widgets that present data on file trends, distribution, and operations.

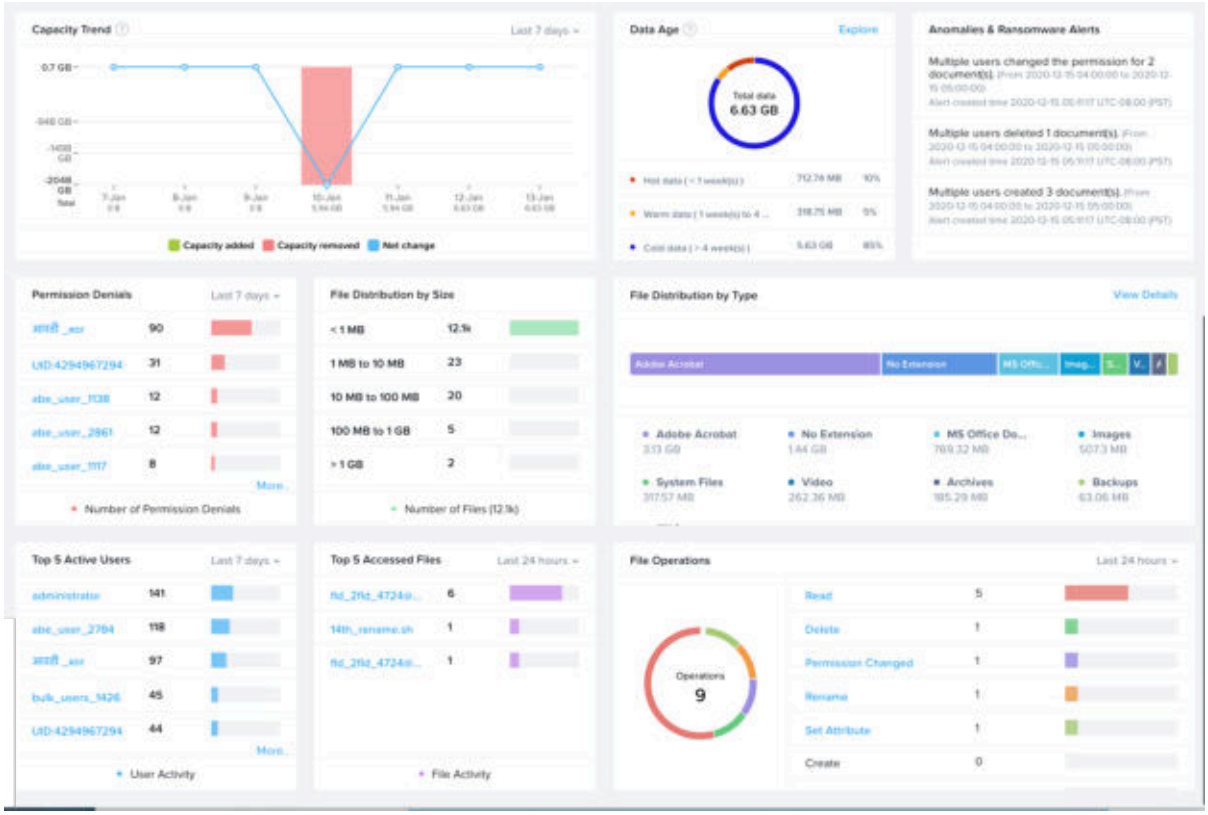


Figure 7: File Analytics Dashboard

Table 1: Dashboard Widgets

Tile Name	Description	Intervals
Capacity trend	Displays capacity trends for the file server including capacity added, capacity removed, and net changes. Clicking an event period widget displays the **Capacity Trend Details** view.	7 days, the last 30 days, or the last 1 year.

Tile Name	Description	Intervals
Data age	Displays the percentage of data by age. Data age determines the data heat, including: hot, warm, and cold.	Default intervals are as follows: • Hot data – accessed within the last week. • Warm data – accessed within 2 to 4 weeks. • Cold data – accessed 4 weeks ago or later.
Anomaly alerts	Displays alerts for configured anomalies and ransomware detection based on blocked file types, see Configuring Anomaly Detection on page 26.	[alert]
Permission denials	Displays users who have had excessive permission denials and the number of denials. Clicking a user displays audit details, see Audit Trails - Users on page 30 for more.	[user id], [number of permission denials]
File distribution by size	Displays the number of files by file size. Provides trend details for top 5 files.	Less than 1 MB, 1–10 MB, 10–100 MB, 100 MB to 1 GB, greater than 1 GB).
File distribution by type	Displays the space taken up by various applications and file types. The file extension determines the file type. See the *File types* table for more details.	MB or GB
File distribution by type details view	Displays a trend graph of the top 5 file types. File distribution details include file type, current space used, current number of files, and change in space for the last 7 or 30 days. Clicking **View Details** displays the **File Distribution by Type** view.	Daily size trend for top 5 files (GB), file type (see the "File Type" table), current space used (GB), current number of files (numeric), change in last 7 or 30 days (GB).
Top 5 active users	Lists the users who have accessed the most files and number of operations the user performed for the specified period. When there are more than 5 active users, the **more** link provides details on the top 50 users. Clicking the user name displays the audit view for the user, see Audit Trails - Users on page 30 for more.	24 hours, 7 days, 1 month, or 1 year.

Tile Name	Description	Intervals
Top 5 accessed files	Lists the 5 most frequently accessed files. Clicking **more** provides details on the top 50 files. Clicking the file name displays the audit view details for the file, see Audit Trails - Files on page 33 for more.	24 hours, 7 days, 1 month, or 1 year.
Files operations	Displays the distribution of operation types for the specified period, including a count for each operation type and the total sum of all operations. Operations include: create, delete, read, write, rename, permission changed, set attribute, symlink, permission denied, permission denied (file blocking). Clicking an operation displays the **File Operation Trend** view.	24 hours, 7 days, 1 month, or 1 year.

Capacity Trend Details

Clicking an event period in the **Capacity Trend** widget displays the **Capacity Trend Details** view for that period. The view includes three tabs **Share/Export**, **Folder**, and **Category**. Each tab includes columns detailing entity details: **Name**. Net capacity change, capacity added, and capacity removed.

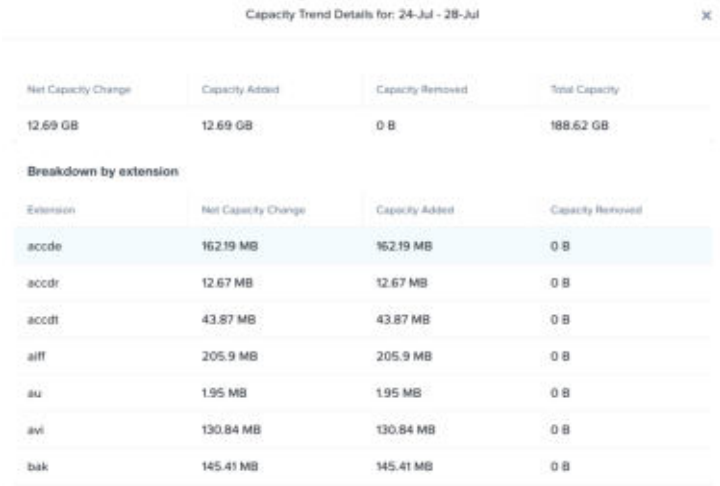


Figure 8: Capacity Trend Details View

Table 2: Capacity Trend Details

Category	Supported File Type
Name	Name of share/export, folder, or category.

Category	Supported File Type
Net capacity change	The total difference between capacity at the beginning and the end of the specified period.
Share name (for folders only)	The name of the share or export that the folder belongs to.
Capacity added	Total added capacity for the specified period.
Capacity removed	Total removed capacity for the specified period.

File Distribution by Type Details

Clicking **View Details** for the **File Distribution by Type** widget displays granular details of file distribution, see the **File Types** table for details.

Figure 9: File Distribution by Type

Table 3: Details of File Distribution Parameters

Category	Supported File Type
File type	Name of file type
Current space used	Space capacity occupied by the file type
Current number of files	Number of files for the file type
Change (in last 30 days)	The increase in capacity over a 30-day period for the specified file type

Table 4: File Types

Category	Supported File Type
Archives	`.cab, .gz, .rar, .tar, .z, .zip`
Audio	`.aiff, .au, .mp3, .mp4, .wav, .wma`

Category	Supported File Type
Backups	.bak, .bkf, .bkp
CD/DVD images	.img, .iso, .nrg
Desktop publishing	.qxd
Email archives	.pst
Hard drive images	.tib, .gho, .ghs
Images	.bmp, .gif, .jpg, .jpeg, .pdf .png, .psd, .tif, .tiff,
Installers	.msi, .rpm
Log Files	.log
Lotus notes	.box, .ncf, .nsf, .ns2, .ns3, .ns4, .ntf
MS Office documents	.accdb, .accde, .accdt, .accdr, .doc, .docx, .docm, .dot, .d
System files	.bin, .dll, .exe
Text files	.csv, .pdf, .txt
Video	.avi, mpg, .mpeg, .mov, .m4v
Disk image	.hlog, .nvram, .vmdk, .vmx, .vmxf, .vmtm, .vmem, .vmsn, .vms

File Operation Trend

Clicking an operation type in the **File Operations** widget displays the **File Operation Trend** view. The **File Operation Trend** view breaks down the specified period into smaller intervals, and displays the number of occurrences of the operation during each interval.

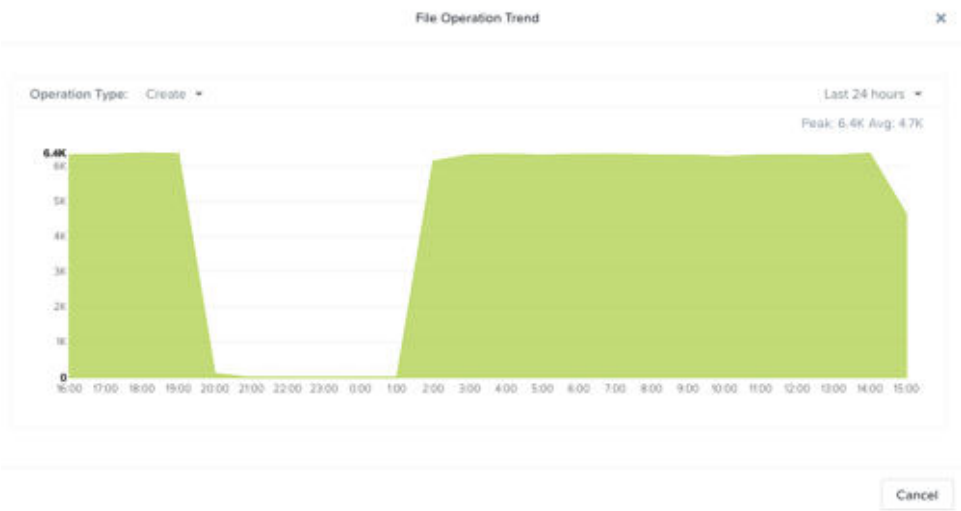


Figure 10: Operation Trend

Table 5: File Operation Trend View Parameters

Category	Description
Operation type	A drop-down option to specify the operation type. See *Files Operations* in the *Dashboard Widgets* table for a list of operation types.
Last (time period)	A drop-down option to specify the period for the file operation trend.
File operation trend graph	The x-axis displays shorter intervals for the specified period. The y-axis displays the number of operations trend over the extent of the intervals.

Health

The Health dashboard displays dynamically updated health information about each File File Analytics component.

The Health dashboard includes the following details:

- **Data Summary** Data summary of all file servers with File Analytics enabled.
- **Host Memory** Percent of used memory on the File Analytics VM (FAVM).
- **Host CPU Usage** Percent of CPU used by the FAVM.
- **Storage Summary** Amount of storage space used on the File Analytics data disk or FAVM disk.
- **Overall Health** Overall health of File Analytics components.
- **Data Server Summary** Data server usage by component.

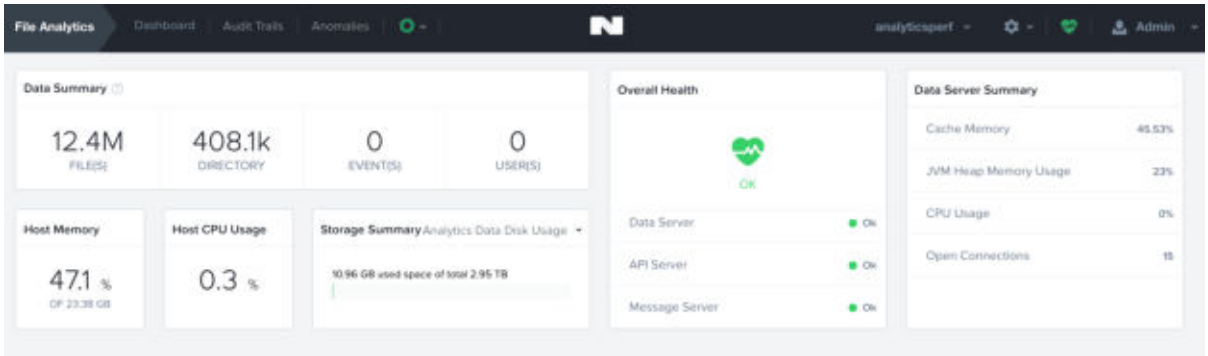


Figure 11: Health Page

Data Age

The Data Age widget in the Dashboard provides details on data heat.

Share-level data is displayed to provide details on share capacity trends. There are three levels of data heat.

- *Hot* – frequently accessed data (last accessed within the last week).
- *Warm* – infrequently accessed data (last accessed within the last 2 to 4 weeks).
- *Cold* – rarely accessed data (last accessed longer than 4 weeks ago).

You can configure the definitions for each level of data heat rather than using the default values.

Configuring Data Heat Levels

Update the values that constitute different data heat levels.

Procedure

1. In the **Data Age** widget, click **Explore**.
2. Click **Edit Data Age Configuration**.
3. Do the following in the **Hot Data** section:
   a. In the entry field next to **Older Than**, enter an integer.
   b. In the dropdown, choose a value for **Week(s)**, **Month(s)**, or **Year(s)**.
4. Do the following in the **Warm Data** section to configure two ranges :
   a. In the first entry field, enter an integer to configure the first range.
   b. In the dropdown, choose a value for **Week(s)**, **Month(s)**, or **Year(s)**.
   c. In the second entry field, enter an integer to configure the second range.
   d. In the dropdown, choose a value for **Week(s)**, **Month(s)**, or **Year(s)**.
5. Do the following in the **Cold Data** section to configure four ranges :
   a. In the first entry field, enter an integer to configure the first range.
   b. In the dropdown, choose a value for **Week(s)**, **Month(s)**, or **Year(s)**.
   c. In the second entry field, enter an integer to configure the second range.
   d. In the dropdown, choose a value for **Week(s)**, **Month(s)**, or **Year(s)**.
   e. In the 3rd entry field, enter an integer to configure the 3rd range.
   f. In the dropdown, choose a value for **Week(s)**, **Month(s)**, or **Year(s)**.
   g. (optional) In the 4th entry field, enter an integer to configure the 4th range.
   h. In the dropdown, choose a value for **Week(s)**, **Month(s)**, or **Year(s)**.
6. Click **Apply**.

> **Note:** The new values do not affect the already calculated heat statistics. File Analytics uses the updated values for future heat calculations..

ANOMALIES

Data panes in the **Anomalies** tab display data and trends for configured anomalies.

The Anomalies tab provides options for creating anomaly policies and displays dashboards for viewing anomaly trends.

> **Note:** Configure an SMTP server to send anomaly alerts, see Configuring an SMTP Server on page 27

You can configure anomalies for the following operations:

- Creating files and directories
- Deleting files and directories
- Permission changes
- Permission denials
- Renaming files and directories
- Reading files and directories

Define anomaly rules by the specifying the following conditions:

- Users exceed an operation count threshold
- Users exceed an operation percentage threshold

Meeting the lower operation threshold triggers an anomaly.

Consider a scenario where you have 1 thousand files, the `operation count threshold` defined as 10, and the `operation percentage threshold` defined as 10%. The count threshold takes precedence, as 10% of 1 thousand is 100, which is greater than the count threshold of 10.

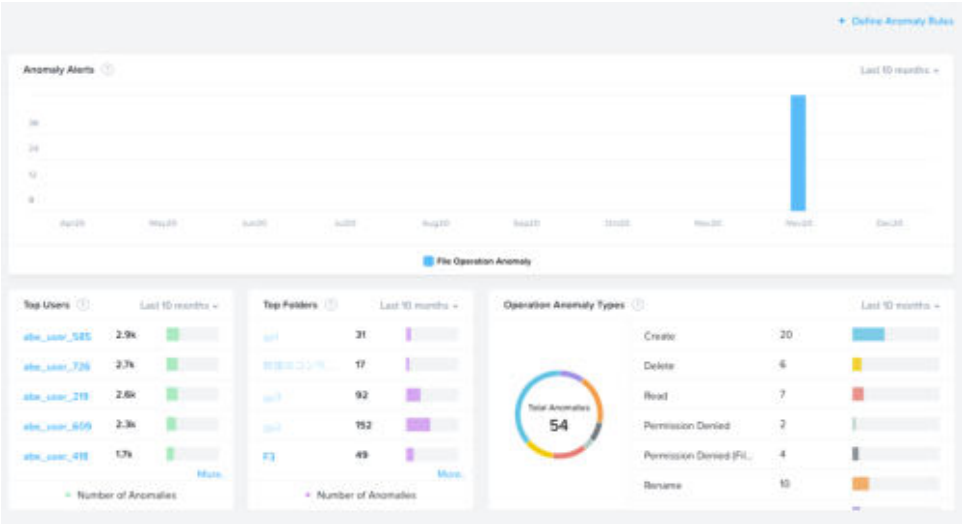


Figure 12: Anomalies Dashboard

Table 6: Anomalies Data Pane Descriptions

Pane Name	Description	Values
Anomaly Trend	Displays the number of anomalies per day or per month.	Last 7 days, Last 30 days, Last 1 year
Top Users	Displays the users with the most anomalies and the number of anomalies per user.	Last 7 days, Last 30 days, Last 1 year
Top Folders	Displays the folders with the most anomalies and the number of anomalies per folder.	Last 7 days, Last 30 days, Last 1 year
Operation Anomaly Types	Displays the percentage of occurrences per anomaly type.	Last 7 days, Last 30 days, Last 1 year

Anomaly Details

Clicking an anomaly bar in the **Anomaly Trend** graph displays the **Anomaly Details** view.

Anomaly Details for: 28-Jul

48 Total Results

1–5 of 48

Anomaly Type	Total User Count	Total Folder Count	Total Operation Count	Time Range
FileRead	93	13244	54507	2020-07-28 15:00:00 - 2020-07-28 16:00:00
Delete	88	5183	6361	2020-07-28 15:00:00 - 2020-07-28 16:00:00
Rename	87	1863	2540	2020-07-28 15:00:00 - 2020-07-28 16:00:00
Create	87	5219	6311	2020-07-28 15:00:00 - 2020-07-28 16:00:00
FileRead	94	13274	54738	2020-07-28 14:00:00 - 2020-07-28 15:00:00

Users Folders

1–10 of 93

Username	Operation Count
kraken\kraken_user16	651
kraken\kraken_user84	649
kraken\kraken_user12	644
kraken\kraken_user78	644
kraken\kraken_user94	644
kraken\kraken_user14	643
kraken\kraken_user42	642
kraken\kraken_user18	639
kraken\kraken_user55	639
kraken\kraken_user8	638

Figure 13: Anomaly Details View

Table 7: Anomalies Details View Total Results Table

Column	Description
Anomaly Type	The configured anomaly type. Anomaly types not configured do not show up in the table.
Total User Count	The number of users that have performed the operation causing the specified anomaly during the specified time range.
Total Folder Count	The numbers of folders in which the anomaly occurred during the specified time range.

Column	Description
Total Operation Count	Total number of anomalies for the specified anomaly type that occurred during the specified time range.
Time Range	The time range for which the total user count, total folder count, and total operation count are specified.

Table 8: Anomalies Details View Users/Folders Table

Column	Description
Username or Folders	Indicates the entity for the operation count. Selecting the **Users** tab indicates operation count for specific users, and selecting the **Folders** tab indicates the operation count for specific folders.
Operation count	The total number of operations causing anomalies for the selected user or folder during the time period for the bar in the **Anomaly Trend** graph.

Configuring Anomaly Detection

Steps for configuring anomaly rules.

About this task

Configure an SMTP server for File Analytics to send anomaly alerts, see Configuring an SMTP Server on page 27. To create an anomaly rule, do the following.

Procedure

1. In the File Analytics web console, click the **gear icon** > **Define Anomaly Rules.** .
2. In the **Anomaly Email Recipients** field, enter a comma-separated list of email recipients for all anomaly alerts and data.

> **Note:** File Analytics sends anomaly alerts and data to recipients whenever File Analytics detects an anomaly.

3. To configure a new anomaly, do the following in the indicated fields:

   a. **Events**: Select a rule for the anomaly from one of the following:

      - Permission changed
      - Permission denied
      - Delete
      - Create
      - Rename
      - Read

      The event defines the scenario type for the anomaly.

   b. **Minimum Operations %**: Enter a percentage value for the minimum threshold.

      File Analytics calculates the minimum operations percentage based on the number of files. For example, if there are 100 files, and you set the minimum operations percentage to 5, five operations within the scan interval would trigger an anomaly alert.

   c. **Minimum Operation Count**: Enter a value for a minimum operation threshold.

      File Analytics triggers an anomaly alert after meeting the threshold.

   d. **User**: Choose if the anomaly rule is applicable for **All Users** or an **Individual** user.

   e. **Type:** the type determines the interval.

      The interval determines how far back File Analytics monitors the anomaly.

   f. **Interval** : Enter a value for the detection interval.

   g. (optional) **Actions**: Click the **pencil icon** to update an anomaly rule. Click the **x** icon to delete an existing rule.

EVENTS	MINIMUM OPERATION %	MINIMUM OPERATION COUNT	USER	TYPE	INTERVAL	ACTIONS
Permission Changed	100	10	All Users	Hourly	1	
Permission Denied	100	1	All Users	Hourly	1	
Delete	100	10	Individual	Hourly	1	
Create	100	1000	Individual	Hourly	1	

Figure 14: Anomaly Configuration Fields

4. Click **Save**.

Configuring an SMTP Server

File Analytics uses a simple mail transport protocol (SMTP) server to send anomaly alerts.

About this task

To configure an SMTP server, do the following:

Procedure

1. In the File Analytics web console, click the **gear icon** > **SMTP Configuration**.

2. In the **SMTP Configuration** window, enter the indicated details in the following fields:

   a. **Hostname Or IP Address**: Enter a fully qualified domain name or IP address for the SMTP server.

   b. **Port**: Enter the port to use.
   The standard SMTP ports are 25 (encrypted), 587 (TLS), and 465 (SSL).

   c. **Security Mode**: Enter the desired security mode from the dropdown list.
   The options are:

      - **NONE** (unencrypted)
      - **STARTTLS** (TTL encryption)
      - **SSL** (SSL encryption)

   d. (If security mode is "NONE" go to step f.)

   e. **User Name** enter a user name for logging into the SMTP server. Depending on the authentication method, the user name may require a domain.

   f. **Password** enter password.

   g. **From Email Address:** enter the email address from which File Analytics will send the anomaly alerts.

   h. **Recipient Email Address:** enter a recipient email address to test the SMTP configuration.

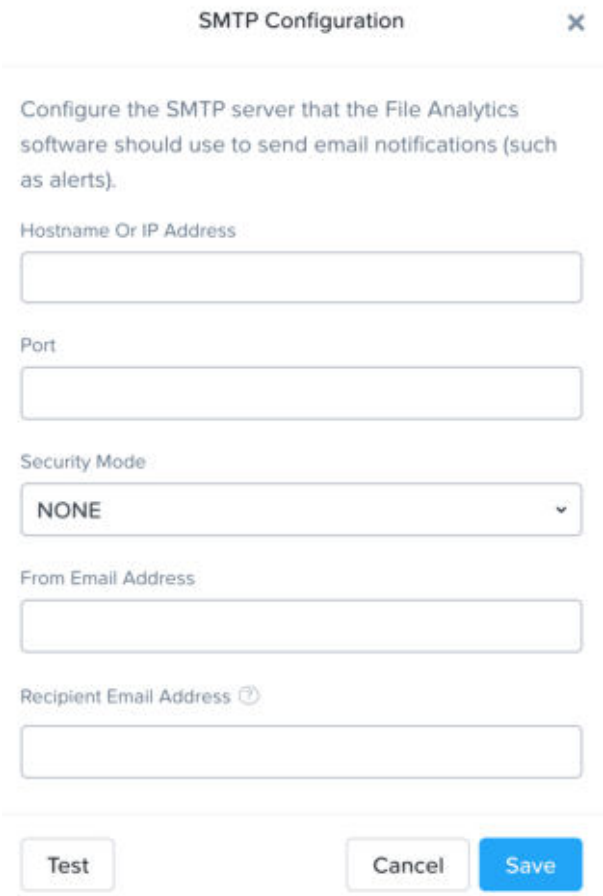


Figure 15: SMTP Configuration

3. Click **Save**.

AUDIT TRAILS

Use audit trails to look up operation data for a specific user, file, folder, or client.

The **Audit Trails** tab includes **Files**, **Folders**, **Users**, and **Client IP** options for specifying the audit type. Use the search bar for specifying the specific entity for the audit (user, folder, file, or client IP).

The results table presents details for entities that match the search criteria. Clicking the entity name (or client IP number) takes you to the Audit Trails dashboard for the target entity.

View Audit Trails

Audit a user, file, client, or folder.

About this task

Procedure

1. Click the **Audit Trails** tab.
2. Select the **Files**, **Folders**, **Users**, or **Client IP** option.
3. Enter the audit trails target into the search bar.
4. Click **Search**.
5. To display audit results in the **Audit Trails** window, click the entity name (or client IP number).

Audit Trails - Users

Details for client IP Audit Trails.

Audit Trails Search - Users

When you search by user in the **Audit Trails** tab, search results display the following information in a table.

- User Name
- Domain
- Last Operation
- Last Operation On
- Share Name
- Operation Date
- Action

Figure 16: Users Search Results

Audit Details Page - Users

Clicking **View Audit** displays the **Audit Details** page, which shows the following audit information for the selected user.

- A **User Events** graph displays various operations the user performed during the selected period and the percentage of time each operation has occurred per total operations during the specified period.
  - The **Filter by operations** dropdown contains operation filters, which you can use to filter the audit by operation type. Operations include:
    - Create File
    - Delete
    - Make Directory
    - Permission Changed
    - Permission Denied
    - Read
    - Remove Directory
    - Rename
    - Set Attribute
    - Write
    - Symlink
  - The **filter bar**, above the User Events graph, displays the filters in use.
  - Use the **From** and **To** fields to filter by date.
- The **Results** table displays operation-specific information. See more details below.
- The **Reset Filters** button removes all filters.

Figure 17: User Audit Details - Events

The **Results** table provides granular details of the audit results. The following data is displayed for every event.

- User Name
- User IP Address

- Operation
- Operation Date
- Target File

Click the **gear icon** for options to download the data as an xls, csv, or JSON file.

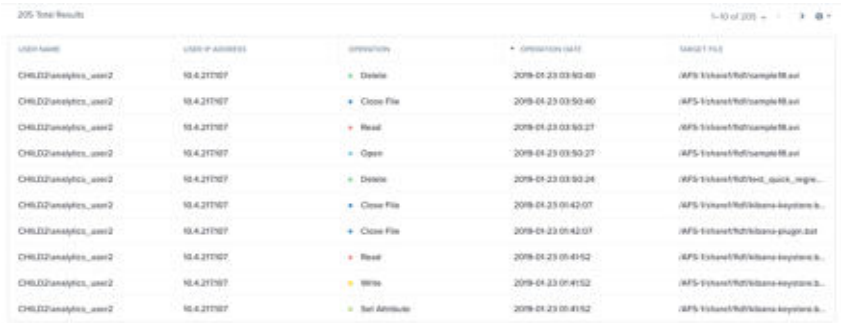


Figure 18: Results Table

Audit Trails - Folders

Dashboard details for folder audits.

The following information displays when you search by file in the Audit Trails tab.

- Folder Name
- Folder Owner Name
- Share Name
- Parent Folder
- Last Operation
- Last Operation By
- Last Operation Date
- Action

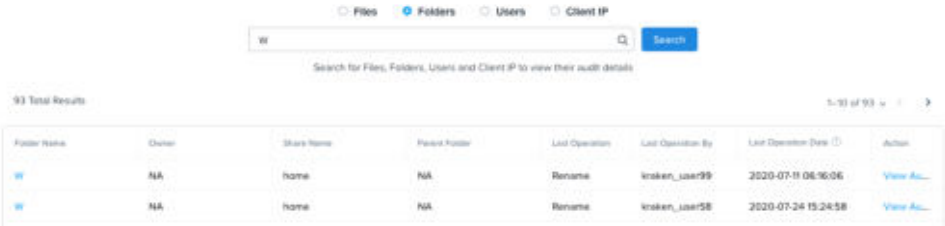


Figure 19: Folders Search Results

The **Audit Details** page shows the following audit information for the selected folder.

- A **Folder Events** graph displays various operations performed on the file during the selected period, and the percentage of time each operation has occurred per total operations during that period.
  - The **Filter by operations** dropdown contains operation filters, which you can use to filter the audit by operation type. Operations include:
    - Select All
    - Make Directory
    - Permission Changed
    - Permission Denied
    - Remove Directory
    - Rename
    - Set Attribute
  - A **filter bar**, above the File Events graph displays the filters in use.
  - Use the **From** and **to** fields to filter by date.
- The **Results** table displays operation-specific details.
- The **Reset Filters** button removes all filters.

The **Results** table provides granular details of the audit results. File Analytics displays the following data for every event.

- User Name
- Client IP
- Operation
- Operation Date

Click the gear icon for options to download the data as a CSV file.

Audit Trails - Files

Dashboards details for file audit.

Audit Trails for Files

When you search by file in the Audit Trails tab, the following information displays:

- File Name
- File Owner Name
- Share Name
- Parent Folder
- Last Operation
- Last Operation By
- Last Operation Date
- Action

Figure 20: Files Search Results

> **Note:** File Analytics does not support regular-expression (RegEx) based search.

The **Audit Details** page shows the following audit information for the selected file.

- A **File Events** graph displays various operations performed on the file during the selected period, and the percentage of time each operation has occurred per total operations during that period.
  - The **Filter by operations** dropdown contains operation filters, which you can use to filter the audit by operation type. Operation types include:
    - Close File
    - Create File
    - Delete
    - Make Directory
    - Open
    - Read
    - Rename
    - Set Attribute
    - Write
    - Symlink
  - A **filter bar**, above the File Events graph displays the filters in use.
  - Use the **From** and **to** fields to filter by date.
- The **Results** table displays operation-specific details.
  - The **Reset Filters** button removes all filters.

Figure 21: Files Audit Details - Events

The **Results** table provides granular details of the audit results. File Analytics displays the following data for every event.

- Username
- Client IP
- Operation
- Operation Date

Click the gear icon for options to download the data as a CSV file.

Figure 22: Results Table

Audit Trails - Client IP

Dashboard details for client IP Audit Trails.

Audit Trails Search - Client IP

When you search by client IP in the **Audit Trails** tab, search results display the following information in a table.

- Client IP
- User Name
- Domain
- Last Operation
- Last Operation On
- Share Name
- Operation Date
- Action

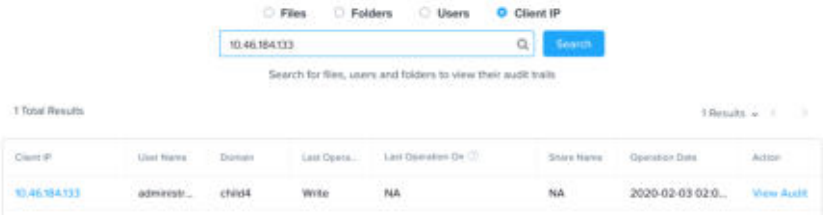


Figure 23: IP Search Results

The **Audit Details** page shows the following audit information for the selected client.

- A **User Events** graph displays various operations performed on the client during the selected period, and the percentage of time each operation has occurred per total operations during that period.
  - The **Filter by operations** dropdown contains operation filters, which you can use to filter the audit by operation type. Operation types include:
    - Create File
    - Delete
    - Make Directory
    - Permission Changed
    - Permission Denied
    - Read
    - Removed Directory
    - Rename
    - Set Attribute
    - Write
    - Symlink
    - Permission Denied (File Blocking)
  - A **filter bar**, above the File Events graph displays the filters in use.
  - Use the **From** and **to** fields to filter by date.
- The **Results** table displays operation-specific details.
  - The **Reset Filters** button removes all filters.

Figure 24: Files Audit Details - Events

The **Results** table provides granular details of the audit results. File Analytics displays the following data for every event.

- User Name
- Operation
- Target File
- Operation Date

Click the gear icon for an option to download the data as a CSV file.

RANSOMWARE PROTECTION

Ransomware protection for your file server.

> **CAUTION:** Ransomware protection helps detect potential ransomware. Nutanix does not recommend using the File Analytics ransomware feature as an all-encompassing ransomware solution.

File Analytics scans files for ransomware in real time, and notifies you through email in the event of a ransomware attack. By using the Nutanix Files file blocking mechanism, File Analytics prevents files with signatures of potential ransomware from carrying out malicious operations. Ransomware protection automatically scans for ransomware based on a curated list of signatures that frequently appear in ransomware files. You can modify the list by manually adding other signatures.

> **Note:** Nutanix does not recommend manipulating the blocked signatures through Nutanix Files.

File Analytics also monitors shares for self-service restore (SSR) policies and identifies shares that do not have SSR enabled in the ransomware dashboard. You can enable SSR through the ransomware dashboard by selecting shares identified by File Analytics.

Ransomware Protection Features

The ransomware dashboard includes panes for managing ransomware protection and self-service restore (SSR).

Ransomware Dashboard

The ransomware dashboard includes two main sections:

- The **SSR Status** pane for viewing, enabling, and managing SSR, see Enabling SSR on page 48.
- The **Vulnerabilities (Infection Attempts)** pane for viewing total vulnerabilities, vulnerable shares, malicious clients, and top recent ransomware attempts.
  - Clicking on the number of **total vulnerabilities** provides a detailed view of recent vulnerabilities.
  - Clicking on the number of **vulnerable shares** provides a detailed view of vulnerable shares.
  - Clicking on the number of **malicious clients** provides a detailed view of malicious clients.
- Click **Settings**, to enable and configure ransomware protection, see Enabling Ransomware Protection on page 46 and Configuring Ransomware Protection on page 47.
- Click **Download (.csv)** to download a list of blocked ransomware signatures.

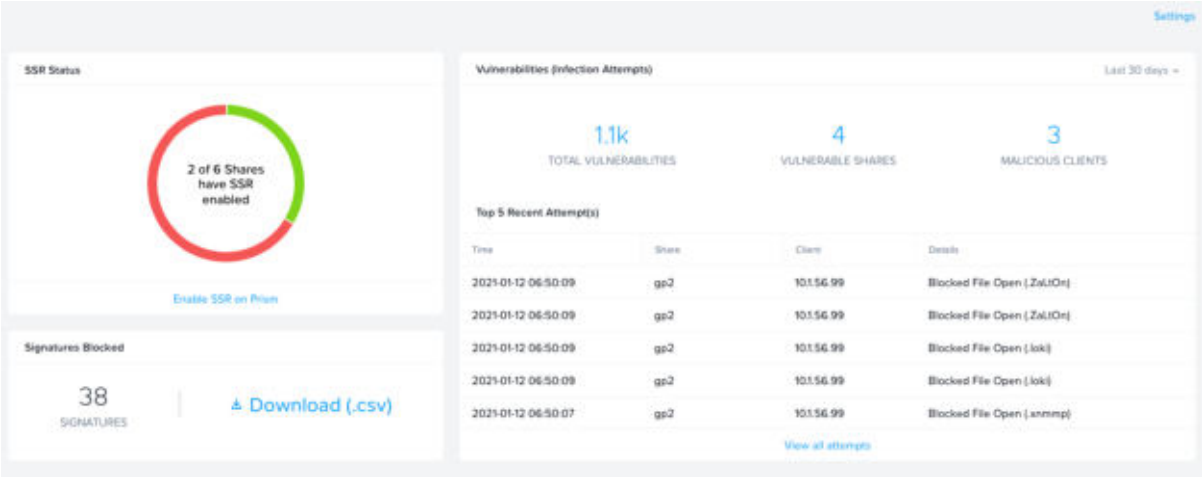


Figure 25: Ransomware Dashboard

Blocked Ransomware Extensions

File Analytics blocks the following ransomware signatures.

Table 9: Table

Extension	Known Ransomware
*.micro	eslaCrypt 3.0
*.zepto	Locky
*.cerber3	Cerber 3
*.locky	Locky
*.cerber	Cerber
*.loli	LOLI
*.mole	CryptoMix (variant)
*.cryp1	CryptXXX
*.axx	AxCrypt
*.onion	Dharma
*.crypt	Scatter
*.osiris	Locky (variant)
*.crypz	CryptXXX
*.ccc	TeslaCrypt or Cryptowall
*.locked	Various ransomware
*.odin	Locky
*.cerber2	Cerber 2
*.sage	Sage
*.globe	Globe

Extension	Known Ransomware
*.good	Scatter
*.exx	Alpha Crypt
*.encrypt	Alpha
*.encrypted	Various ransomware
*.1txt	Enigma
*.ezz	Alpha Crypt
*.r5a	7ev3n
*.wallet	Globe 3 (variant)
*.decrypt2017	Globe 3
*.zzzzz	Locky
*.MERRY	Merry X-Mas
*.enigma	Coverton
*.ecc	Cryptolocker or TeslaCrypt
*.cryptowall	Cryptowall
*.aesir	Locky
*.cryptolocker	CryptoLocker
*.coded	Anubis
*.sexy	PayDay
*.pubg	PUBG
*.ha3	El-Polocker
*.breaking_bad	Files1147@gmail(.)com
*.dharma	CrySiS
*.wcry	WannaCry
*.lol!	GPCode
*.damage	Damage
*.MRCR1	Merry X-Mas
*.fantom	Fantom
*.legion	Legion
*.kratos	KratosCrypt
*.crjoker	CryptoJoker
*.LeChiffre	LeChiffre
*.maya	HiddenTear (variant)
*.kraken	Rakhni
*.keybtc@inbox_com	KeyBTC
*.rrk	Radamant v2
*.zcrypt	ZCRYPT

Extension	Known Ransomware
*.crinf	DecryptorMax or CryptInfinite
*.enc	TorrentLocker / Cryptorium
*.surprise	Surprise
*.windows10	Shade
*.serp	Serpent (variant)
*.file0locked	Evil
*.ytbl	Troldesh (variant)
*.pdcr	PadCrypt
*.venusf	Venus Locker
*.dale	Chip
*.potato	Potato
*.lesli	CryptoMix
*.angelamerkel	Angela Merkel
*.PEGS1	Merry X-Mas
*.R16m01d05	Evil-JS (variant)
*.zzz	TeslaCrypt
*.wflx	WildFire
*.serpent	Serpent
*.Dexter	Troldesh (variant)
*.rnsmwr	Gremit
*.thor	Locky
*.nuclear55	Nuke
*.xyz	TeslaCrypt
*.encr	FileLocker
*.kernel_time	KeRanger OS X
*.darkness	Rakhni
*.evillock	Evil-JS (variant)
*.locklock	LockLock
*.rekt	HiddenTear (variant) / RektLocker
*.coverton	Coverton
*.VforVendetta	Samsam (variant)
*.remk	STOP
*.1cbu1	Princess Locker
*.purge	Globe
*.cry	CryLocker
*.zyklon	ZYKLON

Extension	Known Ransomware
*.dCrypt	DummyLocker
*.raid10	Globe [variant]
*.derp	Derp
*.zorro	Zorro
*.AngleWare	HiddenTear/MafiaWare (variant)
*.shit	Locky
*.btc	Jigsaw
*.atlas	Atlas
*.EnCiPhErEd	Xorist
*.xxx	TeslaCrypt 3.0
*.realfs0ciety@sigaint.org.fs0ciety	Fsociety
*.vbransom	VBRansom 7
*.exotic	Exotic
*.crypted	Nemucod
*.fucked	Manifestus
*.vvv	TeslaCrypt 3.0
*.padcrypt	PadCrypt
*.cryeye	DoubleLocker
*.hush	Jigsaw
*.RMCM1	Merry X-Mas
*.unavailable	Al-Namrood
*.paym	Jigsaw
*.stn	Satan
*.braincrypt	Braincrypt
*.ttt	TeslaCrypt 3.0
*._AiraCropEncrypted	AiraCrop
*.spora	Spora
*.alcatraz	Alcatraz Locker
*.reco	STOP/DJVU
*.crypte	Jigsaw (variant)
*.aaa	TeslaCrypt
*.pzdc	Scatter
*.RARE1	Merry X-Mas
*.ruby	Ruby
*.fun	Jigsaw
*.73i87A	Xorist

Extension	Known Ransomware
*.abc	TeslaCrypt
*.odcodc	ODCODC
*.crptrgr	CryptoRoger
*.herbst	Herbst
*.comrade	Comrade
*.szf	SZFLocker
*.pays	Jigsaw
*.antihacker2017	Xorist (variant)
*.rip	KillLocker
*.rdm	Radamant
*.CCCRRRPPP	Unlock92
*.bript	BadEncriptor
*.hnumkhotep	Globe 3
*.helpmeencedfiles	Samas/SamSam
*.BarRax	BarRax (HiddenTear variant)
*.magic	Magic
*.noproblemwedecfiles	Samas/SamSam
*.bitstak	Bitstak
*.kkk	Jigsaw
*.kyra	Globe
*.a5zfn	Alma Locker
*.powerfulldecrypt	Samas/SamSam
*.vindows	Vindows Locker
*.payms	Jigsaw
*.lovewindows	Globe (variant)
*.p5tkjw	Xorist
*.madebyadam	Roga
*.conficker	Conficker
*.SecureCrypted	Apocalypse
*.perl	Bart
*.paymts	Jigsaw
*.kernel_complete	KeRanger OS X
*.payrms	Jigsaw
*.paymst	Jigsaw
*.lcked	Jigsaw (variant)
*.covid19	Phishing

Extension	Known Ransomware
*.ifuckedyou	SerbRansom
*.d4nk	PyL33T
*.grt	Karmen HiddenTear (variant)
*.kostya	Kostya
*.gefickt	Jigsaw (variant)
*.covid-19	Phishing
*.kernel_pid	KeRanger OS X
*.wncry	Wana Decrypt0r 2.0
*.PoAr2w	Xorist
*.Whereisyourfiles	Samas/SamSam
*.edgel	EdgeLocker
*.adk	Angry Duck
*.oops	Marlboro
*.theworldisyours	Samas/SamSam
*.czvxce	Coverton
*.crab	GandCrab
*.paymrss	Jigsaw
*.kimcilware	KimcilWare
*.rmd	Zeta
*.dxxd	DXXD
*.razy	Razy
*.vxlock	vxLock
*.krab	GandCrab v4
*.rokku	Rokku
*.lock93	Lock93
*.pec	PEC 2017
*.mijnal	Minjal
*.kobos	Kobos
*.bbawasted	Bbawasted
*.rlhwasted	RLHWasted
*.52pojie	52Pojie
*.FastWind	Fastwind
*.spare	Spare
*.eduransom	Eduransom
*.RE78P	RE78P
*.pstKll	pstKll

Extension	Known Ransomware
*.erif	
*.kook	
*.xienvkdoc	
*.deadfiles	
*.mnbzr	
*.silvertor	
*.MH24	
*.nile	
*.ZaCaPa	
*.tcwwasted	
*.Spade	
*.pandemic	
*.covid	
*.xati	
*.Zyr	
*.spybuster	
*.ehre	
*.wannacry	WannaCry
*.jigsaaw	
*.boop	
*.Back	
*.CYRAT	
*.bmd	
*.Fappy	
*.Valley	
*.copa	
*.horse	
*.CryForMe	
*.easyransom	
*.nginxhole	
*.lockedv1	Lockedv1
*.ziggy	Ziggy
*.booa	Booa
*.nobu	Nobu
*.howareyou	Howareyou
*.FLAMINGO	Flamingo

Extension	Known Ransomware
*.FUSION	Fusion
*.pay2key	Pay2Key
*.zimba	Zimba, Dharma
*.luckyday	Luckyday
*.bondy	Bondy
*.cring	Cring
*.boom	Boom
*.judge	Judge
*.LIZARD	LIZARD
*.bonsoir	Bonsoir
*.moloch	Moloch
*.14x	14x
*.cnh	CNH
*.DeroHE	DeroHE

Enabling Ransomware Protection

Enable ransomware protection on your file server.

About this task

Procedure

1. Go to **dropdown menu** > **Ransomware**.
2. In the message banner, click **Enable Ransomware Protection**.
3. (optional) Click **Configure SMTP to add recipients**.

> **Note:** This option appears only if you have not configured a simple mail transfer protocol (SMTP) server, see Configuring an SMTP Server on page 27.

4. Under **Ransomware Email Recipients**, add at least one email address. If there is a ransomware attack, File Analytics sends a notification to the specified email address.

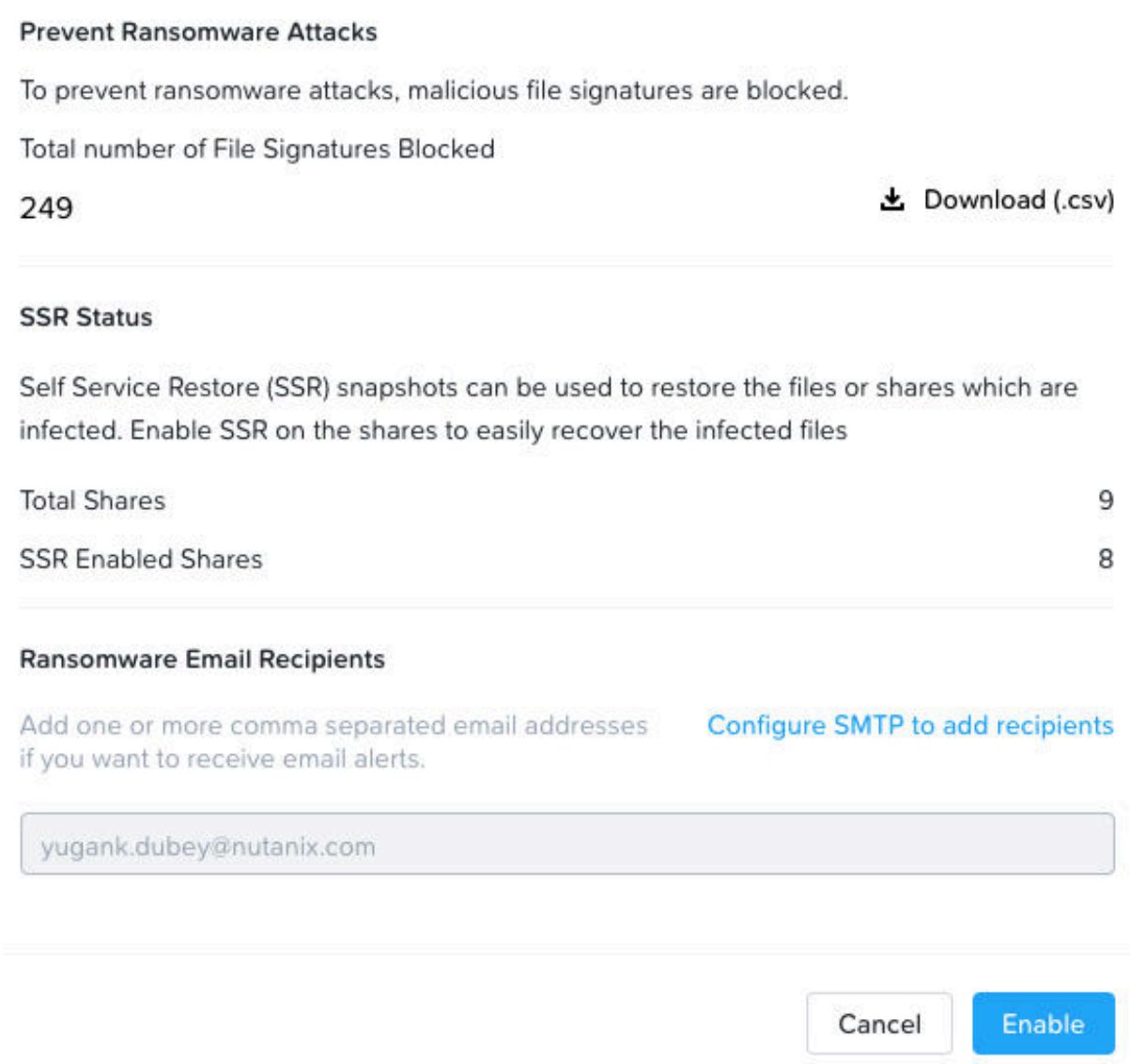


Figure 26: Enable Ransomware

5. Click **Enable**.

Configuring Ransomware Protection

Configure ransomware protection on file servers.

About this task

Do the following to add signature to the blocked extension list.

Procedure

1. Go to **dropdown menu** > **Ransomware** > > **Settings**.

2. (optional) Under **Search for blocked File Signatures**, enter ransomware signatures in the *.`(signature)` format.

   a. To check that the signature has been blocked, click **Search**.

   b. If the signature has not been blocked, click **Add to Block List**.

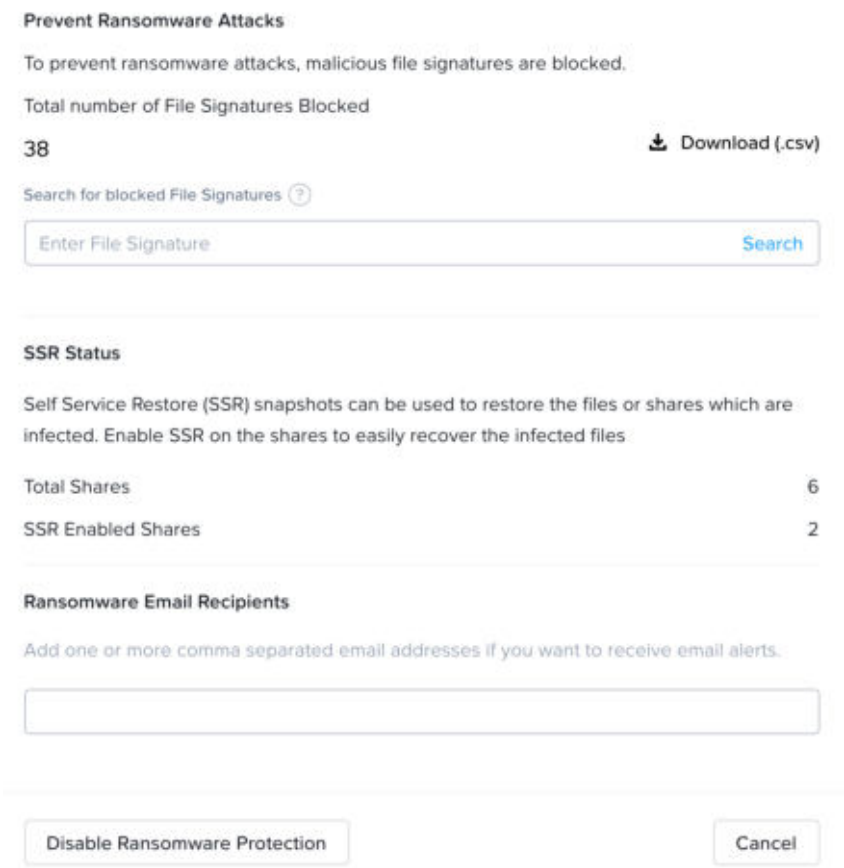


Figure 27:

3. (optional) click **Download (.csv)** to download a list of blocked ransomware signatures.

4. (optional) under **Ransomware Email Recipients**, add a comma separated list of email addresses. If there is a ransomware attack, File Analytics sends a notification to the specified email address.

5. (optional) click **Disable Ransomware Protection**, to disable the ransomware protection feature.

Enabling SSR

Enable self-service restore on shares identified by File Analytics.

About this task

File Analytics scans shares for SSR policies.

Procedure

1. Go to **dropdown menu** > **Ransomware**.

2. Click **Enable SSR on Prism**.
3. Check the box next to the shares for which to enable SSR.

Figure 28:

4. Click **Enable SSR**.

REPORTS

Generate a report for entities on the file server.

Create a report with custom attribute values or use one of the File Analytics pre-configured report templates. To create a custom report, you must specify the entity, attributes, operators for some attributes, attribute values, column headings, and the number of columns.

The reports page displays a table or previously generated reports. You can rerun existing reports rather than creating a template. After creating a report, download it as a JSON or CSV file.

Reports Dashboard

The reports dashboard includes options to create, view, and download reports.

The **Reports** dashboard includes options to create a report, download reports as a JSON, download reports as a CSV, rerun reports, and delete reports.

The reports table includes columns for the report name, status, last run, and actions.

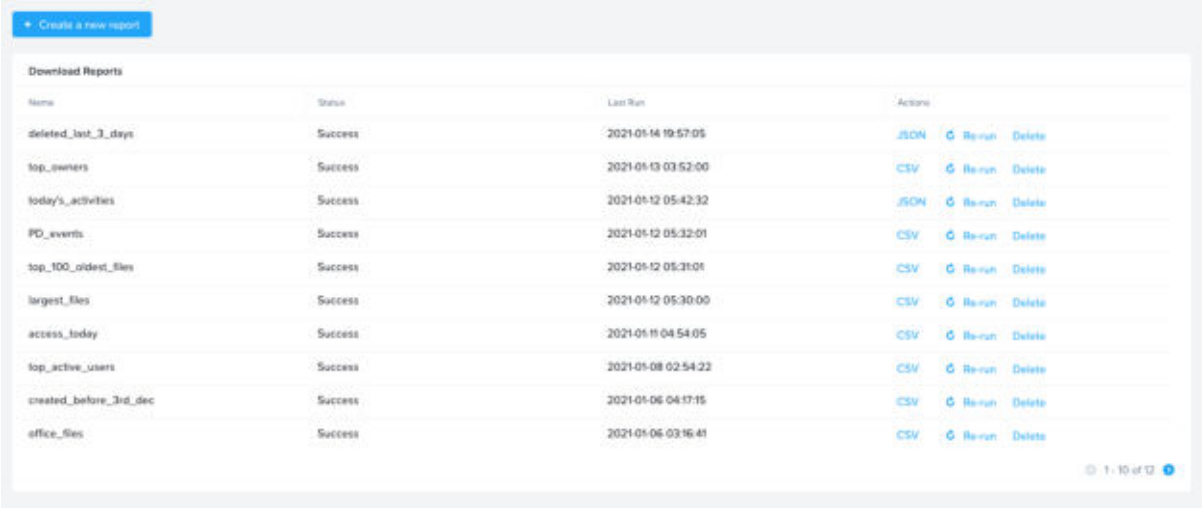


Figure 29: Reports Dashboard

Clicking **Create a new report** takes you to the report creation screen, which includes a **Report builder** and a **Pre-canned Reports Templates** tabs. The tabs include report options and filters for report configuration.

Both tabs include the following elements:

- The **Define Report Type** section includes an **Entity** drop-down menu to select an entity.
- The **Define Filters** section includes an **Attribute** drop-down menu and an option to add more attributes by clicking **+ Add filter**.
- The **Add/remove columns in this report** in your report section displays default columns. Clicking the columns field lets you add addition columns to the report. Clicking the **x** next to the column name removes it from the report.
- The **Define number of maximum rows in this report** section includes a **Count** section to specify the number of rows in the report.

Table 10: Report Builder – Filter Options

Entity	Attributes (filters)	Operator	Value	Column
Events	event_date	• equal_to • greater_than • greater_than_equal_to • less_than • less_than_equal_to	(date)	• audit_path (object path) • audit_objectname (object name) • audit_operation (operation) • audit_machine_name (source of operation) • audit_event_date (event date in UTC) • audit_username (user name)
	Event_operation	N/A	• file_write • file_read • file_create • file_delete • rename • directory_create • directory_delete • SecurityChange (permission change) • set_attr • sym_link	

Entity	Attributes (filters)	Operator	Value	Column
Files	Category	• equal_to • greater_than • greater_than_equal_to • less_than • less_than_equal_to	(date)	• object_name (file name) • share_UUID (share name) • object_owner_name (owner name) • object_size_logical (size) • file_type (extension) • object_creation_date (creation date in UTC) • last_event_date (access date in UTC) • share_UUID (share name) • fileserver_protocol • object_ID (file id) • object_last_operation_name (last operation) • audit_username (last operation user • object_last_operation_name (last operation) • file_path (file path)
	Extensions	N/A	(type in value)	
	Deleted	N/A	Last (number of days from 1 to 30) days	
	creation_date	• equal_to • greater_than • greater_than_equal_to • less_than • less_than_equal_to	(date)	
	access_date	• equal_to • greater_than • greater_than_equal_to • less_than • less_than_equal_to	(date)	
	Size	• equal_to • greater_than • greater_than_equal_to • less_than • less_than_equal_to	(number) (file size) File size options: • B • KB • MB • GB • TB	

Entity	Attributes (filters)	Operator	Value	Column
Folders	Deleted	N/A	Last (number of days from 1 to 30) days	• object_name (Dir name) • object_owner_name (owner name) • object_creation_date (creation date in UTC) • last_event_date (access date in UTC) • share_UUID (share name) • object_last_operation_name (last operation) • audit_username (last operation user) • File server protocol • object_ID (file id) • file_path (Dir path)
	creation_date	• equal_to • greater_than • greater_than_equal_to • less_than • less_than_equal_to	(date)	
Users	last_event_date	• equal_to • greater_than • greater_than_equal_to • less_than • less_than_equal_to	(date)	• user_login_name (user name) • Last operation • last_event_date (access date in UTC) • last_operation_audit_path

Table 11: Pre-Canned Reports – Filters

Entity	Pre-canned report template	Columns
Events	• PermissionDenied events • Permission Denied (file blocking) events	• audit_path (object path) • audit_objectname (object name) • audit_operation (operation) • audit_machine_name (source of operation) • audit_event_date (event date in UTC) • audit_username (user name)
Files	• Largest Files • Oldest Files • Files not accessed for last 1 year • Files accessed in last 30 days	• object_name (file name) • share_UUID (share name) • object_owner_name (owner name) • object_size_logical (size) • file_type (extension) • object_creation_date (creation date in UTC) • last_event_date (access date in UTC) • share_UUID (share name) • fileserver_protocol • object_ID (file id) • object_last_operation_name (last operation) • audit_username (last operation user • object_last_operation_name (last operation) • file_path (file path)
Users	• Top owners with space consumed • Top active users • All users	• user_login_name (user name) • Last operation • last_event_date (access date in UTC) • last_operation_audit_path

Creating a Custom Report

Create a custom report by defining the entity, attribute, filters, and columns.

About this task

Procedure

1. Go to **dropdown menu** > **Reports**.
2. Click **Create a new report**.
3. In the **Report Builder** tab, do the following.
   a. In the **Define Report Type** section, select an entity from the drop-down menu.
   b. In the **Define Filters** section, select an attribute from the attributes dropdown.
   c. Under **Value**, specify the values for the attribute (some attributes also require to specify an operator in the **Operator** field).
   d. (optional) click **+ Add filter** to add more attributes.
   e. In the **Add/Remove column in this report** section, click **x** for the columns you want to remove.
   f. In the **Define maximum number of rows in this report** section, type in , or use the **-**and **+** buttons, to specify the number of rows in your report. This value indicates the number of records in the report.
4. Click **Run Preview**.
   The **Report Preview** section populates.
5. Click **Generate report**.
   a. Select either the **CSV** or **JSON** option.

Create a Pre-Canned Report

Use one of the pre-canned File Analytics templates for your report.

Procedure

1. Go to **dropdown menu** > **Reports**.
2. Click **Create a new report**.
3. In the **Pre-Canned Reports Templates** tab, do the following.
   a. In the **Define Report Type** section, select an entity from the drop-down menu.
   b. In the **Define Filters** section, select an attribute from the attributes dropdown.
   c. In the **Add/Remove column in this report** section, click **x** for the columns you want to remove.
   d. In the **Define maximum number of rows in this report** section, type in , or use the **-**and **+** buttons, to specify the number of rows in your report. This value indicates the number of records in the report.
4. Click **Run Preview**.
   The **Report Preview** section populates.

5. Click **Generate report**.

   a. Select either the **CSV** or **JSON** option.

FILE ANALYTICS OPTIONS

You can get more insight into the usage and contents of files on your system by configuring and updating File Analytics features and settings. Some options include scanning the files on your file server on demand, updating data retention, and configuring data protection.

Updating Data Retention

The data retention period determines how long File Analytics retains event data.

About this task

Follow the steps as indicated to configure data retention.

Procedure

1. In File Analytics, click **gear icon** > **Update Data Retention**.
2. In the **Data Retention Period** drop-down, select the period for data retention.
3. Click **Update**.

Scanning the File System

Once enabled, File Analytics scans the metadata of all files and shares on the system. You can perform an on-demand scan of shares in your file system.

About this task

To scan shares, perform the following task:

Procedure

1. In File Analytics, click the **gear icon**.
2. In the drop-down list, click **Scan File System**.

Figure 30: Scan File System Option

3. In the list of shares, select the target shares for the scan.

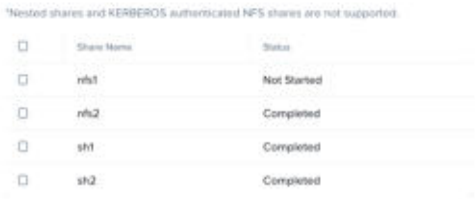


Figure 31: Select Scan Targets

4. Click **Scan**.
   The status of the share is **In Progress**. Once the scan is complete, the status changes to **Completed**.

Blacklisting

Blacklist users, file extensions, and client IPs.

About this task

Use the blacklisting feature to block audit events from being performed on specified file extensions or by specified users and clients.

Procedure

1. Click the **gear icon** > **Define Blacklisting Rules**.
2. Click the pencil icon in the user, file extension, or client IP row.
3. Add a comma separated list of entities that you want blocked.
4. Click save in the updated row.

Managing File Categories

File Analytics uses the file category configuration to classify file extensions.

About this task

The capacity widget in the dashboard uses the category configuration to calculate capacity details.

Procedure

1. Click **gear icon** > **Manage File Category**.
2. To create a category, click **+ New Category**. (Otherwise, move on to step 3).
   a. In the **Category** column, name the category.
   b. In the **Extensions** column, specify file extensions for the category.
3. To delete an existing category, click the **x icon** next to the category. (Otherwise, move on to step 4)
4. To modify an existing category, click the **pencil icon** next to the category and modify the specified file extensions.
5. Click **save**.

Data Protection

Configure File Analytics disaster recovery (DR) using Prism Element.

File Analytics only supports async disaster recovery. File Analytics does not support NearSync and metro availability.

Create an async protection domain, configure a protection domain schedule, and configure remote site mapping. The remote site must have symmetric configurations to the primary site. The remote site must also deploy File Analytics to restore a File Analytics VM (FAVM).

The *Data Protection* section in the *Prism Web Console Guide* provides more detail on the disaster recovery process.

Configuring Disaster Recovery

To set up disaster recovery for File Analytics, create an async protection domain, configure a protection domain schedule, and configure remote site mapping.

About this task

By default, the File Analytics volume group resides on the same container that hosts vDisks for Nutanix Files.

Procedure

1. If you have not done so already, configure a remote site for the local cluster.

   See the *Configuring a Remote Site (Physical Cluster)* topic in the *Prism Web Console Guide* for this procedure.

2. Create an async DR protection domain for the File Analytics volume group as the entity. The volume group name is `File_Analytics_VG`.

   See *Configuring a Protection Domain (Async DR)* in the **Prism Web Console Guide**.

3. In the **Schedule** tab, click the **New Schedule** button to add a schedule.

   Add a schedule, as File Analytics does not provide a default schedule. See *Creating a Protection Domain Schedule (Files) Nutanix Files Guide.*

4. Configure local and remote container mapping.

   See the *Configuring Disaster Recovery (Files)* section in the *Nutanix Files Guide* for steps to configure mapping between local and remote containers.

5. Create a protection domain schedule.

   See *Creating a Protection Domain Schedule (Files)* in the *Nutanix Files Guide*.

Activating Disaster Recovery

Recover a File Analytics VM (FAVM) after a planned or unplanned migration to the remote site.

About this task

Perform the following tasks on the remote site.

Procedure

1. Fail over to the protection domain for disaster recovery activation.

   See the *Failing Over a Protection Domain* topic in the *Prism Web Console Guide*.

2. Fail back the protection domain to the primary site.

   See the *Failing Back a Protection Domain* topic in the *Prism Web Console Guide*.

Deploying File Analytics on a Remote Site (AHV)

Deploy a File Analytics VM (FAVM) after a planned or unplanned (disaster) migration to the remote site.

About this task

To perform disaster recovery, deploy and enable File Analytics on the remote site. Restore the data using a snapshot of the volume group from the primary FAVM.

Before you begin

Ensure that the Nutanix Files and AOS versions match the versions on the remote and primary sites.

About this task

Run the following commands from the command prompt inside the FAVM.

Procedure

1. Deploy a new File Analytics instance on the remote site, see Deploying File Analytics on page 7.

   > **CAUTION:** Do not enable File Analytics.

   The remote site requires an iSCSI data service IP address to configure the FAVM on the remote site. This procedure deploys a new volume group File_Analytics_VG and deletes in a subsequent step.

2. On the remote site, create a volume group by restoring the snapshot of the `File_Analytics_VG`.

   See *Restoring an Entity from a Protection Domain* in *Data Protection and Recovery with Prism Element*. For the **How to Restore** step, use the **Create new entities** option, and specify a name in the **Volume Group Name Prefix** field. The restored volume group name format is *`prefix`*-File_Analytics_VG.

3. To configure the FAVM on the remote, follow these steps:

> **CAUTION:** If the IP address of the File Analytics VM has changed on the remote site, contact Nutanix Support before proceeding.

a. Log on to the FAVM with SSH.

> **Tip:** See KB 1661 for default credential details.

b. To discover all storage devices accessed by the FAVM, run the following commands.

```
nutanix@favm$  sudo blkid
```

c. Copy the `cvm.config` file to the temporary files directory.

```
nutanix@favm$ cd /mnt/containers/config/common_config /tmp
```

d. Stop the File Analytics services.

```
nutanix@favm$  sudo systemctl stop monitoring
```

```
nutanix@favm$  docker stop $(docker ps -q)
```

```
nutanix@favm$  sudo systemctl stop docker
```

e. Unmount the volume group.

```
nutnix@avm$ sudo umount /mnt
```

f. Detach the volume group File_Analytics_VG from the FAVM.
See the "Managing a VM (AHV)" topic in the *Prism Web Console Guide*.

g. Attach the cloned volume group *`prefix`*-File_Analytics_VG to the FAVM.
See "Managing a VM (AHV)" in the *Prism Web Console Guide*.

h. Restart the AVM to discover the attached volume group.

```
nutanix@avm$ sudo reboot
```

i. Log on to the FAVM with SSH.

> **Tip:** See KB 1661 for default credential details.

j. Discover all storage devices accessed by the FAVM.

```
nutanix@favm$  sudo blkid
```

```
/dev/sr0: UUID="2019-06-11-12-18-52-00" LABEL="cidata" TYPE="iso9660"
/dev/sda1: LABEL="_master-x86_64-2" UUID="b1fb6e26-a782-4cf7-b5de-32941cc92722"
 TYPE="ext4"
```

The FAVM discovers the attached volume group and assigns to the `/dev/sdb` device.

k. Delete the deployed volume group File_Analytics_VG.
See the "Modifying a Volume Group" topic in the *Prism Web Console Guide*.

l. Rename the restored volume group *`prefix`*-File_Analytics_VG to File_Analytics_VG.
See the "Modifying a Volume Group" topic in the *Prism Web Console Guide*.

m. Create a backup of the `cvm.config` file.

```
nutanix@favm$ mv /mnt/containers/config/common_config/cvm.config \
```

```
/mnt/containers/config/common_config/cvm_bck.config
```

n. Copy the `cvm.config` file from the `/tmp` directory to `/common_config/` on the FAVM.

```
nutanix@favm$ mv /tmp/cvm.config /mnt/containers/config/common_config/
```

o. Reconfigure the password of the user on Prism for internal FAVM operations. Specify a passphrase for *`new password`*. File Analytics uses the password only for internal communication between Prism and the FAVM. You must issue the same command twice.

```
nutanix@favm$ sudo python /opt/nutanix/analytics/bin/reset_password.py --user_type=prism
 \
 --password='new password' --local_update
```

```
nutanix@favm$  sudo python /opt/nutanix/analytics/bin/reset_password.py --
user_type=prism \
--password='new password' --prism_user=admin --prism_password='Prism admin password'
```

p. In File Analytics, go to **gear icon** > **Scan File System** to check if a file system scan can be initiated.

> **Note:** If you receive errors, disable and re-enable File Analytics, see "Disabling File Analytics" and "Enabling File Analytics."

Deploying File Analytics on a Remote Site (ESXi)

Deploy a File Analytics VM (FAVM) after a planned or unplanned (disaster) migration to the remote site.

About this task

To perform disaster recovery, deploy and enable File Analytics on the remote site. Restore the data using a snapshot of the volume group from the primary FAVM.

Before you begin

Ensure that the Nutanix Files and AOS versions match the versions on the remote and primary sites.

About this task

Run the following commands from the command prompt inside the FAVM.

Procedure

1. Deploy a new File Analytics instance on the remote site, see Deploying File Analytics on page 7.

   > **CAUTION:** Do not enable File Analytics.

   The remote site requires an iSCSI data service IP address to configure the FAVM on the remote site. This procedure deploys a new volume group File_Analytics_VG and deletes in a subsequent step.

2. On the remote site, create a volume group by restoring the snapshot of the `File_Analytics_VG`.

   See *Restoring an Entity from a Protection Domain* in *Data Protection and Recovery with Prism Element*. For the **How to Restore** step, use the **Create new entities** option, and specify a name in the **Volume Group Name Prefix** field. The restored volume group name format is *`prefix`*-File_Analytics_VG.

3. In the **Storage Table** view, go to the **Volumes** tab.

   a. Copy the target IQN prefix from the **Volume Group Details** column.

   > **Tip:** Click the tooltip to see the entire IQN prefix.

4. To configure the FAVM on the remote, follow these steps:

> **CAUTION:** If the IP address of the File Analytics VM has changed on the remote site, contact Nutanix Support before proceeding.

a. Log on to the FAVM with SSH.

> **Tip:** See KB 1661 for default credential details.

b. To discover all storage devices accessed by the FAVM, run the following commands.

```
nutanix@favm$  sudo blkid
```

c. Copy the `cvm.config` file to the temporary files directory.

```
nutanix@favm$ cd /mnt/containers/config/common_config/ /tmp
```

d. Stop the File Analytics services.

```
nutanix@favm$  sudo systemctl stop monitoring
nutanix@favm$  docker stop $(docker ps -q)
nutanix@favm$  sudo systemctl stop docker
```

e. Unmount and log off from all iSCSI targets.

```
nutnix@avm$ sudo umount /mnt
nutnix@avm$ sudo /sbin/iscsiadm -m node -u
```

f. Remove the disconnected target records from the discoverydb mode of the FAVM.

```
nutanix@favm$  sudo /sbin/iscsiadm -m node –o delete
```

g. Discover all storage devices accessed by the FAVM.

```
nutanix@favm$  sudo blkid
/dev/sr0: UUID="2019-06-11-12-18-52-00" LABEL="cidata" TYPE="iso9660"
```

```
/dev/sda1: LABEL="_master-x86_64-2" UUID="b1fb6e26-a782-4cf7-b5de-32941cc92722"
 TYPE="ext4"
```

The output does not show the `/dev/sdb` device.

h. Get the File Analytics Linux client iSCSI initiator name.

```
nutanix@favm$  sudo cat /etc/iscsi/initiatorname.iscsi
```

The output displays the initiator name.

```
InitiatorName=iqn.1991-05.com.redhat:8ef967b5b8f
```

i. Copy the iSCSI initiator name.

j. Remove the iSCSI initiator name from the client whitelist of the deployed volume group File_Analytics_VG.

See the "Modifying a Volume Group" topic in the *Prism Web Console Guide*.

k. Whitelist the AVM client on the cloned volume group *`prefix`*-File_Analytics_VG using the iSCSI initiator name of the AVM client.

See the "Modifying a Volume Group" topic in the *Prism Web Console Guide*.

l. Let the Analytics initiator discover the cluster and its volume groups.

```
nutanix@favm$  sudo /sbin/iscsiadm --mode discovery --type sendtargets --portal
  data_services_IP_address:3260
```

Clicking the Nutanix cluster name in Prism displays cluster details including the data service IP address. The output displays the restored iSCSI target from step 2.

m. Connect to the volume target by specifying IQN prefix.

```
nutanix@favm$  sudo /sbin/iscsiadm --mode node --targetname iqn_name --
portal data_services_IP_address:3260,1 --login
```

n. Restart the FAVM to restart the iSCSI host adapters, which allows the discovery of the attached volume group.

```
nutanix@favm$  sudo reboot
```

o. Log on to the FAVM with SSH.

> **Tip:** See KB 1661 for default credential details.

p. Discover all storage devices accessed by the FAVM.

```
nutanix@favm$  sudo blkid
```

The FAVM discovers the attached iSCSI volume group and assigns to the `/dev/sdb` device.

```
/dev/sr0: UUID="2019-06-11-12-18-52-00" LABEL="cidata" TYPE="iso9660"
/dev/sda1: LABEL="_master-x86_64-2" UUID="b1fb6e26-a782-4cf7-b5de-32941cc92722"
 TYPE="ext4"
/dev/sdb: UUID="30749ab7-58e7-437e-9a09-5f6d9619e85b" TYPE="ext4"
```

q. Delete the deployed volume group File_Analytics_VG.

See the "Modifying a Volume Group" topic in the *Prism Web Console Guide*.

r. Rename the restored volume group *`prefix`*-File_Analytics_VG to File_Analytics_VG.

See the "Modifying a Volume Group" topic in the *Prism Web Console Guide*.

s. Create a backup of the `cvm.config` file.

```
nutanix@favm$ mv /mnt/containers/config/common_config/cvm.config \
```

```
/mnt/containers/config/common_config/cvm_bck.config
```

t. Copy the `cvm.config` file from the `/tmp` directory to `/common_config/` on the FAVM.

```
nutanix@favm$ mv /tmp/cvm.config /mnt/containers/config/common_config/
```

u. Reconfigure the password of the user on Prism for internal FAVM operations. Specify a passphrase for *`new password`*. File Analytics uses the password only for internal communication between Prism and the FAVM. You must issue the same command twice.

```
nutanix@favm$ sudo python /opt/nutanix/analytics/bin/reset_password.py --user_type=prism
 \
 --password='new password' --local_update
```

```
nutanix@favm$  sudo python /opt/nutanix/analytics/bin/reset_password.py --
user_type=prism \
--password='new password' --prism_user=admin --prism_password='Prism admin password'
```

v. In File Analytics, go to **gear icon** > **Scan File System** to check if a file system scan can be initiated.

> **Note:** If you receive errors, disable and re-enable File Analytics, see "Disabling File Analytics" and "Enabling File Analytics."

COPYRIGHT

Copyright 2022 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

Files 4.0

Nutanix Files User Guide

March 14, 2022

Contents

Introduction to Nutanix Files 5
- Files Deployment 7
- Port Requirements 7
- Prerequisites 7
- Upgrades 11
  - Upgrading Files 12

Files Overview 15
- File Server View in Prism 15
- Files Console 17
  - Dashboard View 18
  - Monitoring View 19
  - Shares View 25
  - Data Management View 29
  - Alerts & Events View 29
  - Tasks View 32
  - Configuration View 33
- File Analytics 35

File Server Management 36
- Creating a File Server 36
  - Deleting a File Server 52
- File Server Updates 53
  - Updating the Network Configuration 54
  - Scaling FSVMs 57
  - Updating Memory and vCPU Resources 58
  - Updating File Server Basics 59
- Logging Onto A File Server VM 60
  - Changing an FSVM Password 60
  - Setting Timezones 61
- Starting a Files Cluster 61
  - Stopping a Files Cluster 62

Share and Export Management 63
- Creating a Share (SMB) 64
- Creating an Export (NFS) 70
- Creating a Multi-Protocol Share or Export 78
  - Multi-Protocol Support for Files 85
- Modifying a Share or Export 87
  - Deleting a Share or Export 88
- Accessing Home Shares 89
  - Accessing User Home Shares (Advanced) 89
- Continuously Available Shares (SMB Only) 90
  - Enabling Continuous Availability 90
- Connected Shares 90
  - Connecting a Share 91

Disconnecting Shares....91
Nested Shares and Exports....92
Durable SMB File Handles....93
Enabling SMB Symlinks....94
Managing Limited Local Users (SMB Only)....94
Configuring Backup for Distributed Shares....95
Setting Directory-Level Quotas....96
Share Migration....96
Migrating Shares (SMB Only)....96
Re-attempt Share Migration (SMB Only)....98

Directory Service and Domain Management....99
Joining a Domain....99
Leaving a Domain....100
Updating Domain Name System (DNS) Entries....101
Disjoint Domains....104
Configuring Disjoint Domains....104
Updating Directory Services....105
Setting AD Machine Account Password Expiry....105
Authentication....106
Authorization....106

User Management....108
User Mapping....108
Configuring User Mapping....109
Managing Roles....114
Managing REST API Roles....116
Authorizing a REST API User....116
Quotas....117
Managing Quotas....120

Files Options....124
Cloning....124
Cloning a File Server....124
Encryption....126
Files Data Collection....126
Access-Based Enumeration (SMB only)....126
File Blocking....127
Blocking Files on a File Server....127
Antivirus (AV) Scanning (SMB Only)....128
Configuring Antivirus Scanning (SMB Only)....130
Antivirus Tab....131
Files REST APIs....133

Performance Optimization....134
Managing Performance Optimization....134
Unblocking Rebalancing....137
Workload Optimization....137
Modifying the Workload Type....137
File System Compression....138

Data Management **139**
Data Protection and Recovery 139
Configuring Disaster Recovery 140
Activating Disaster Recovery 143
High Availability 145
Smart Tiering 145
Self-Service Restore 145
Enabling Self-Service Restore 146
Adding Snapshot Schedules 147
Retrieving Files (SMB Only) 149
Retrieving Files (NFS Only) 149
Deleting SSR Snapshots 149
Setting Custom Snapshot Times 150

Security Hardening **151**

Troubleshooting **156**
Invalid Mounts After Authentication Change 156
Client Access Denial (NFS Protocol) 156
Clients Cannot Mount Shares 156
Client Side Network Mapping 156
Connecting to Authentication Services 156
Constraint Violation 157
DNS Missing SRV Records 157
Domain Controller Issues 157
Finding IP Addresses 157
Identifying the Share Owner 157
Invalid Credential 158
NLM Locks 158
Network Cannot Expand 158
NTLM Authentication Issues 159
Share Copying 159
Stale Statistics 159
Time Difference 159
Unsuccessful Authentication 160

Copyright **161**

INTRODUCTION TO NUTANIX FILES

Nutanix Files (Files) is a software-defined, scale-out file storage solution that lets you share files in a centralized and protected location to eliminate the requirement of a third-party file server.

Files uses a scale-out architecture that provides file services to clients through the Server Message Block (SMB) and Network File System (NFS) protocols. Files combines one or more file server VMs (FSVMs) into a logical file server instance sometimes referred to as a Files cluster. You can create multiple file servers within a single Nutanix cluster.

Files creates a volume group (VG) for every FSVM to provide stable storage for persistent states and audit events. During a service outage, the states, storage, and events of a VG fail-over to another FSVM. Files also creates a dedicated container for every file server instance. If you choose to delete a file server, you can delete the container in Prism the fact.

> **Tip:** Solutions Documentation offers tech notes that include performance best practices, sizing recommendations, migration guidance, and an in-depth technical overview of Files architecture.

Files offerings also include File Analytics, for statistics and monitoring of file servers, and the Files Manager for a unified control plane of all file servers. For more information on these products, see the *File Analytics Guide* and the *Files Manager Guide*.

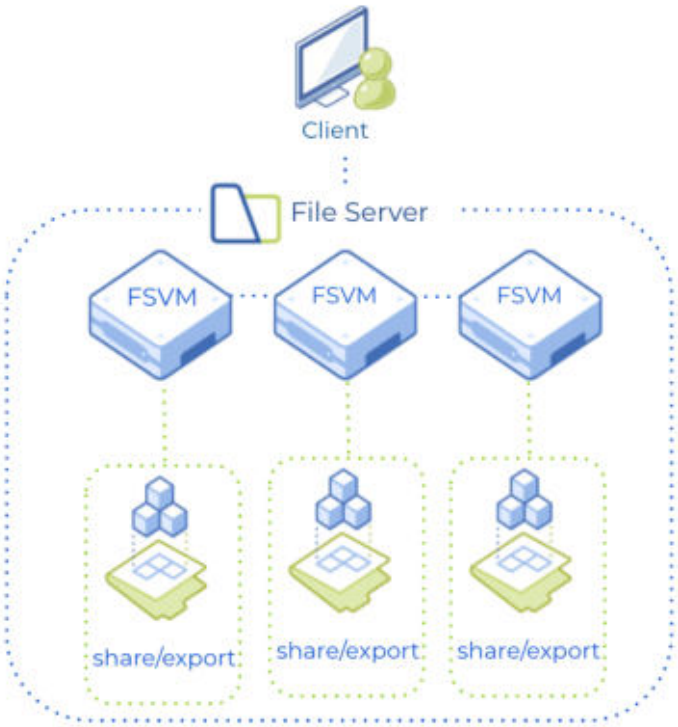


Figure 1: File Server Components

File Shares and Exports

Shares (SMB) and exports (NFS) encapsulate file directories. There are two types of shares or exports:

Distributed: A distributed share ("home") or export ("sharded") spreads data across all of the FSVMs on the file server to improve performance and scalability of client connections.

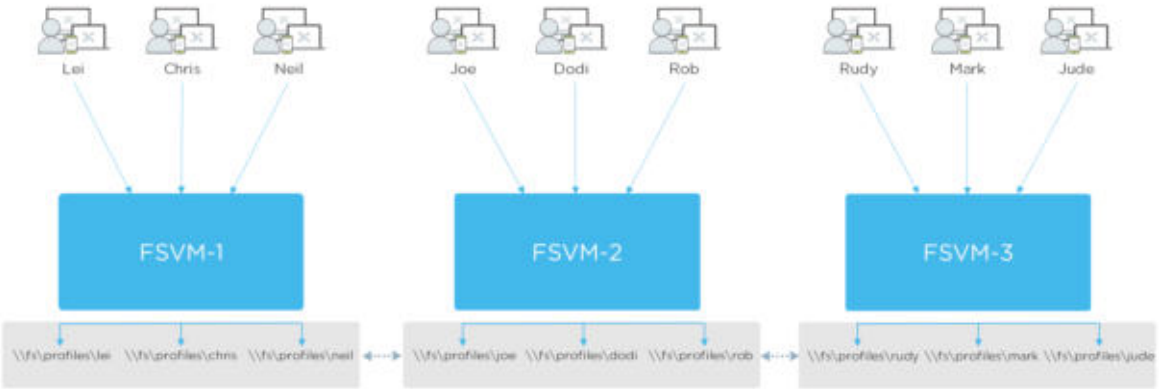


Figure 2: Distributed Share or Export

Standard: A standard share ("general purpose") or export ("non-sharded" or "non-distributed") contains all of the data on a single FSVM. A standard share or export serves data and connections from a single FSVM.

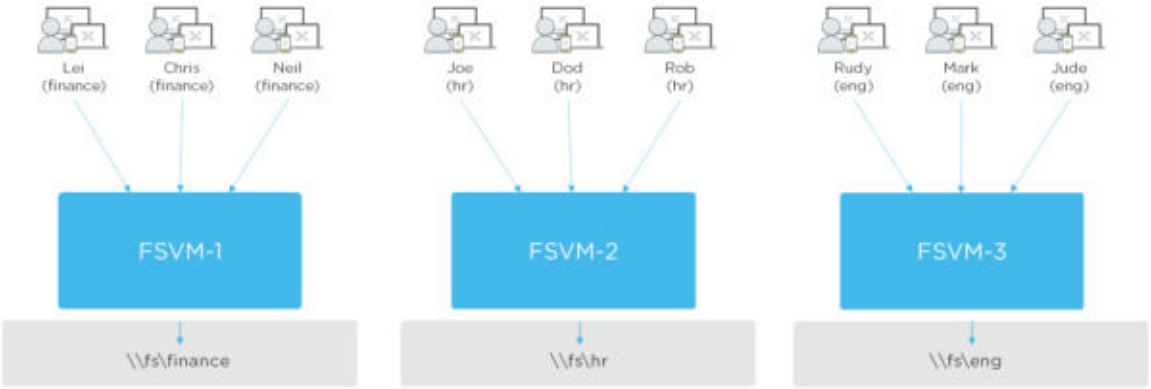


Figure 3: Standard Share or Export

Features

Files includes the following salient features:

- SMBv2, SMBv3, NFSv3 and NFSv4 protocol support.
- Multi-protocol support, see Multi-Protocol Support for Files on page 85.
- AHV and ESXi hypervisor support.
- High Availability for both VMs and data, see High Availability on page 145.
- Load balancing through scale-up and scaleout, see Performance Optimization on page 134.
- Data management including tiering and disaster recovery. See Data Protection and Recovery on page 139 and the *Files Manager Guide* for share-level data replication with SmartDR.

You can also use Files storage for the deployment of Kubernetes clusters on Karbon using the CSI Volume Driver. Refer to Karbon and CSI Volume Driver documentation for details.

For a description of features added with every major Files release, refer to the *Nutanix Files Release Notes*.

Networking

Files uses storage and client networks.

- **Storage network**: The storage network enables communication between the FSVMs and the Controller VMs.
- **Client network**: The client network enables communication between the clients and the FSVMs, allowing clients to access the Files shares. Files also uses the client network to communicate with the directory services.

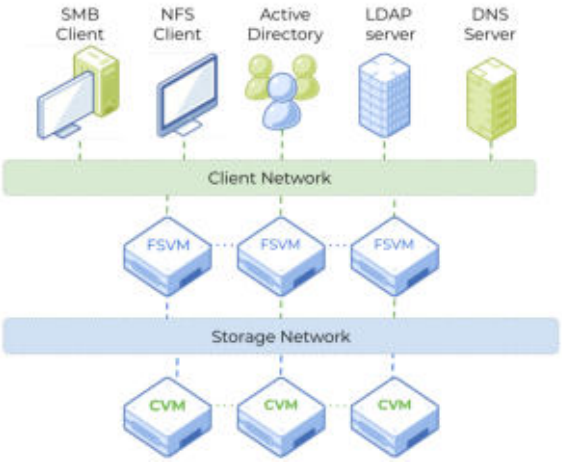


Figure 4: Files Networking

Files Deployment

Files deployment overview and requirements.

To deploy Files in a Nutanix cluster, do the following:

1. Satisfy the prerequisites and port requirements, see Prerequisites on page 7 and Port Requirements on page 7.
2. Install the Files software, see Upgrading Files on page 12.
3. Create a file server instance (Files cluster), see Creating a File Server on page 36.
4. Create one or more file shares (SMB) or exports (NFS), see Share and Export Management on page 63.

Port Requirements

Files has various firewall requirements depending on the protocols and services being used.

The Port Reference provides detailed port information for Nutanix products and services, including port sources and destinations, service descriptions, directionality, and protocol requirements.

Prerequisites

Review this section carefully to ensure you have satisfied the prerequisites before attempting to deploy Files.

Requirements

Do the following before deploying Files.

- Install the File Server Module (FSM) from the **Downloads** page on the **Support Portal**, or upgrade to a compatible version using the Life-Cyle Manager (LCM). See Upgrades on page 11.
- Configure and define the storage network.

- Configure and define the client network.
- Have at least one network (two networks recommended).
- Set up a network time protocol (NTP) server.
- If you use Active Directory for user authentication, have credentials of the domain administrator or a user with delegated permissions.
- If using SMB shares, enable the distributed file system (DFS) for Windows clients (on by default).
- If you plan to use LDAP for NFS with permissions required for search, have credentials of the bind distinguished name (DN).
- [ESXi clusters only] register all ESXi host nodes in the AOS cluster to the same vCenter.
- Have an assigned iSCSI Data Services IP configured for the clusters.

Limitations

File servers require the following minimum configurations.

- A minimum of four vCPUs per host.
- A minimum of 12 GiB of memory per host.
- For each file server, the number of CVMs must equal to or be greater than the number of file server VMs (FSVMs) to ensure availability if there is a node failure.

> **Note:** Refer to *Files Release Notes* for release-specific details on supported configurations and software compatibility.

Network Requirements

The storage network requires at least one more IP address than the number of FSVMs. The client network requires the same number of IP addresses as the number of FSVM nodes.

- Storage network: Number of FSVMs + 1 (available IP addresses)
- Client network: Number of FSVMs

Single-FSVM deployments require one IP address for the storage network and one IP address for the client network.

If the client and storage networks are separate, they must be on different subnets. If you use the same network for both client and storage, then IP addresses must be unique. Clients on the same subnet as the storage network cannot access the shares or exports.

Required Information

Collect the following information before deploying Files.

Table 1: Network Time Protocol

NTP Server	Used for the time synchronization between the file server and AD service.

Table 2: Domain Name System

DNS server names	Files uses DNS to resolve FSVM names and access external services.

Table 3: Active Directory (Optional)

Active Directory (AD)	Windows AD domain name.
AD admin account	Admin with domain administrator or delegated permissions.

Table 4: LDAP (Optional)

LDAP URI	LDAP server name or IP address (with optional port number).
Base DN	Distinguished name of the entry where to start the search for records.
Bind DN	Distinguished name of entry to use to perform search (optional when anonymous bind is disallowed).
Bind password	Password to use for bind DN.

Table 5: iSCSI Data Services

iSCSI data services IP address	Files uses iSCSI to connect the storage to the FSVMs. See the *Nutanix Volumes Guide* for more information about iSCSI storage.

Table 6: Managed Networks

Storage network	The VLAN that connects the Controller VM to the FSVM.
Client network	The VLAN that connects the FSVM to the AD and DNS.

Table 7: Unmanaged Networks

Storage network gateway	The VLAN that connects the Controller VM to the FSVM.
Storage network subnet	The storage network subnet value.
Storage network IP address range	If there is more than one FSVM, the number of IP addresses is the total number of FSVMs (one FSVM per node) plus one more address. For example, in a three-node cluster you will need four IP addresses, and in a four-node cluster you will need five. Single-FSVM deployments require only one storage network IP address.

Client network gateway	VLAN connects the FSVM to the AD and DNS.
Client network subnet	Subnet of client network.
Client network IP address range	One IP address for each FSVM in the file server. For example, a three FSVM file server needs three IP addresses.

Table 8: Active Directory Parameters

Provide the following parameters to join the domain.

Element	Definition	Required
Domain Name	The fully qualified domain name.	Required
Organizational unit	**Note:** By default, Files creates the computer account in the Computers container. The forward slash (/) is not allowed in the organizational unit. The organizational unit (OU) contains the computer account Files creates. List the AD users permissions for the machine account on the OU. In an organization with complex hierarchies, create the computer account in a specified container by using a forward slash mark to denote hierarchies (for example, organizational_unit/inner_organizational_unit).	Optional
Password	The password for the account used to connect to the AD server. Files uses the password to authenticate to AD and to create the Files computer account.	Required

Element	Definition	Required
Preferred domain controller	**Note:** • Files does not support read-only domain controllers (RODC) for joining domains because RODC cannot create machine accounts. • If the preferred domain controller is not reachable, the enabling AD operation fails. Files discovers a local domain controller for all communications. If you do not configure Files to a specific site, then it uses a domain-level domain controller. This parameter lets you specify a preferred domain controller Files uses for the join domain operation. The preferred domain controller must be reachable and allow write access privileges to the domain controller.	Optional
User name	The user name that authenticates on the Active Directory on the domain. A domain user authenticates to the file server on the domain controller and creates Files computer accounts, related SPN entries, and Files DNS entries (when using Microsoft DNS). If you are in the same domain, you can use the user principal name (UPN) or SamAccountName. If you are outside of the domain but in the same forest, then use the UPN. **Note:** • Files only allows the "at" symbol (@) symbol to specify the domain. Files does not allow UPNs with multiple @ symbols. For example, "user@mydomain.com" is valid but "user@name@mydomain.com" is not valid. • Files does not allow the forward slash (/) and the backward slash (\) in the UPN.	Required

Upgrades

Upgrade Files and the required components using the Life Cycle Manager.

Nutanix recommends performing Files upgrades during non-business hours. Upgrading Files can cause a brief FSVM downtime and temporary disconnection of connected clients.

Starting with AOS 5.17 and Life Cycle Manager (LCM) 2.3.1.1, perform Files upgrades through LCM. For earlier versions, see Upgrading Files on page 12.

Upgrading Files requires upgrading to a compatible AOS and File Server Module (FSM) version in LCM. The FSM manages the Files life-cycle and encompasses the Files GUI components. Files relies on AOS for some control plane components.

> **Tip:** The *Nutanix Files Release Notes* provide details on updates in each Files and FSM version.

To check the current version of Files or the FSM, and to upgrade to later versions, perform the inventory check in LCM. For steps on performing inventory and upgrades in LCM, refer to the *Life Cycle Manager Guide*.

Upgrading Files

Upgrade the Files software from the Nutanix portal for use.

About this task

Follow the steps as indicated to upgrade Files through Prism. For upgrades with AOS 5.17 or later and Life Cycle Manager (LCM) 2.3.1.1 and later, perform inventory and upgrades in LCM (refer to the *Life Cycle Manager Guide*).

> **Note:**
>
> - [ESXi only] When performing one-click hypervisor upgrades that have Files, disable the anti-affinity rules on all FSVMs. After the hypervisor successfully upgrades, enable the anti-affinity rules on the FSVMs.
> - ESXi hosts that belong to multiple vSphere clusters or are deployed across multiple datacenters might experience limitations. See KB 5369 for more information.

Procedure

1. Log into Prism Element (PE) with your credentials.
2. Click the **gear icon** > **Upgrade Software**.
3. In the **Settings** menu, click **Upgrade Software**.
4. In the **Upgrade Software** window, select the **File Server** tab.

5. Complete the upgrade process.

   » Download the Files version you want to upgrade to (step 6).

   » Upload the Files binary from a different source (continue to step 7).

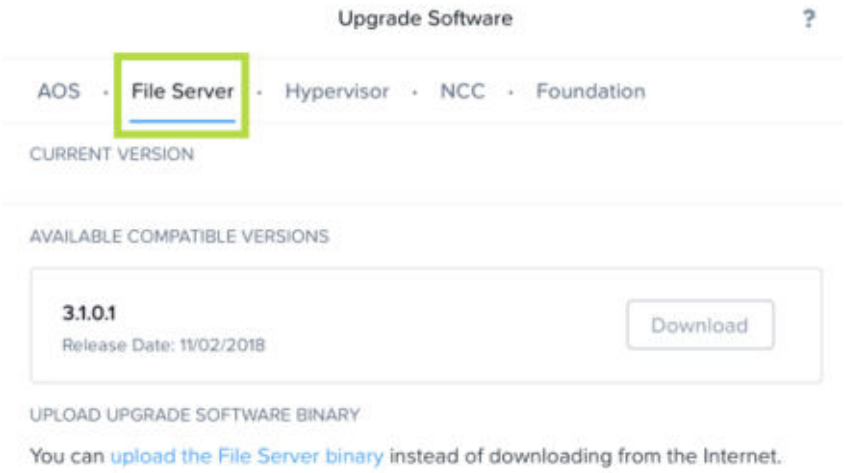


Figure 5: Upgrade Software: File Server tab

6. To download software, click the **Download** button for the target upgrade version.

   a. Once the download is complete, click **Continue**.

   b. In the **New File Server: Pre-check** dialog box, review the requirements and best practices.

   c. Click **Continue.**

   d. In **Upgrade Software** window, click the **Upgrade** button.

   Existing file servers upgrade to the selected Files version. This upgrade takes a few minutes. Once the upgrade is complete, a message indicates the number of file servers successfully upgraded.

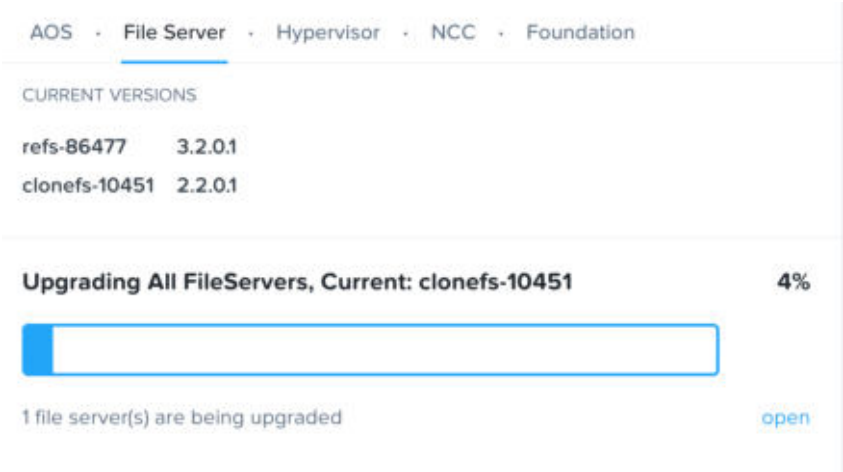


Figure 6: File Server: Upgrading

7. To upgrade using uploaded software, click the **upload the File Server binary** link, and perform the following steps as indicated.

   a. **File server metadata file**: click the **Choose File** button and select the target Files metadata file.

Figure 7: Upload Software Binary

   b. **File server binary file**: click the **Choose File** button and select the target Files binary file.

   c. To upload the upgrade files, click the **Upload Now** button.

      Existing file servers upgrade to the selected Files version. This upgrade takes a few minutes. Once the upgrade is complete, a message indicates the number of file servers successfully upgraded.

FILES OVERVIEW

Manage file servers and shares from the Files Console and the **File Server** view in Prism Element (PE).

The **Files Server** view in Prism Element is the landing page for Files. The **File Server** view provides an overview of your file servers on a PE cluster and basic file server management options.

The Files Console is a GUI for comprehensive management of file servers and shares. To access the Files Console from **File Server** view in PE, click **Launch Files Console** next to the target file server.

File Server View in Prism

The **File Server** view in Prism Element provides basic information about each file server in the cluster.

To get to the **File Server** view, select **File Server** from the pull-down list in Prism Element.

File Server View Layout

The **File Server** view includes the following sections and options:

- Action buttons to create a file server (see Creating a File Server on page 36), to configure the network (see Updating the Network Configuration on page 54), and to deploy File Analytics (see File Analytics).
- An entities table displays information about each file server. You can filter the table contents by entering a string in the search field located above the table.

  > **Note:** The **Recommendations** column is disabled. See **Recommendations** in the Files Console dashboard.

- A **File Summary** pane displays high-level details about file servers on the cluster, and, after selecting a file server, the **File Server Details** pane displays summary information for that file server.
- A **gear icon** with options to download the table content in the CSV or JSON format.
- A list of operations to perform on the file server that includes the following: **Launch Files Console** (see Files Console on page 17, **Clone** (see Cloning on page 124), **Update** (see File Server Updates on page 53), **Protect** (see Data Management on page 139), and **Delete** (see Deleting a File Server on page 52). You can perform some of these operations through the Files Console.

  > **Note:** The values for the parameters do not account for features applied by AOS or space used at the storage container level.

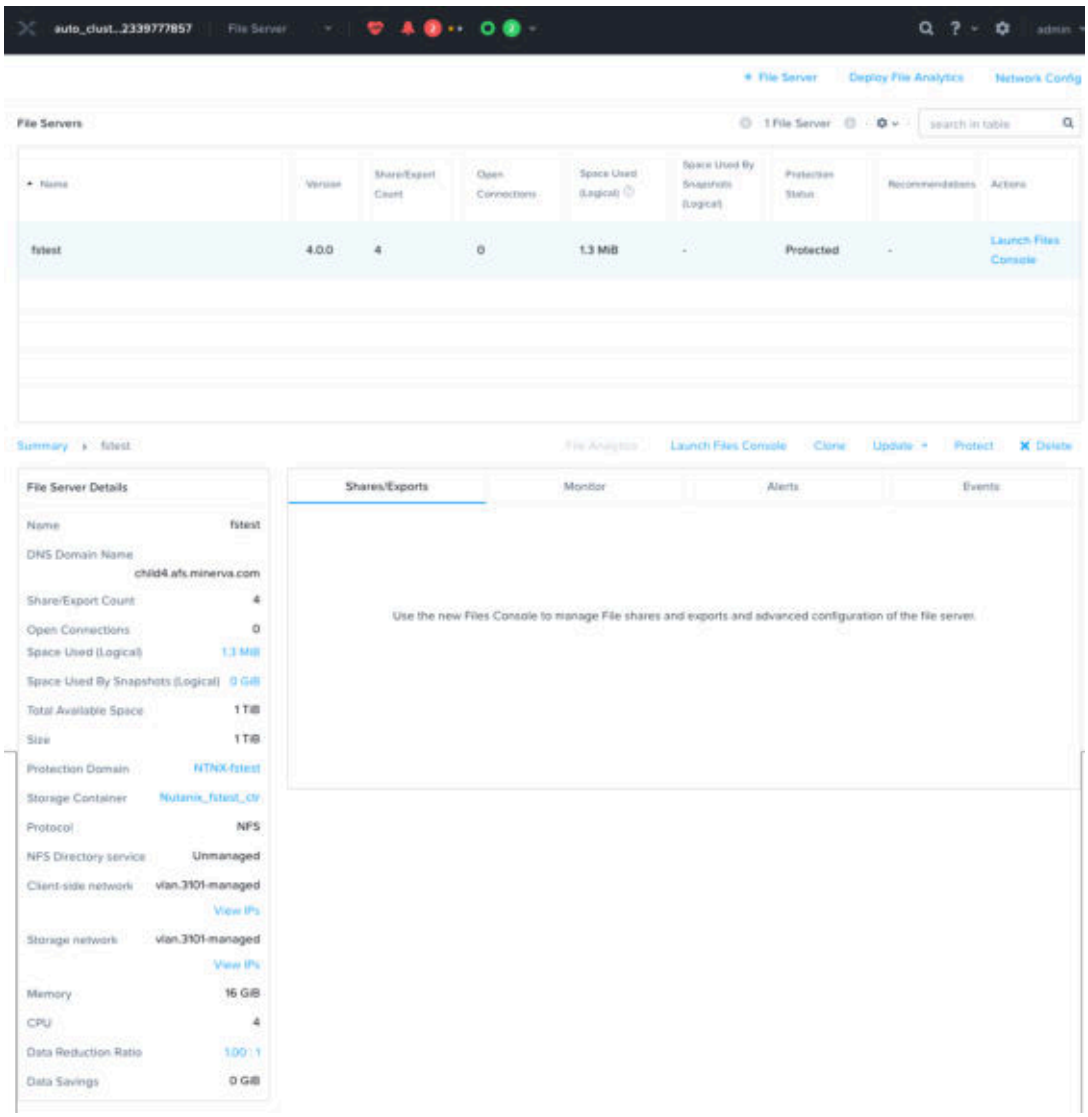


Figure 8: File Server View

Details Pane

Selecting a file server in the table presents detailed information in the **File Server Details** pane. The following tables describe the fields.

Table 9: File Server Details Fields

The parameters described in this table represent values from the perspective of the file server.

Parameter	Description	Values
Name	The file server name.	(name)
DNS domain name	The name of the domain that the file server is registered to. "Not Protected" indicates that the file server is not currently in a protection domain.	(DNS domain name)
Open connections	The number of open connections.	(integer)

Parameter	Description	Values
Share/Export Count	The total number of shares or exports.	(integer)
Space used	The total amount of storage space used within the file server currently.	xx [GiB|TiB]
Space used by snapshots	The amount of space used within the file server to store snapshots currently.	xx [MiB|GiB|TiB]
Total available space	The amount of available (unused) storage space currently on the file server.	xx [MiB|GiB|TiB]
Size	The size of the file system of the file server.	xx [TiB]
Protection domain	The name of the protection domain that includes this file server. Clicking the name displays the **Data Protection** view for that protection domain. "Not Protected" indicates that the file server is not in a protection domain. See the *Data Protection and Recovery with Prism Element* guide for information on protection domains.	(protection domain name)
Storage container	The name of the storage container of the file server. Clicking the name displays the **Storage Container** view for that storage container.	(storage container name)
Protocol	The protocols used by the file server.	SMB, NFS, or both
SMB directory service	The SMB protocol always uses Active Directory as the directory service.	Active Directory
NFS directory service	The NFS protocol has multiple options for the directory service.	Unmanaged, Active Directory, or LDAP
Client-side network	The name of the network used by clients.	(network name)
Storage network	The name of the network used for storage.	(network name)
Memory	Used memory.	GiB
CPU	Total CPUs.	Numerical
Data reduction ratio	Data reduced using file-system-level compression.	x:x (numerical)
Data savings	Amount of data saved using file-system-level compression.	MiB

Alerts Tab

The **Alerts** tab displays a table of alerts for the selected file server. You can also see alert details in the Files Console, see Alerts & Events View on page 29.

Events Tab

The **Events** tab displays a table of events for the selected file server. You can also see events details in the Files Console, see Alerts & Events View on page 29.

Files Console

The Files Console provides administrative tools and dynamically updated information for a single file server and its shares.

Access the Files Console from the **File Server** view in Prism Element (PE).

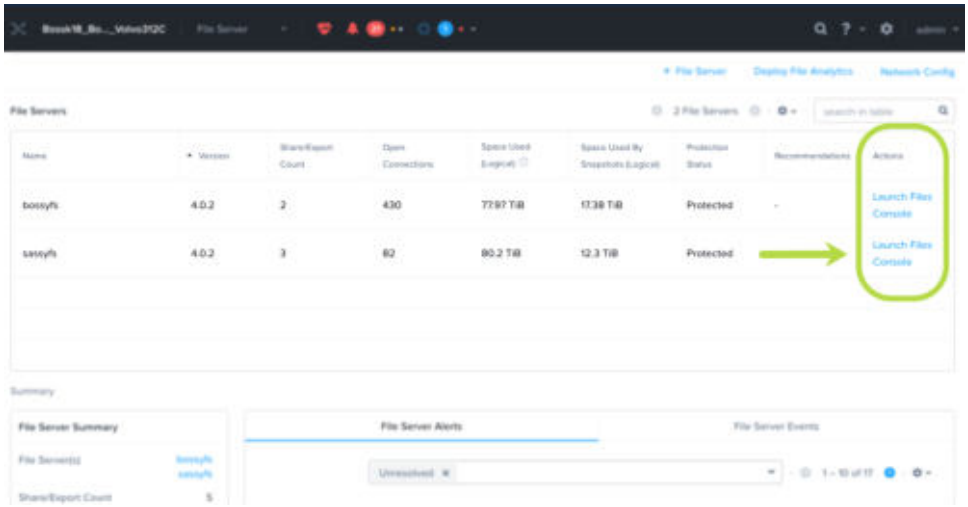


Figure 9: Launch the Files Console

The Files Console consists of the following primary tabs:

- The **Dashboard** tab is the home page in the files console provides an overview of file server dataDashboard View on page 18.
- The **Shares** tab provides detailed information on every share on the file server, see Share Details View on page 27.
- The **Data Management** tab provides options for configuring disaster recovery, self-service restore, and Smart Tiering, se Data Management on page 139.
- The **Alerts & Events** tab provides details of file server events and alerts on the file server with an option to acknowledge each occurrence, see Alerts & Events View on page 29.
- The **Tasks** view displays a list of recent tasks and the current status of each task, s Tasks View on page 32.
- The **Configuration** tab includes configuration options for the file server and a **Platform** view that provides a configuration summary, see Configuration View on page 33.

Dashboard View

This **Dashboard** view is the landing page in the Files Console.

Dashboard

The **Dashboard** tab includes the following elements.

- A **Capacity Summary** pane that visualizes the data usage on the file server.
- A **File Server Health** pane indicates the health status of the file server.
- A **Performance Summary** pane that consists of a graph that displays current throughput, current total IOPS, and current latency data.
- A **Data Lens** pane indicates whether you have or have not enabled Data Lens on the file server, see the *Nutanix Data Lens User Guide* for more details.
- A top **Top Shares** pane includes a drop-down option to sort top-shares by storage used, connections, and files.

- A **Features** pane that lists the features enabled on the file server.
- A **Recommendations** pane lists recommendations for improving the file server performance.

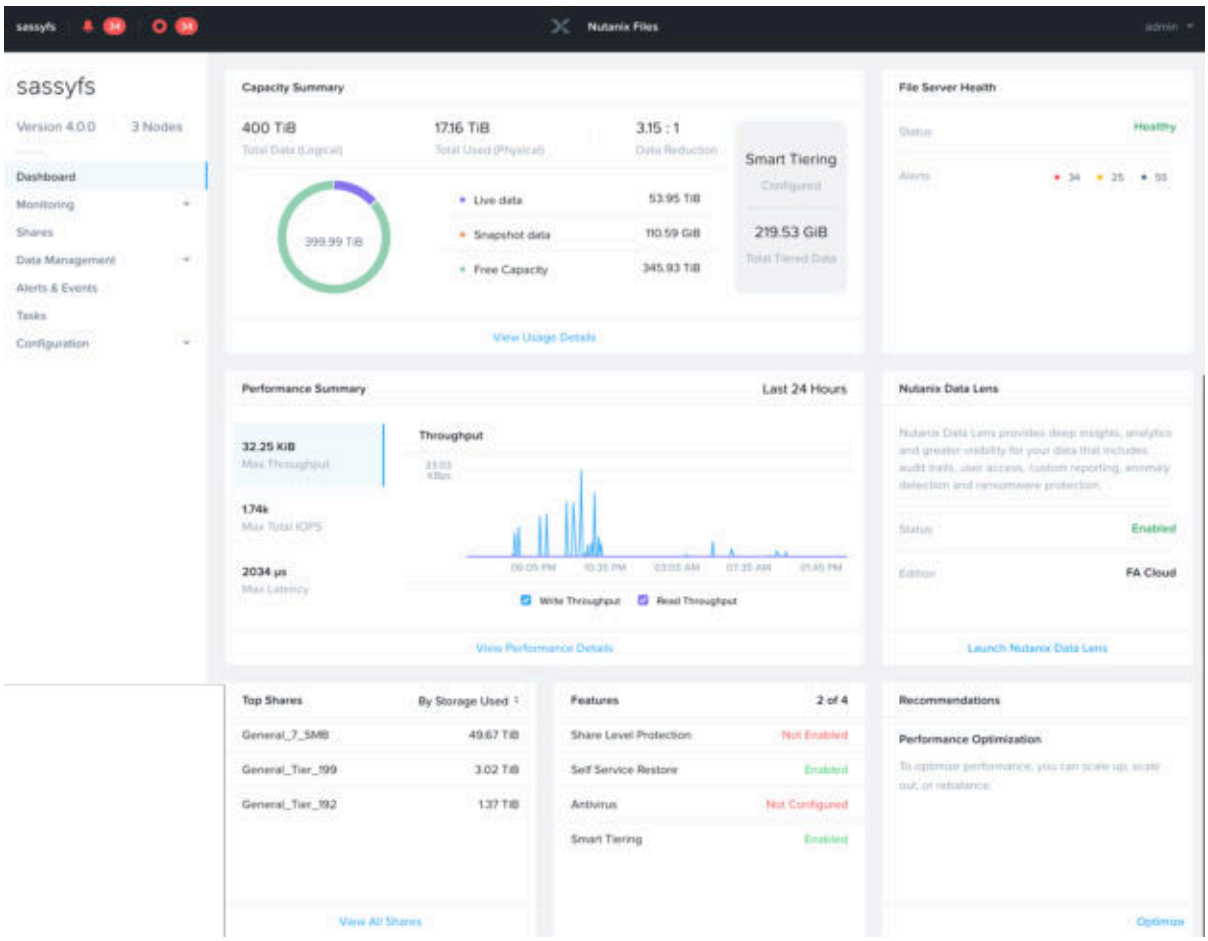


Figure 10: Dashboard View

Monitoring View

The **Monitoring** tab includes subtabs with granular monitoring details.

Usage Tab

The **Usage** tab displays these graphs.

- The **Storage Used** graph displays a rolling time interval monitor of the storage space used for data and snapshots on the file server. Hovering over the data displays the value for the time specified on the horizontal axis. To isolate a data set, check or uncheck the **Spaced Used by Dataset** and **Space Used For Snapshots** boxes.
- The **Open Connections** graph displays a rolling time interval monitor of the number of open connections on the file server.
- The **Number of Files** graph displays a rolling time interval monitor of the total number of files in the file server. Hovering over the data displays the value for the time specified on the horizontal axis.
- The **Top Shares by Current Capacity** graph indicates the top shares using the most storage capacity.

- The **Top Shares by Current Connections** graph indicates the top shares with the most current open connections.

Pull down lists above the graphs let you sort the data. You can select the time interval (last week, last 24 hours, last 6 hours, or last 3 hours). You can also select to display data for all shares or for specific shares.

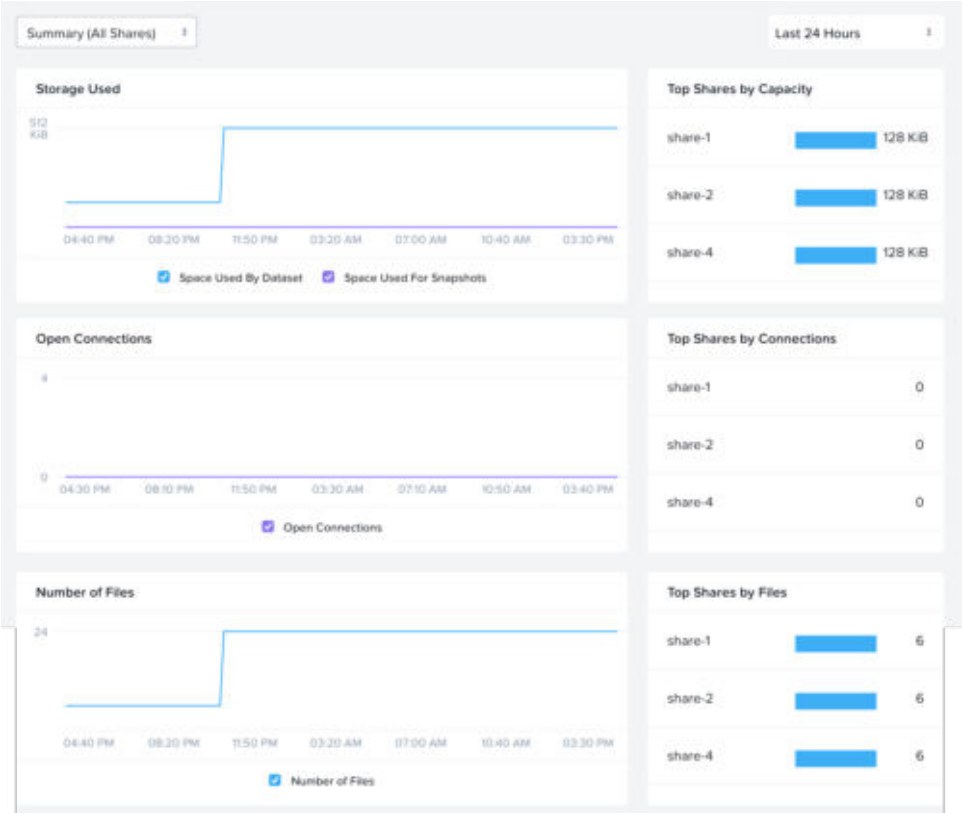


Figure 11: File Server Usage Tab

Performance Tab

The **Performance** tab includes the following elements.

- The **Latency** graph displays average latency across a rolling time interval monitor. Hovering over the data displays the value for the time specified on the horizontal axis. Selecting the **Show I/O and Metadata Breakdown** option above the graphs adds **Write Latency**, **Read Latency,** and **Metadata Latency** data set options.
- The **Throughput** graph displays average throughput. Hovering over the data displays the value for the time specified on the horizontal axis. Selecting the **Show I/O and Metadata Breakdown** option above the graphs adds **Write Throughput** and **Read Throughput** data options.
- The **IOPS** graph displays total I/O operations per second. Hovering over the data displays the value for the time specified on the horizontal axis. Selecting the **Show I/O and Metadata Breakdown** option above the graphs adds **Write IOPS**, **Read IOPS**, and **Metadata IOPS** data options.
- The **Top Shares by Current Latency**pane displays the shares with the most latency.
- The **Top Shares by Current Throughput** pane displays the shares with the most throughput.
- The **Top Shares by Current IOPS** pane displays the shares with the most I/O operations per second.

Pull-down lists above the graphs let you sort the data. You can select the time interval (last week, last 24 hours, last 6 hours, or last 3 hours), and you can select to display data for all shares or for specific shares.

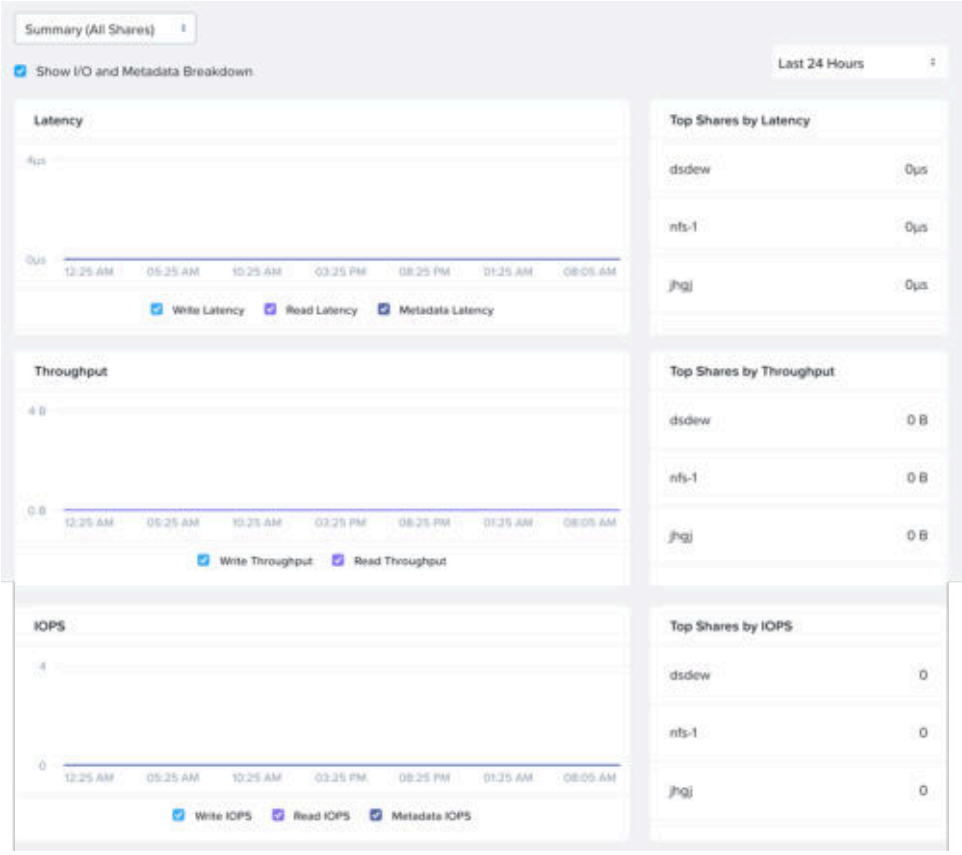


Figure 12: File Server Performance Tab

Antivirus Tab

> **Note:** This tab is only visible with SMB or multi-protocol shares.

The **Antivirus** tab displays antivirus scanning information (see Antivirus (AV) Scanning (SMB Only) on page 128) in a set of subtabs:

The **ICAP Servers** tab displays a pane that lists the configured ICAP servers and panes with the following details:

- The **All ICAP Servers** pull-down list lets you choose to scan data for all ICAP servers or for a specific server based on its IP address.
- The **ICAP Server** pane indicates the scanned server.
- The **Connection Status** pane indicates whether the server connected to Files.
- The **Number of Shares** pane indicates the number of shares and exports scanned.
- The **Average Latency** graph displays the average response latency.
- The **Files Scanned** and **Data Processed** switches reveal the **Files Scanned** or the **Data Processed** graphs. The **Files Scanned** graph displays the number of files scanned by the server. The **Data Processed** graph displays the amount of data processed by the server.

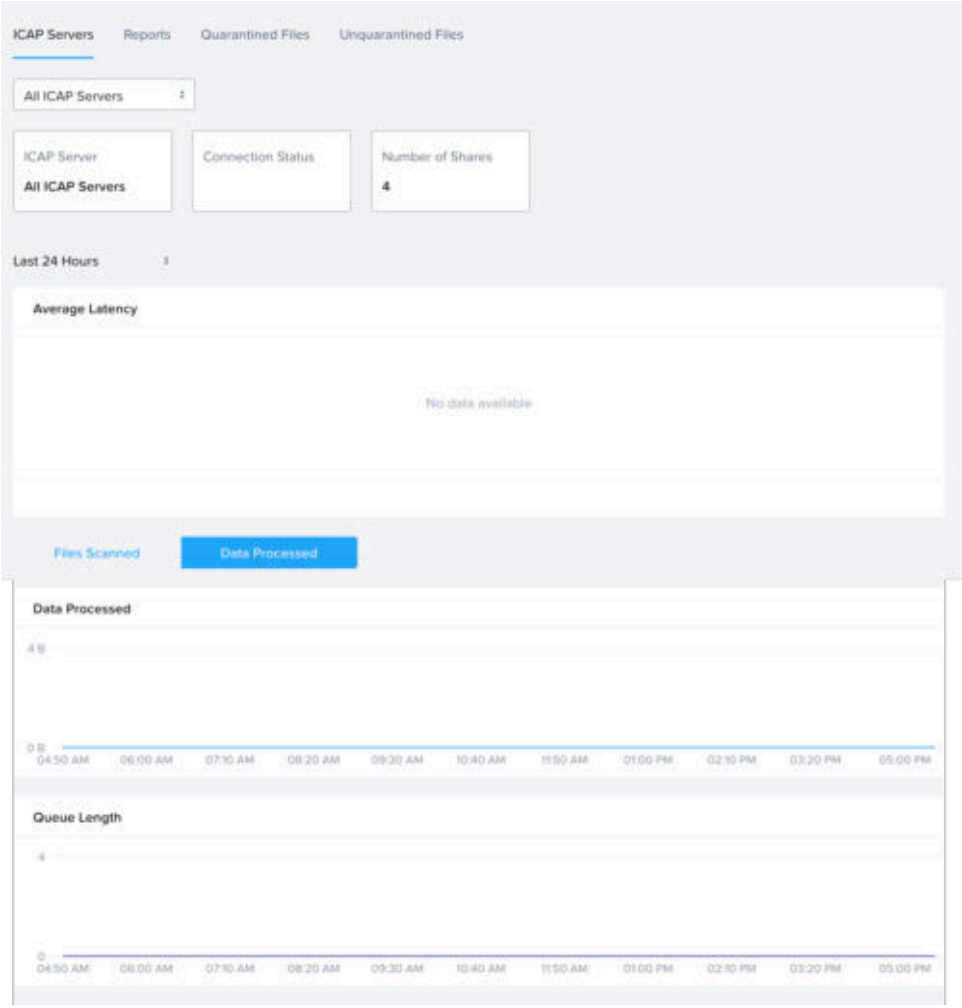


Figure 13: Antivirus Tab: ICAP Servers

The **Reports** tab provides two tables, one summarizing the latest scan and a second listing the identified threats. The following table describes the fields. The first table provides a scan summary, which includes the following fields:

- **Scan Period**: the antivirus scan period (24, six, or 3 hours).
- **Total Files Scanned**: the number of files scanned.
- **Threats Detected**: the number of threats detected. The **Events** table details each threat.
- **Files Cleaned**: the number of files cleaned.
- **Files Quarantined**: the number of files quarantined.
-

The second table provides event details, which include the following columns:

- **Share/Export**: the name of the share or export in which the affected file resides.
- **File Path**: the path to the affected file.
- **Threat Description**: describes the detected threat.

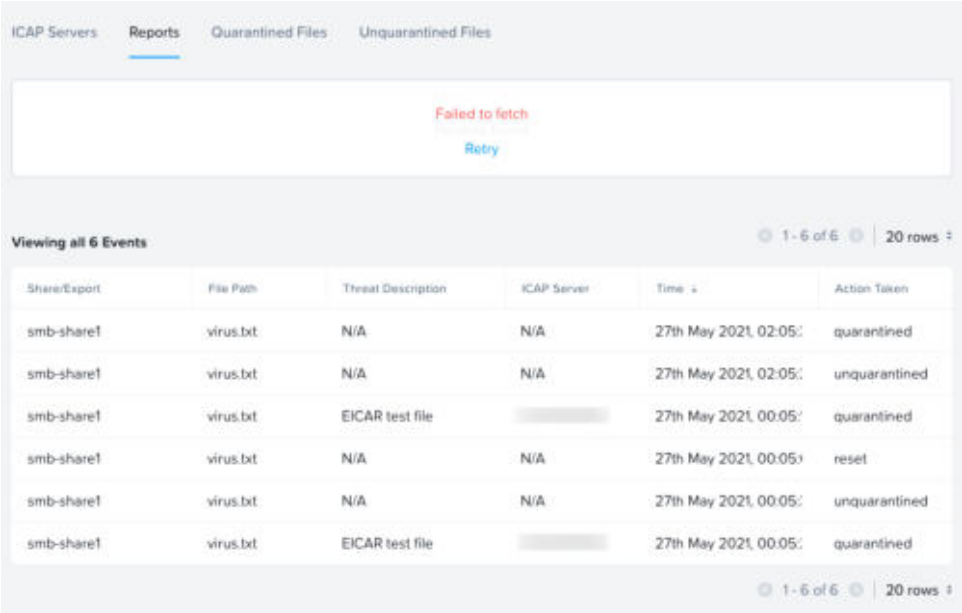
- **ICAP server**: the IP address of the ICAP server that detected the threat.
- **Time**: the time when the threat Files detected the threat.
- **Action Taken**: the action taken to address the threat (quarantined, unquarantined, reset).
- 

Figure 14: Antivirus Tab: Reports

Quarantined Files and **Unquarantined Files** tabs include tables that describe each of the quarantined or unquarantined files and an action pull-down menu.

Table 10: Quarantine Fields

Parameter	Description	Values
Share/Export name	The name of the share or export where the affected file resides.	(name)
File path	The path to the affected file.	(file path)
Threat description	Describes the detected threat.	(text string)
ICAP server	The name of the ICAP server that detected the threat.	(server name)

Parameter	Description	Values
Scan time	The time when the file was quarantined (unquarantined).	(time)

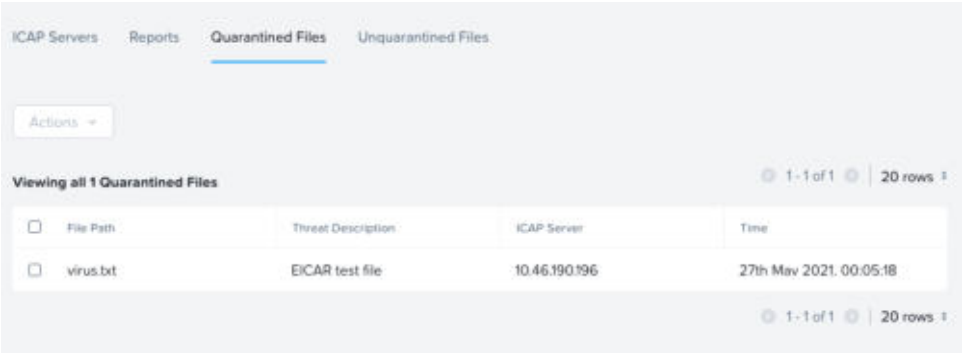


Figure 15: Antivirus Tab: Quarantined Files

File Server VMs

The **File Server VMs** tab displays the following graphs:

- The **Load Average** graph displays a rolling time interval monitor of the CPU usage on the file server as a percentage of total available CPU. Placing the cursor anywhere on the horizontal axis displays the value then.
- The **Memory Usage** graph displays a rolling time interval monitor of the memory usage on the file server in GB. Placing the cursor anywhere on the horizontal axis displays the value then.

You can select the time interval (last week, last 24 hours, last 6 hours, or last 3 hours). You can also select to display data for all shares or for specific shares.

Figure 16: File Server Usage Tab

Shares View

The **Shares** tab in the Files Console.

The **Shares** view provides a list of all shares on the file server. Clicking a share name in the share table goes to the share details views, which includes several more tabs: **Summary**, **Snapshots**, **Quota Policies**, **Antivirus**, and **Metrics**. (Continue to the next sections for more details on the additional tabs).

The **Shares** tab includes the following elements.

- The **Create a New Share** action button.
- A table with information on each share on the file server. By default, the table displays the **General** view; use the **View By** dropdown menu to switch the table to the **Metrics** view.

The **General** view consists of information described in the following table.

Table 11: Shares - General

Column	Description
Name	The name of the share or export. Clicking the share opens the share **Summary** in the share details view.
Share/Export path	The file path to the share or export.
Protocol type	The primary protocol of the share or export (NFS or SMB).
Share type	The data distribution type of the share. *Standard* shares and exports contain all data on a single FSVM. *Distributed* shares and exports load balance data across all FSVMs of the file server.
Share protection	The disaster recovery policy status on the share. See Data Protection and Recovery on page 139.
Self-service restore	The status of self-service restore for the share or export (enabled or disabled). See Self-Service Restore on page 145.
Compression	The status of file-system level compression: the green checkmark icon indicates enabled compression, and the gray **x** icon indicates disabled compression. See File System Compression on page 138

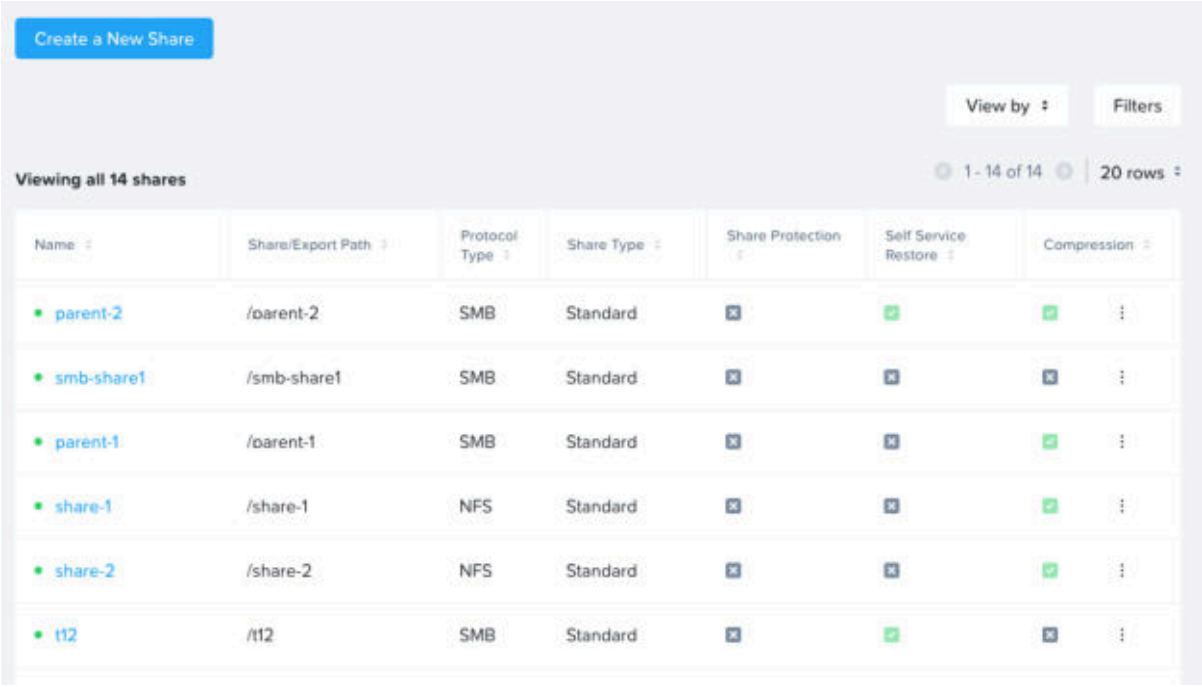


Figure 17: Shares- General

The **Metrics** view consists of the following columns and details:

Table 12: Shares- Metrics

Column	Description
Name	Name of the share.
Share/Export path	Path to the share or export.
Space used	The sum of space used by data, space used by snapshots (logical), and space used by file metadata.
Space used by snapshots	The space used by self-service restore snapshots.
Connections	The average number of open connections on the share.
IOPS	The average number of input and output operations per second.
Throughput	The average throughput.
Latency	The average latency.

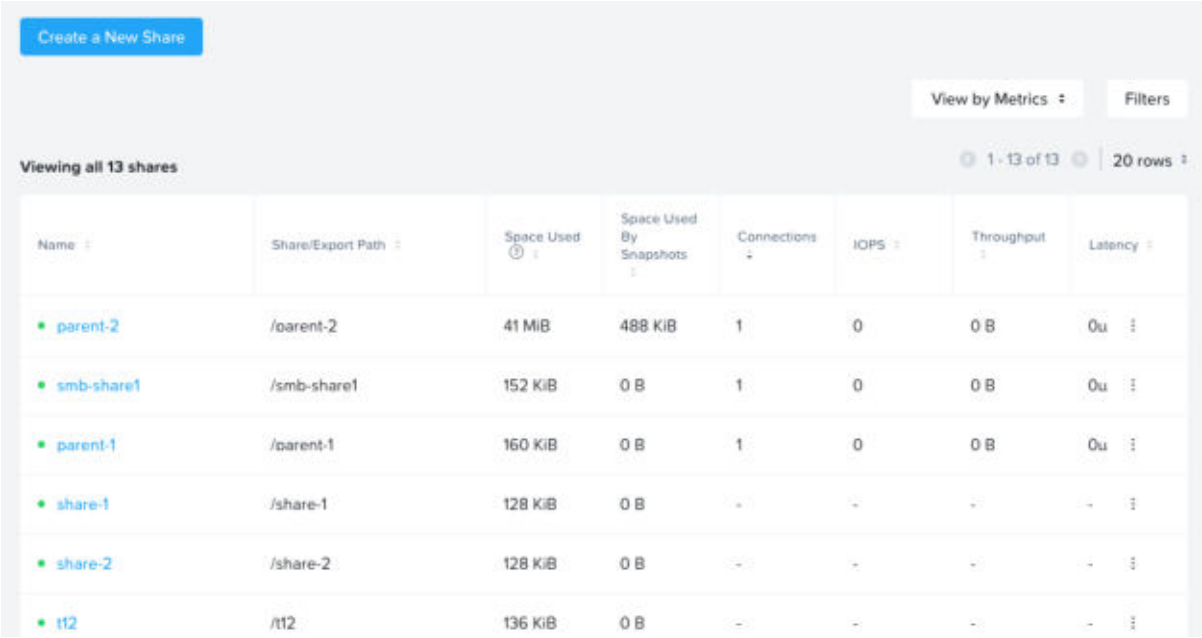


Figure 18: Shares - Metrics

Share Details View

Clicking the name of a share in the Files Console opens the share details view, which includes several tabs that detail share properties and operations.

The share details view consists of the **Summary**, **Snapshots**, **Quota Policies**, **Antivirus**, and **Metrics** tabs.

Summary

The **Summary** tab includes the following elements:

- An **Actions** dropdown menu includes options to update some of the share configurations.
- A **Capacity Summary** pane visualizes the share capacity used by snapshot and actual data.
- A **Share Properties** pane includes details on the configuration of the share (see the "Share Properties" table for more details).
- A **Performance Summary** pane consists of a graph that displays current throughput, current total IOPS, and current latency data.
- A **Features** pane lists the features enabled on the share.

Table 13: Share Properties

Parameter	Description	Values
Name	The share or export name.	(share/export name)
Description	The description of the share provided during share creation.	(description text)
Share path	The file path to the share or export.	(share path)
Mount path	The mount path to the share or export.	(mount path)

Parameter	Description	Values
Primary protocol	The primary protocol of the share.	[NFS | SMB]
Multi-protocol access	Indicates the multi-protocol access status on the share.	[enabled | disabled]
Share type	The data distribution type of the share. *Standard* shares and exports contain all data on a single FSVM. *Distributed* shares and exports load balance data across all FSVMs of the file server.	[standard | distributed]
Share compression	The status of file-system level compression, see File System Compression on page 138	[enabled | disabled]
Blocked file types	The status of file blocking, see File Blocking on page 127.	[enabled | disabled]
Message encryption	Indicates enabled SMB3 message encryption. See, Encryption on page 126.	[enabled | disabled]
File system compression	Indicates enabled file-system level compression. See File System Compression on page 138	[enabled | disabled]

Snapshots

The **Snapshots** tab includes a table that displays all snapshots of the share. The table includes column with the following details:

- The **Create Time** indicates the time that Files took the snapshot.
- The **Snapshot ID** indicates the unique identified for the snapshot.
- The **Total Space** indicates the size of the snapshot.
- The **Reclaimable Space** indicates the amount of space that you can recover by deleting the snapshot.

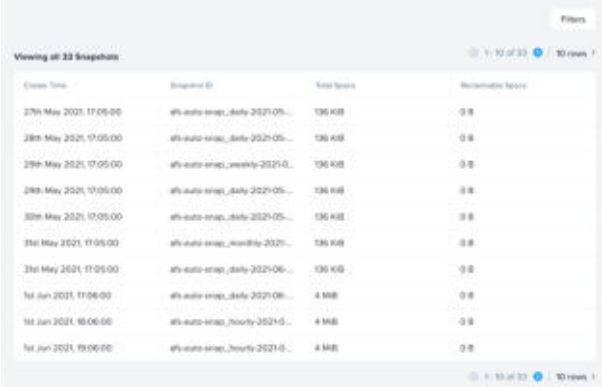


Figure 19: Snapshots

Quota Policies

The **Quota Policies** tab includes a **New Quota Policy** button and a table that displays all quota policies on the share.

Antivirus

The **Antivirus** tab includes the following subtabs: **Infected Files**, **Quarantined Files**, **Unquarantined Files**.

The **Infected Files** tab includes a list of infected files.

The **Quarantined Files** tab includes a list of infected files.

The **Unquarantined Files** tab includes a list of files that were previously quarantined. An administrator removed the file from quarantine. An unquarantined file typically indicates a false-positive.

Metrics

The **Metrics** tab includes two more tabs: **Usage** and **Performance**.

The **Usage** tab includes the following dynamically updated graphs:

- **Storage Usage**: Displays the amount of storage used over time.
- **Open Connections**: Displays the number of open connections over time.
- **Number of Files**: Displays the number of files on the share over time.

The **Performance** tab includes an option to **Show I/O and Metadata Breakdown** and the following dynamically updated graphs:

- **Latency**: Displays latency on the share over time.
- **Throughput**: Displays throughput on the share over time.
- **IOPS**: Displays the number of input and output operations per second over time.

Data Management View

Manage file server data in the **Data Management** view.

The **Data Management** view consists of the following subtabs.

- The **Protection** tab includes more tabs for **Disaster Recovery** and **Self-Service Restore**.
  - In the **Disaster Recovery** tab, configure Smart disaster recovery (DR) or protection-domain-based DR. See Data Protection and Recovery on page 139.
  - In the Self-Service Restore tab, configure snapshot schedules. See Self-Service Restore on page 145.
- The **Smart Tiering** tab, includes an option to configure Smart Tiering using Data Lens. See Smart Tiering on page 145.

Alerts & Events View

The **Alerts & Events** tab in the Files Console.

The **Alerts & Events** view consists of the **Alerts** and the **Events** tab.

Alerts Tab

The **Alerts** tab includes the following elements:

- A search bar to filter by alert name.
- A table displaying a list of recent alerts.

Table 14: File Server Alerts Fields

Parameter	Description	Values
(selection box)	To select the alert, click this box. Clicking the **Acknowledge** or **Resolve** buttons acknowledges or resolves all the selected alerts.	n/a
Title	Description of the alert.	(string)
Impact type	Displays the impact category.	[availability | capacity| configuration | performance]
Acknowledged	Displays the acknowledgment status.	[auto | yes | no | (unspecified)]
Resolved	Displays the resolution status.	[auto | yes | no | (unspecified)]
Source entity	Displays the entity name (**File Server**) to which this alert applies. Clicking the name displays the details for that file server.	(entity name)
Severity	Displays the severity level of this condition. There are three levels: **Critical** A "critical" alert is one that requires immediate attention, such as a failed Controller VM. **Warning** A "warning" alert is one that might need attention soon, such as an issue that could lead to a performance problem. **Informational** An "informational" alert highlights a condition to be aware of, for example, a reminder that the support tunnel is enabled.	Critical, warning, informational
Create time	Displays the date and time when the alert occurred.	(time and date)

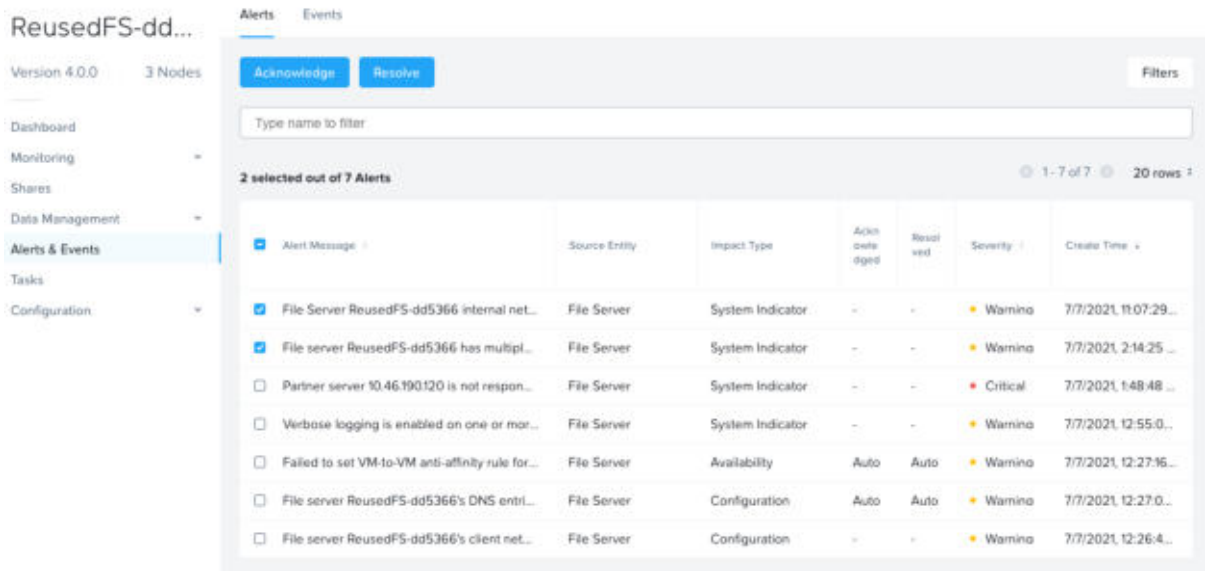


Figure 20: Alerts

The **Events** tab displays a table of events across all file servers. The following table describes the event table fields.

Table 15: File Server Events View Fields

Parameter	Description	Values
Title	Displays the event title and indicates related entities.	(message text)
Entities	Displays the type of entity (**File Server**, **Share**) to which the event applies. A comma-separated list appears if it applies to multiple entities. Clicking the entity name displays the details for that file server, share, or export.	[share, file server]
Event type	Displays the category for the event.	[storage, user action]
(create time)	Displays the date and time when the event occurred.	(time and date)

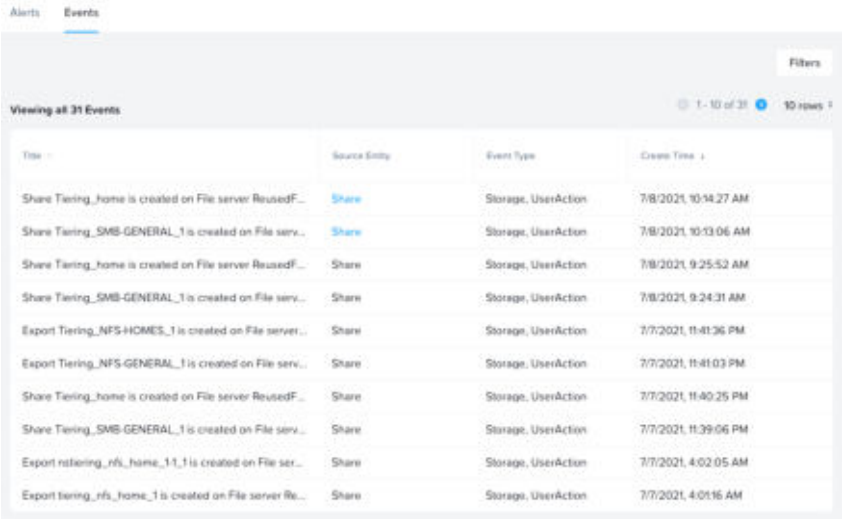


Figure 21: Events Tab

Tasks View

The **Tasks** view in the Files Console.

The **Tasks** view indicates the tasks running on the file server and includes the following elements:

- A **tasks table** that lists each administrative operation initiated on the file server.
- A **filters** menu to filter using pre-configured filters.

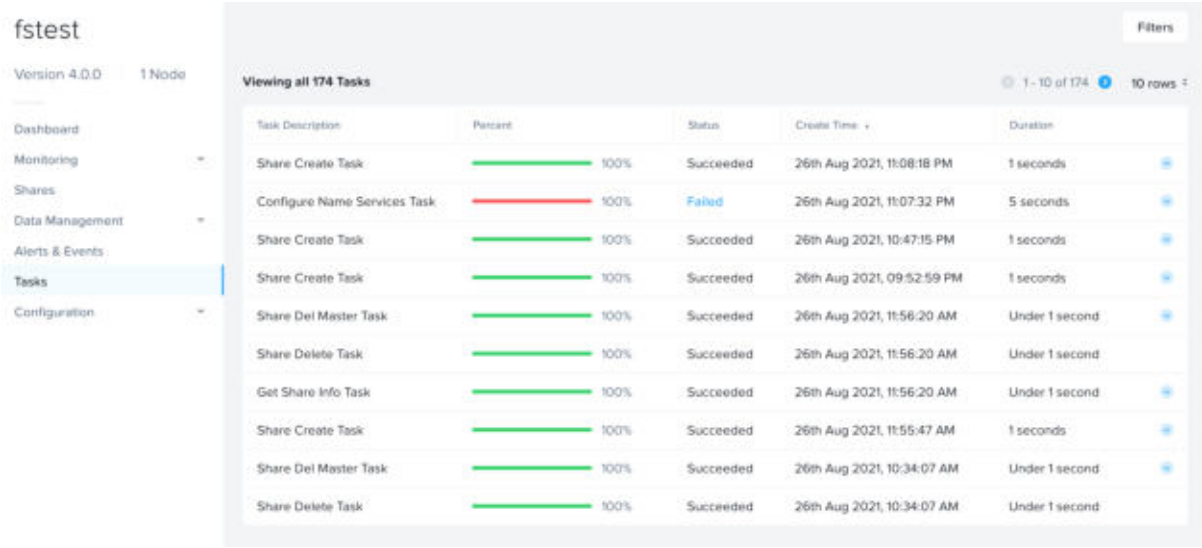


Figure 22: Tasks View

The following tables describe the columns in the **Tasks** table.

Table 16: Table

Column	Description	Values
Task description	Describes the operation that triggered the task.	(name of operation such as "create share")
Percent	Indicates the current completeness of the task as a percentage.	(0%-100%)
Status	Indicates the status of the task.	[succeeded, running, failed, queued]
Create time	Indicates when the task began.	(x) [seconds, minutes, hours, days]
Duration	Indicates how long the task has been running.	(x) [seconds, minutes, hours, days]

Configuration View

The **Configuration** view in the Files Console.

The **Configuration** view includes the following tabs:

- Authentication
- Blocked file types
- Manage roles
- Update DNS entries
- Antivirus
- Platform

The **Platfrom** view includes the following elements:

- An **Update** drop-down menu includes options to update file server basics, scale up/ scale down, and update the DNS and NTP servers.
- The **Configuration Summary** provides details about the configuration of the file server.
- The **Files Cluster** diagram provides a visual diagram of the file server configuration.

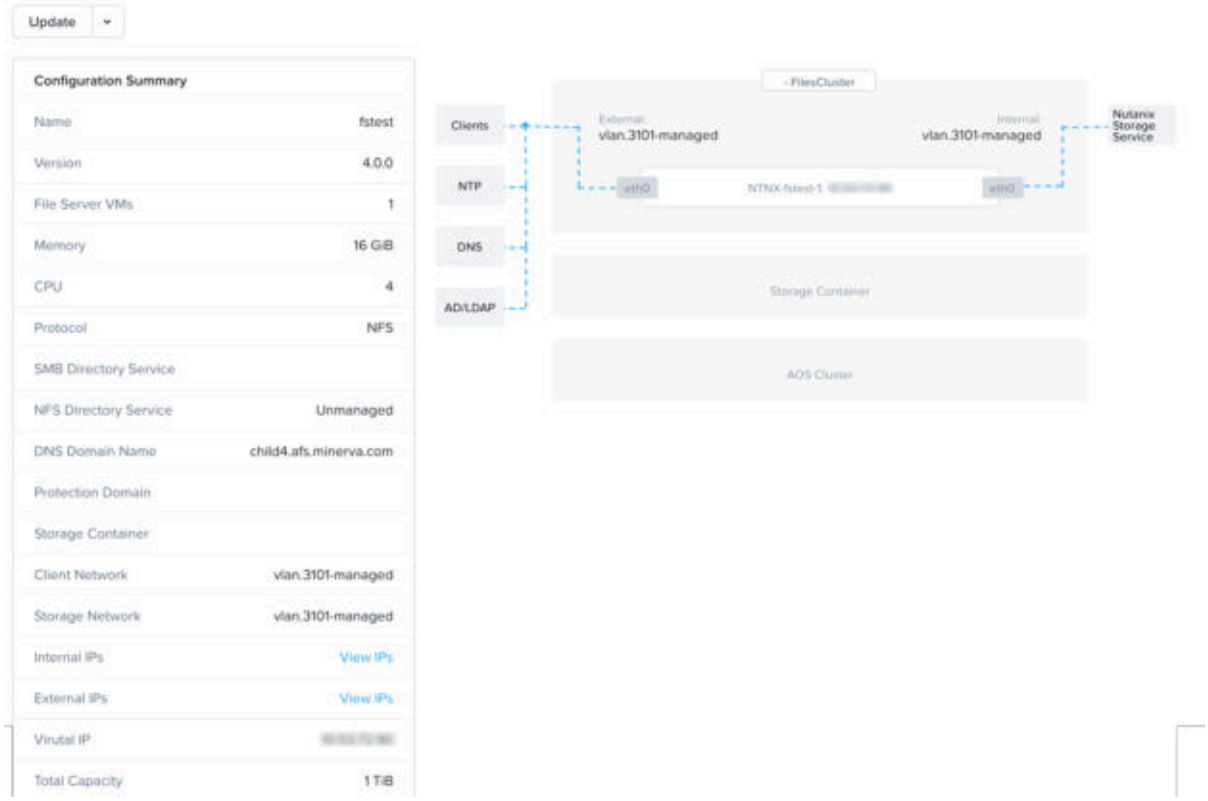


Figure 23: Platform View

Table 17: Configuration Summary

Parameter	Description
Name	The name of the file server.
Version	The Files version.
File Server VMs	Number of file server VMs on the file server.
Memory	Maximum configured memory.
CPU	Maximum configured CPU.
Protocol	The primary protocol and, when applicable, the secondary protocol (see Multi-Protocol Support for Files on page 85).
SMB directory service	The configured directory service for SMB shares.
NFS directory service	The configured directory service for NFS shares.
DNS domain name	The name of the domain name system (DNS) for the file server.
Protection domain	The name of the protection domain.
Storage container	The name of the storage container.
Client network	The name of the client network.

Parameter	Description
Storage network	The name of the storage network.
Internal IPs	The internal IP addresses of the file server.
External IPs	The external IP addresses of the file server.
Virtual IPs	The virtual IP addresses of the file server.
Total capacity	The maximum capacity for the file server.
Data savings	The amount of data saved on the file server.

File Analytics

File Analytics provides data and statistics on the operations and contents of a file server.

Once deployed, Files adds a File Analytics VM (FAVM) to the Files cluster. A single FAVM supports all file servers in the cluster, but you must enable Analytics separately for each file server. Files protects the data on the FAVM and keeps it in a separate volume group.

For deployment steps and administrative guidance, refer to the *File Analytics Guide*.

FILE SERVER MANAGEMENT

Create a new file server and manage its configuration.

Creating and managing a file server requires administrator privileges. You can perform most file server management tasks in the File Server view in Prism Element and through the Files Console. You can start, stop, and manage file server VMs (FSVMs) through the VM tab in Prism Element, see "VM Management" in the *Prism Web Console Guide* for more details.

A file server usually consists of three or more FSVMs combined into a logical group. You cannot reduce file servers of three FSVMs, and you cannot expand single-FSVM file servers.

Single-FSVM deployments have the following limitations:

- No High Availability (HA) support. Single-FSVM deployments do not include other FSVMs that an out-of-service FSVM can fall back on, see High Availability on page 145.
- No distributed share support. Since distributed shares and exports spread data across multiple FSVMs, single-FSVM deployments only support standard shares.

> **Important:** When the file server capacity or the container capacity reaches 100 percent, all shares or exports within the file server become read-only and Files blocks write privileges. Prism displays alerts when the file server capacity or the container capacity reaches 90 percent and when either reaches 100 percent capacity.

Creating a File Server

Follow the procedure to create a file server with one or multiple file server VMs (FSVMs).

About this task

> **Note:** Provide two static IP addresses for each new FSVM, one for the client-side network and one for the storage network.

Procedure

1. Go to the **File Server** view in Prism Element (PE) (see File Server View in Prism on page 15) and click the **+ File Server** button.

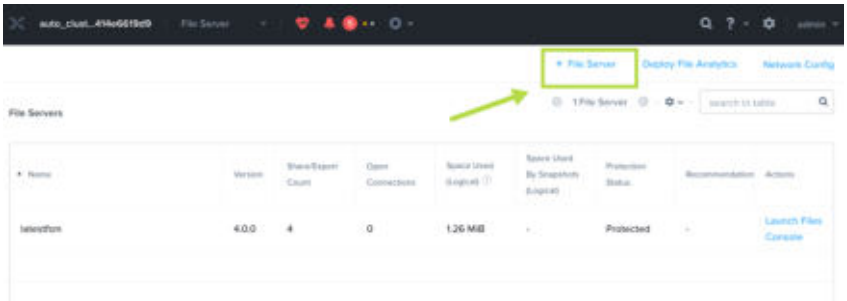


Figure 24: + File Server button

2. If a **New File Server: Pre-Check** window appears, review the displayed information and address any unsatisfied prerequisites before continuing.

   Files checks your current environment and either verifies it satisfies the prerequisites or identifies where it does not meet the requirements (see Prerequisites on page 7). A blue check mark indicates a satisfied prerequisite.

Any unchecked items need attention. The checks are for the Files software version, available data services IP addresses, available networks, and (in the ESXi case) whether vCenter registered the cluster.

> **Note:** To use the high-availability (HA) and DRS features, add all ESXi hosts under the same ESXi cluster and place one FSVM per ESXi host.

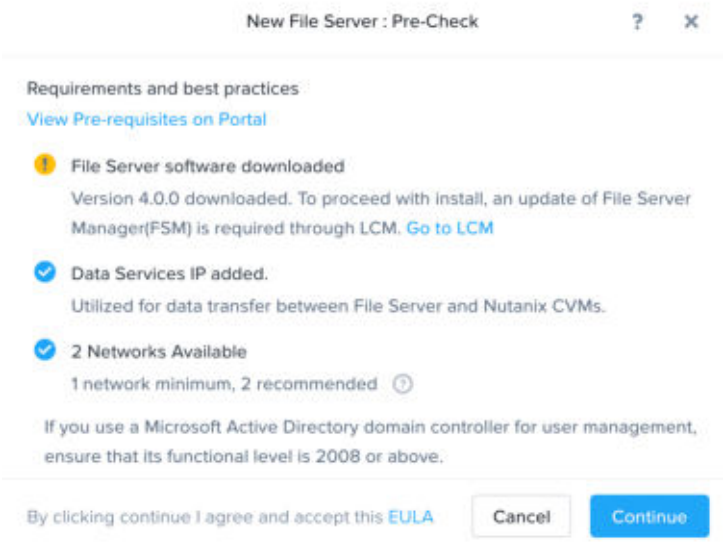


Figure 25: New File Server: Pre-Check Window

3. The **Create File Server** window appears and displays the **Basics** tab. Do the following in the indicated fields:

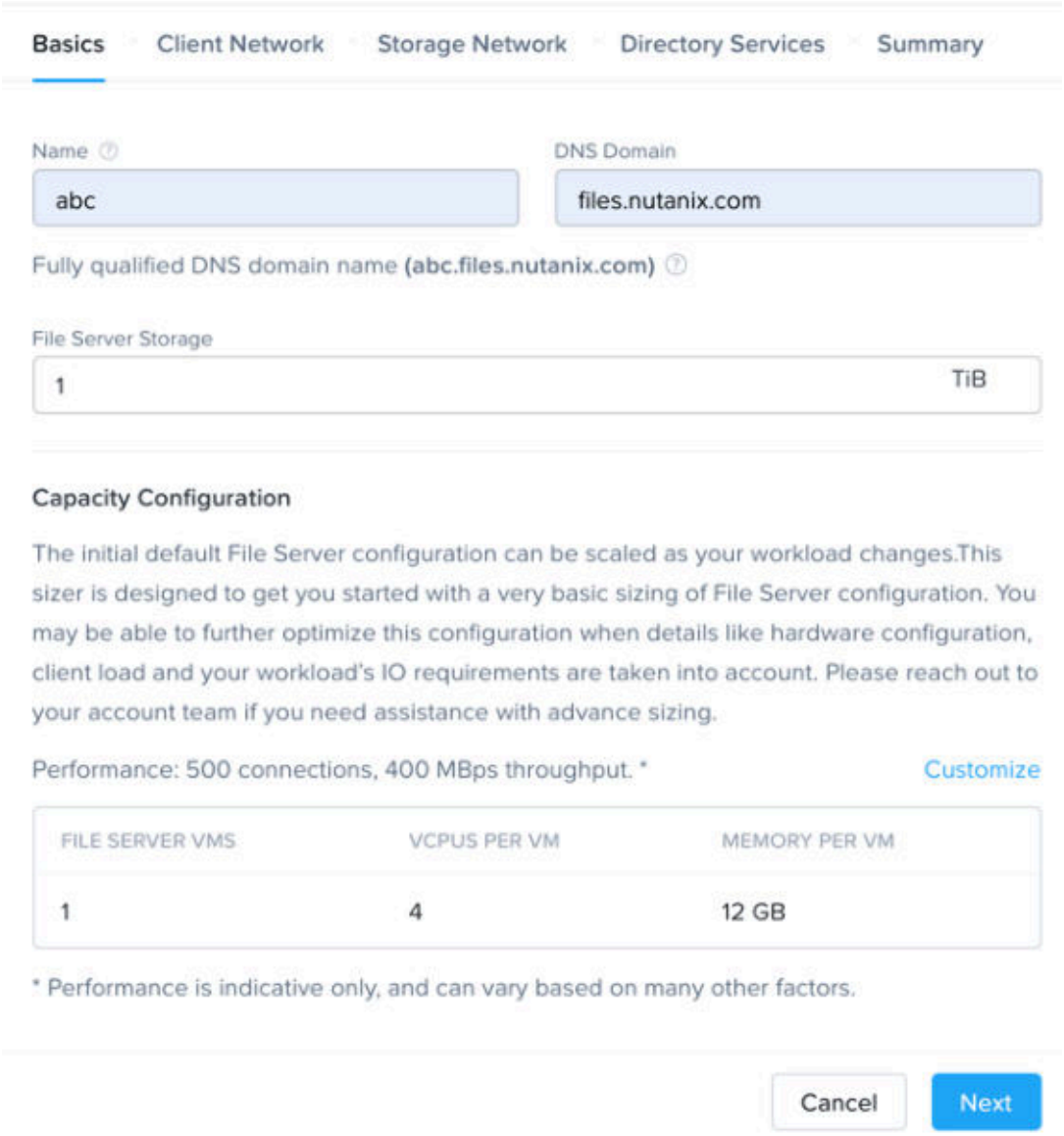


Figure 26: Create File Server: Basics Tab

a. **Name**: Enter a name for the file server.

   Clients use the file server name to access the file server. The fully qualified name (file server name + domain) must be unique.

b. **DNS Domain**: Enter a fully qualified domain.

c. **File Server Storage**: Enter the file server total storage size (minimum 1 TiB).

d. **Capacity Configuration**: Files automatically recommends the number of file server VMs, vCPUs per VM, and memory per VM. Review the performance values (connections and throughput) and then do one of the following:

> **Note:** For single-FSVM deployments, the number of suggested FSVMs is always one.

- If you are satisfied with the recommended configuration, click **Next** button.
- To change the configuration, click **Customize**. The **File Server Capacity Configuration** window displays. To change the configuration based on performance requirements, enter the target number of connections and throughput amount (in MBps) in the **SMB Concurrent Connections** and **NFS Throughput in MBps** fields and then click **Save**.

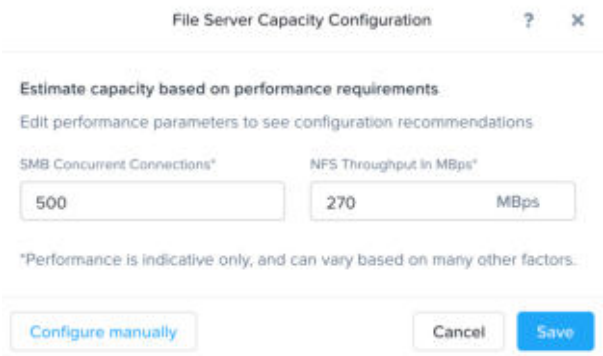


Figure 27: File Server Capacity Configuration window

- To manually specify the FSVM parameters, click**Customize** > **Configure manually**. The **File Server Capacity Configuration** window displays. Enter the number and capacity of the FSVMs in the **Number of File Server VMs**, **VCPUs Per VM**, and **Memory Per VM** (in GIB) fields, and then click **Save**.

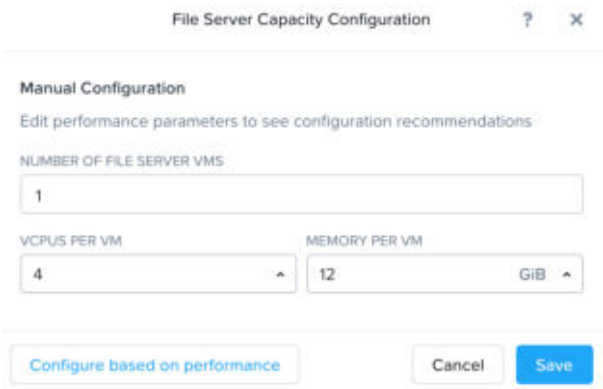


Figure 28: File Server Capacity Configuration window

4. In the **Client Network** tab, do the following in the indicated fields:

> **Note:** If clients or the AD domain controllers are in the same subnet as the controller VM (CVM) or the storage network, configure the same client-side and storage-side networks. Otherwise, use separate client and storage

networks. If you use the same network for both clients and storage, use unique IP addresses; otherwise clients on the same subnet as the storage network cannot access the shares.

a. **VLAN** (AHV) or **Port Group** (ESXi): Select the target VLAN or port group from the drop-down list.

After selecting the target, configured network parameters display if the target is a managed network.

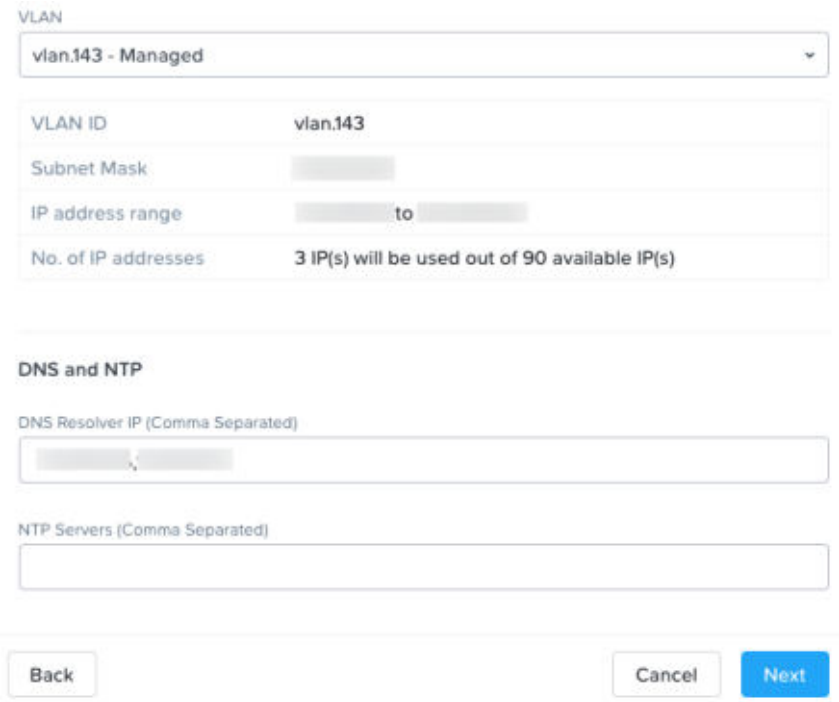


Figure 29: Client Network Details (managed VLAN)

b. For unmanaged networks, the following fields appear:

- **Subnet Mask**: Enter the subnet mask value.
- **Gateway**: Enter the IPv4 gateway IP address.
- **# IP addresses required**: Click the **+ IP Addresses** button to display a row for entering IP addresses. Enter the starting IP address in the **From** field, the ending IP address in the **To** field, and click **Save**. (Usually, Files requires three IP addresses. Files requires one IP address for single-FSVM deployments.)
- **Add IPv6**: Check this box to enable the IPv6 protocol.
- **Prefix Length**: Enter an integer for the prefix length (the default IPv6 prefix is 64).
- **Gateway**: Enter the alpha-numeric, colon-separated IPv6 gateway address.
- Click the **+ IP Addresses** button to display a row for entering IP addresses. Enter the starting IP address in the **From**, and enter the total number of IP addresses in the **No. of IPs** field (usually Files requires three IP addresses). Click **Save**.

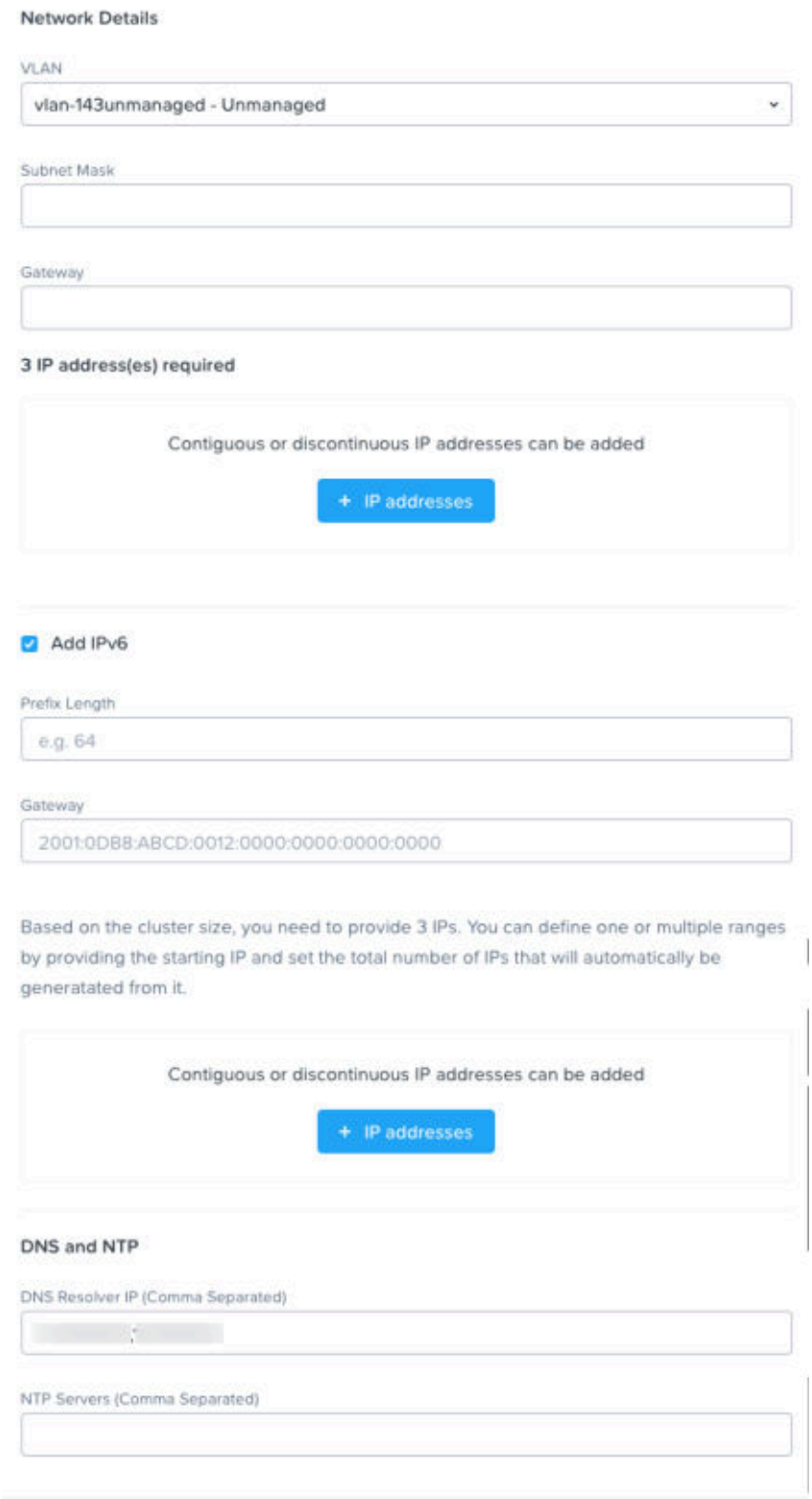


Figure 30: Client Network Details (unmanaged VLAN)

c. **DNS Resolver IP**: Enter IP addresses for one or more DNS servers (IPv4 only). Use a comma-separated list for multiple entries.

> **Note:** The DNS Resolver IP and the NTP servers fields can pre-populate based on the target VLAN or port group.

d. **NTP Servers**: Enter the server names or IP addresses for the NTP servers. Use a comma-separated list for multiple entries.

e. When all the entries are correct, click **Next**.

5. In the **Storage Network** tab, do the following in the indicated fields:

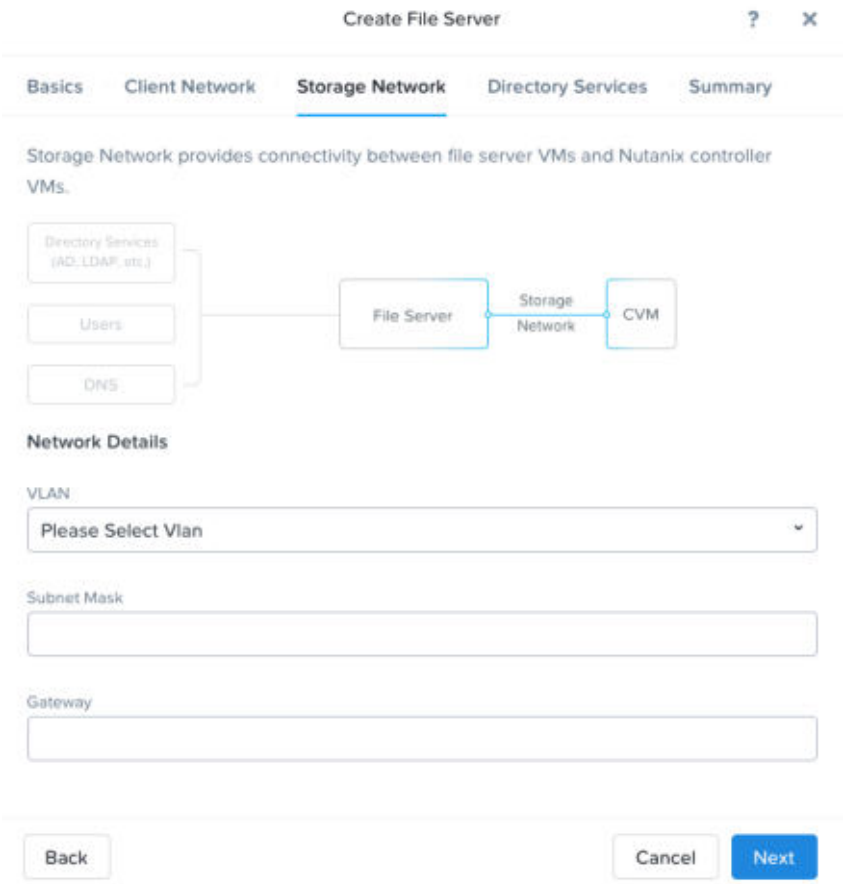


Figure 31: Create File Server: Storage Network Tab

a. **VLAN** (AHV) or **Port Group** (ESXi): Use the pull-down list to select the desired VLAN or port group for the storage network.

   Once you select the target, if the target is a managed network, Files displays configured network parameters.

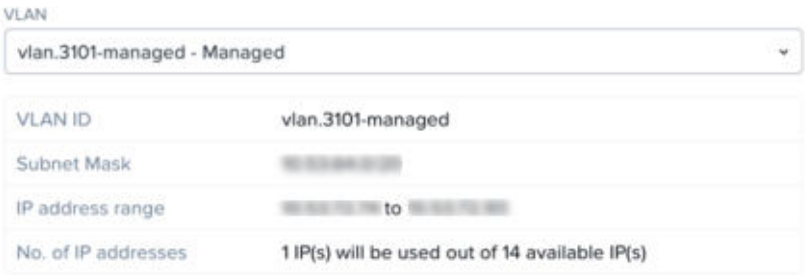


Figure 32: Storage Network Details (managed VLAN selected)

b. For unmanaged networks, the same fields appear as for the client network (**Subnet Mask**, **Gateway**, and **# IP addresses required**). Enter the appropriate values for the storage network. In this case, Files normally requires four IP addresses.

c. When all the entries are correct, click **Next**.

6. In the **Directory Services** tab, select one or more protocols to use (check the **Use SMB Protocol** box, **Use NFS Protocol** box, or both boxes).

> **Note:** You can skip this step and select the protocols later, but you cannot use the file server until this step is complete.

> **Note:** Files supports using LDAP for AD.

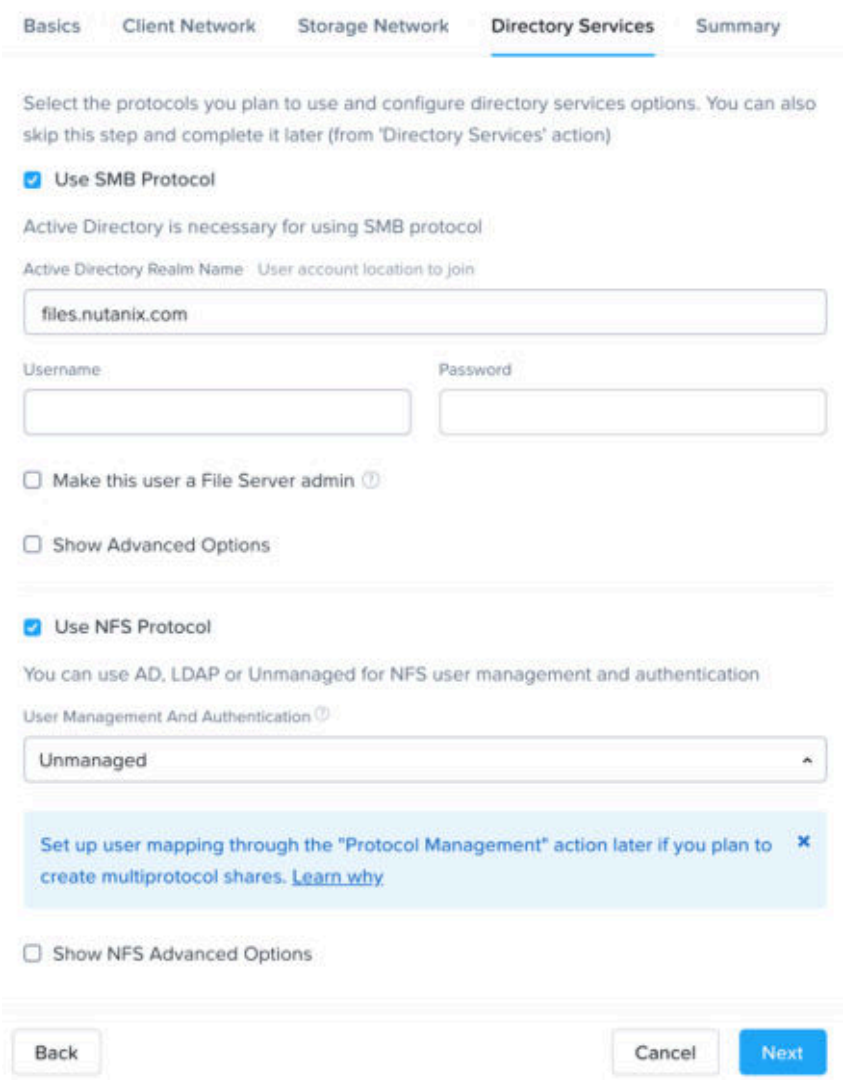


Figure 33: Create File Server: Directory Services Tab

7. If you selected the SMB protocol, do the following in the indicated fields:

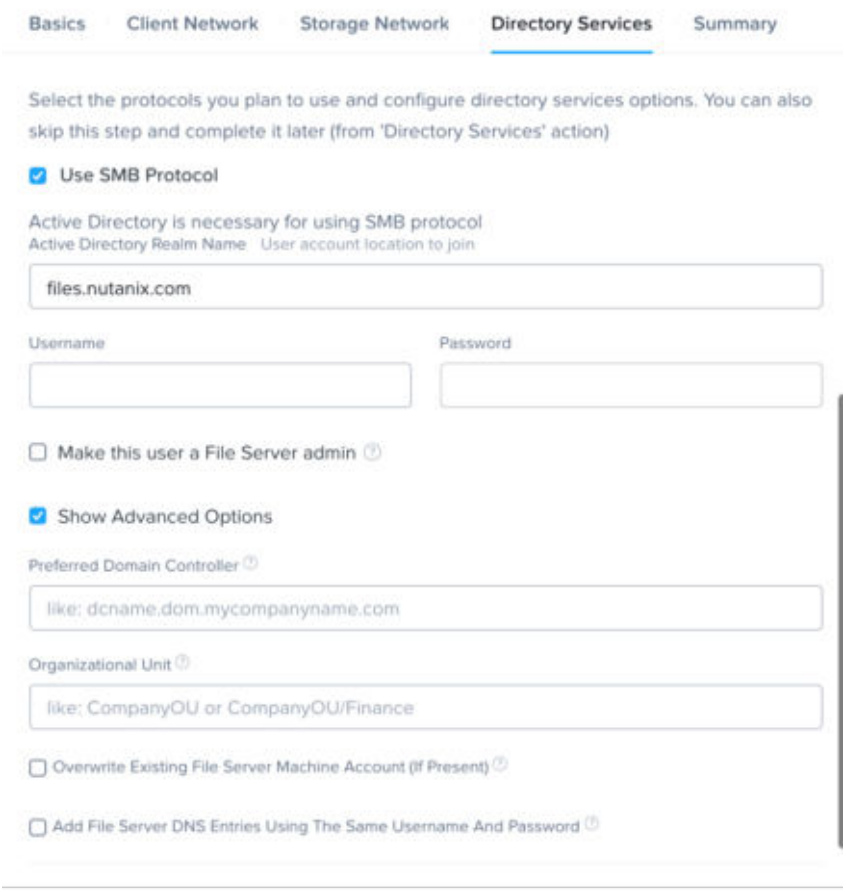


Figure 34: SMB Fields

a. **Active Directory Realm Name**: Enter the Active Directory realm name for join domain operations.

b. **Username**: Enter an AD username in the `domain|username` format or in the UPN format `username@ADrealm`. You must have an administrator account or the following required permissions for the relevant AD organizational unit (OU) realm:

   - Create computer objects
   - Read/Write service principal names

   Note:

   - Do not remove the file-server computer object in AD, as that can cause file-server services to be disrupted.
   - If the DNS domain name of the file server is different from the AD domain name, the following permission is also required:
     - Read/Write DNS host name

c. **Password**: Enter the password for the username.

d. **Make this user a File Server admin**: Check this box if you want to make this user an administrator for the file server.

e. **Show Advanced Options** (optional): Check this box to display three more fields to enter your preferred domain controller and organizational unit name for identification purposes.

f. [advanced option] **Preferred Domain Controller**: Specify a domain controller for join-domain operations in a multi-DC Active Directory configuration. Files discovers and uses a site local domain controller by default.

g. [advanced option] **Organizational Unit**: Enter your preferred organizational unit name. Files creates the machine account in the "Computers" OU container by default. If there is a **Sub-OU** field, use the following format: `parentcompanyOU/sub-ou1/sub-ou2/sub-oux`.

> **Note:** The default machine account password expiry period is 0, meaning the password does not expire. To update the machine account password expiry period, see Setting AD Machine Account Password Expiry on page 105.

h. [advanced option] **Overwrite Existing Files Machine Account (if present)**: Check this box to overwrite an existing machine account during the join-domain operation if one exists with the same name as the file server.

i. [advanced option] **Add Files Server DNS Entries Using the Same Username And Password**: Check this box to use AD credentials for adding DNS entries.

   Ensure that the AD user account has DNS admin rights (when using Microsoft DNS).

8. If you selected the NFS protocol, select the authentication method to use from the pull-down list in the **User Management and Authentication** field and then do the following:

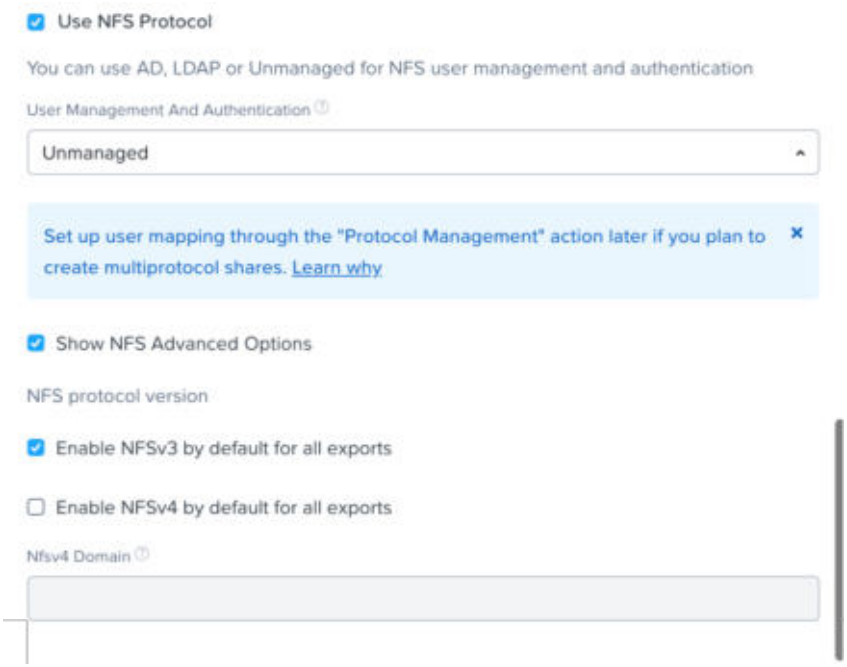


Figure 35: NFS Fields (Unmanaged + advanced options)

a. **User Management and Authentication: Unmanaged**: If you select unmanaged as the authentication method, there is nothing else to specify here.

b. **User Management and Authentication: LDAP**: If you select LDAP as the authentication method, there are several more fields:

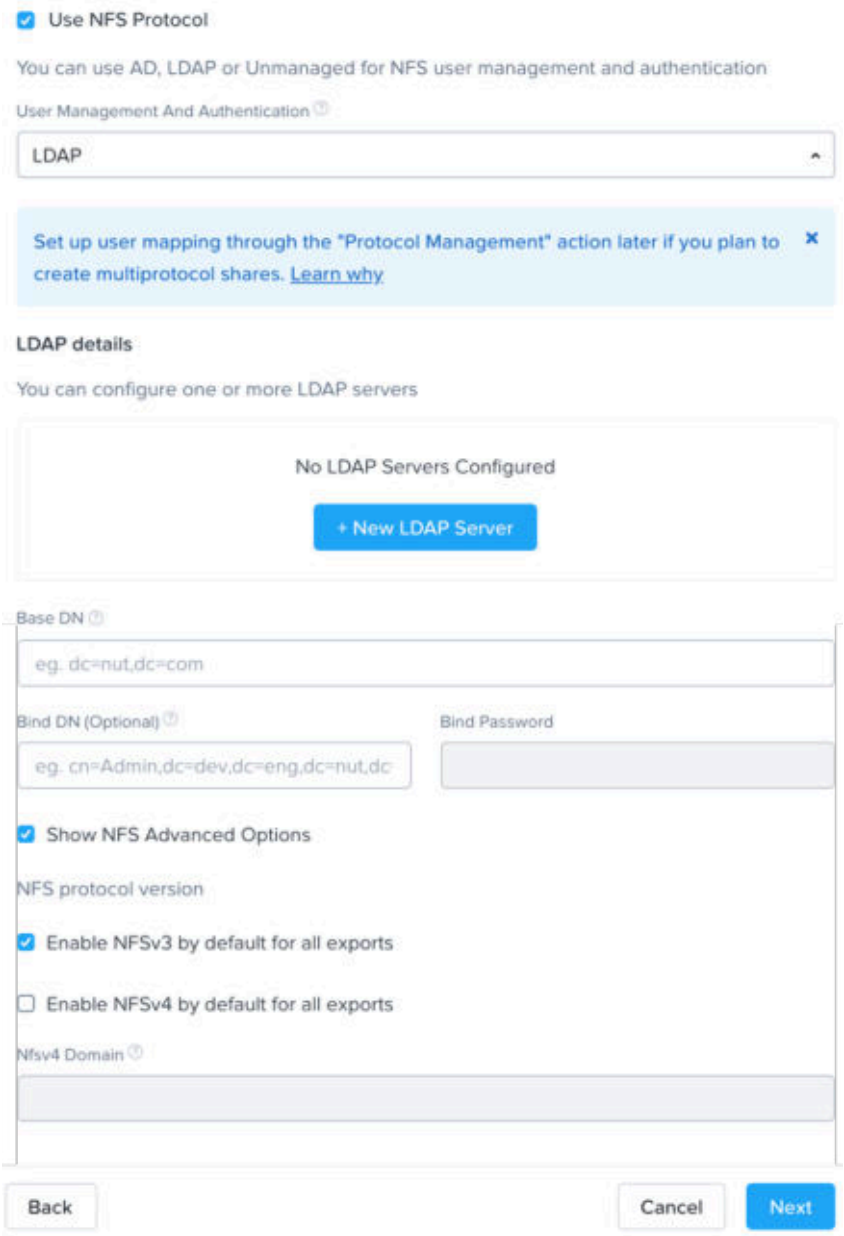


Figure 36: NFS Fields (LDAP)

- **LDAP details**: This section displays available LDAP (and LDAPS) servers. If there is no data or the target server is not listed, click the **+ New LDAP Server** button. A line appears in the table. Enter the URI address for the server in the **Server URI** column and then click **Save** in the **Action** column for that

row. Repeat for any additional LDAP servers. For LDAPS servers, use the LDAPS format `ldaps://ldap-servername.company.com:port`.

- **Base DN**: Enter the base distinguished name (search starting point).
- **Bind DN** (optional): Enter the distinguished name at which to bind to the server.
- **Bind password**: Enter the bind password.

If you configured LDAP or LDAPS settings, you may skip the next step.

c. **User Management and Authentication: Active Directory**: If you select Active Directory as the authentication method, there are several extra fields:

> **Note:** When NFS uses AD, you do not need to configure user mapping for multi-protocol access.

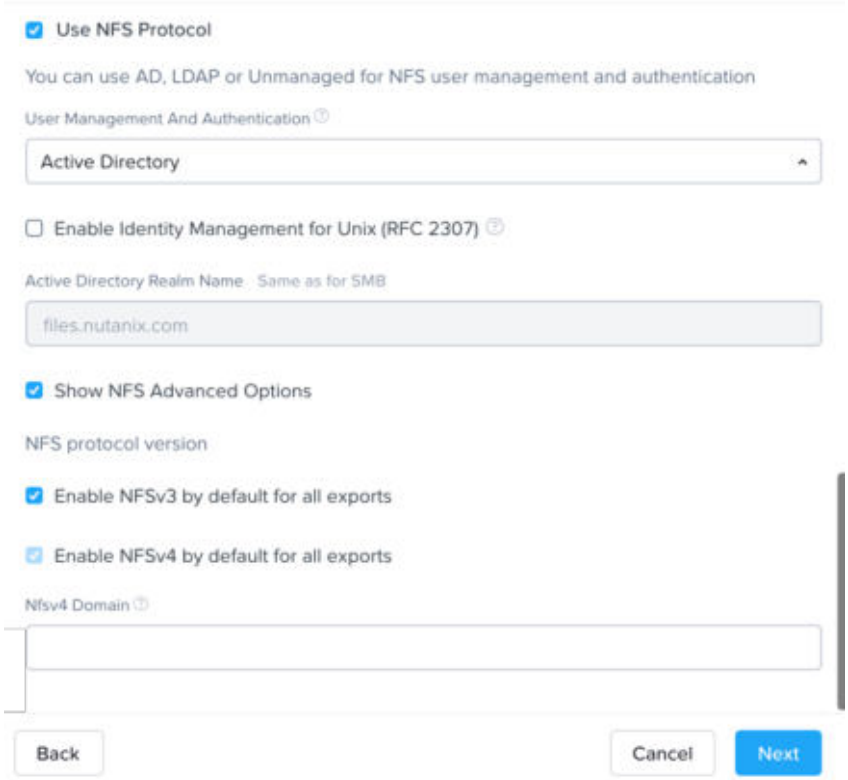


Figure 37: NFS Fields (Active Directory)

- **Enable Identity Management for Unix (RFC 2307)**: Check this box if you have RFC 2307 configured for Active Directory.
- **Active Directory Realm Name**: Displays the Active Directory realm name (read-only).

> **Note:** If you previously configured SMB, the following fields do not appear because they are already configured for the SMB protocol.

- **Username**: Enter an AD username in the `domain|username` format or in the UPN format `username@ADrealm`. You must have an administrator account or the following required permissions for the relevant AD organizational unit (OU) realm:
  - Create computer objects

- Read/Write service principal names

> **Note:** If the DNS domain name of the file server is different from the AD domain name, the following permission is also required:
>
> - Read/Write DNS host name

- **Password**: Enter the password for the user.
- **Show Active Directory Advanced Options** (optional): Check this box to display the following three more fields.
- **Preferred Domain Controller**: Specify a specific domain controller for join-domain operations in a multi-DC Active Directory configuration. Files discovers and uses a site local domain controller by default.
- **Organizational Unit**: Enter your preferred organizational unit name. Files creates the machine account in the "Computers" OU container by default.

> **Note:** The default machine account password expiry period is 0, meaning the password does not expire. To update the machine account password expiry period, see Setting AD Machine Account Password Expiry on page 105.

- **Overwrite Existing Files Machine Account (if present)**: Check this box to overwrite an existing machine account during the join-domain operation if one exists with the same name as the file server.

d. **Show NFS Advanced Options** (optional): By default Files supports both NFSv3 and NFSv4 protocols for exports. Check this box to modify the default NFS protocol version for all exports on the file server.

- **Enable NFSv3 by default for all exports**: Un-check to disable NFSv3 by default for all exports.
- **Enable NFSv4 by default for all exports**: Un-check to disable NFSv4 by default for all exports.
- In the **NFSV4 Domain** field, enter the NFSv4 domain name. Files uses the DNS domain name to map NFSv4 names to UIDs and GIDs; clients and servers must agree on the mapping.

e. When all the fields in the **Directory Services** tab are correct, click **Next**.

9. In the **Summary** tab, review the displayed information.

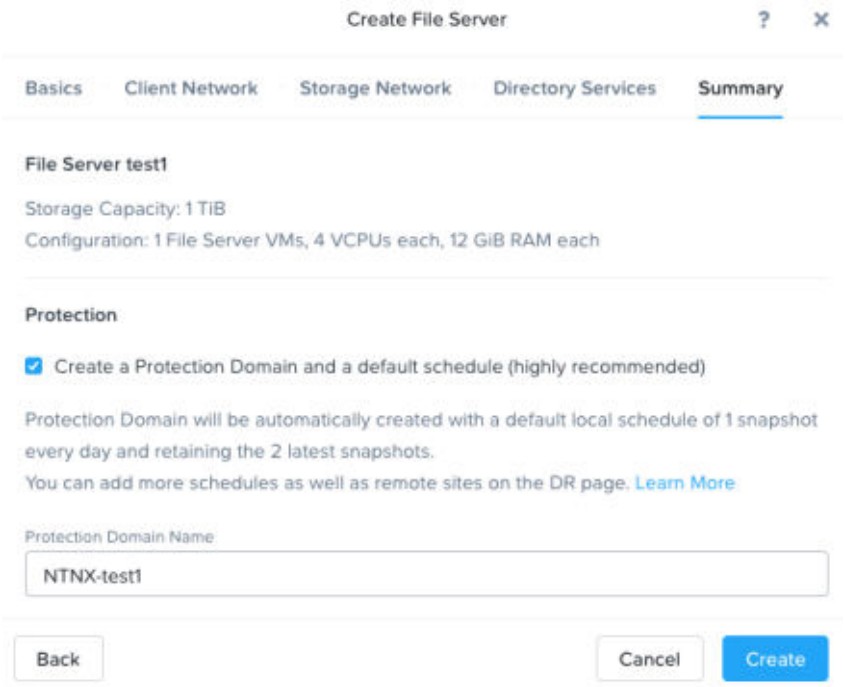


Figure 38: Create File Server: Summary Tab

a. Review and verify the file server information you entered.

b. The default protection domain name is `NTNX-`*`file_server_name`*, but you can change that name by entering a new name in the **Protection Domain Name** field.

c. When all the information is correct, click **Create**.

Creating the file server begins. You can monitor progress through the **Tasks** page. (Either select **Tasks** from the main menu or click the tasks icon and then the **View All Tasks** link.)

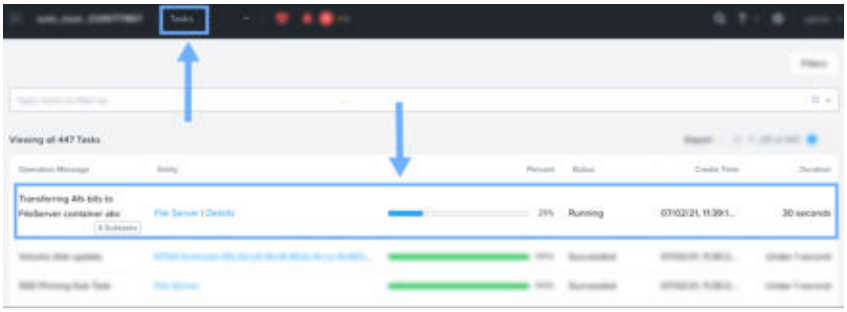


Figure 39: Task Dashboard

What to do next

- If you choose to add DNS entries, see Updating Domain Name System (DNS) Entries on page 101.

- On deployments using the ESXi hypervisor, disable VM monitoring. See the VMware vSphere documentation for procedures describing how to enable or disable VM monitoring.

Deleting a File Server

Delete a file server in Prism.

About this task

Do not delete FSVMs or power off a file server prior to deleting in. If you delete FSVMs, you will need to force delete the file server.

To delete a file server, do the following:

> **CAUTION:** Deleting a file server is permanent, and none of the deleted files can be recovered.

Procedure

1. Go to the File Server view (see File Server View in Prism on page 15), select the target file server, and click the **Delete** action button.
2. In the dialog box, choose one of the following options:
   - » Retain all the related entities.
   - » Delete specific related entities.

   Deleting a file server deletes all the underlying user data related to that file server. However, depending on your AOS version, you may not see the option to delete the storage container and protection domain created for that file server. You can delete that storage container and protection domain manually.

3. Click the **Delete** button.

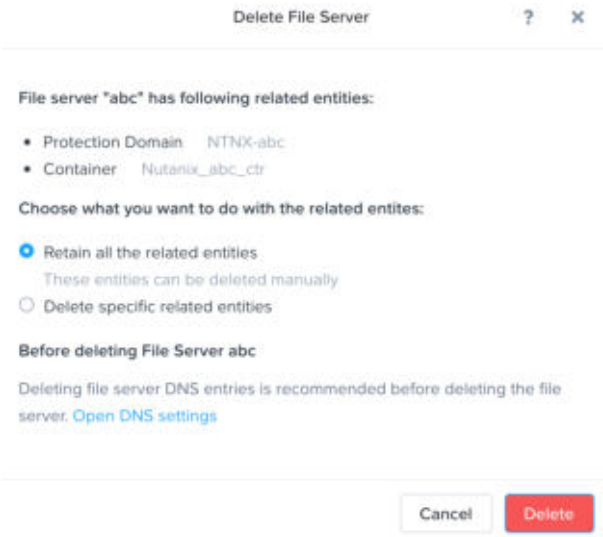


Figure 40: Delete Window

   a. If you decide to delete specific related entities, choose which entities to delete from the following options (otherwise continue to the next step):

      - **Snapshots, schedules, and protection domain**
      - **Container**

4. (Optional) delete file server DNS entries. Click **Open DNS settings**, see Updating Domain Name System (DNS) Entries on page 101.

5. Click **Delete**.

File Server Updates

Update or expand the configuration of an existing file server.

You can modify an existing file server in the following areas:

- Update the name, domain, or storage capacity, see Updating File Server Basics on page 59.
- Update the number of file server VMs (FSVMs), see Scaling FSVMs on page 57.

  Note: You cannot update the number of FSVMs on single-FSVM deployments.

  Note: Number of CVMs must be equal to or greater than the number of FSVMs.

- Update the vCPU count and memory size for each FSVM, see Updating Memory and vCPU Resources on page 58.
- Update the network configuration (client network, storage network, domain name server (DNS), NTP server).
  - Go to the file server dashboard (see File Server View in Prism on page 15), select the target file server, and click **Network Config**. Follow the steps as indicated in File Server Updates on page 53.

- Update blocked file types, see Blocking Files on a File Server on page 127.

Updating the Network Configuration

Change the client or storage network IP address configuration of a file server (also referred to as "re-IP").

Before you begin

Stop a file server before updating the network to reduce traffic during the update process, see Stopping a Files Cluster on page 62.

About this task

To configure a virtual network for guest VMs, see "Configuring a Virtual Network for Guest VM Interfaces" in the *Prism Web Console Guide*. To modify network connections, see "Modifying Network Connections" in the *Prism Web Console Guide*.

To update the client or storage network configuration, follow the steps as indicated.

> **Note:** Updating the network configuration requires a temporary downtime. Nutanix recommends performing network updates during off-peak hours.

Procedure

1. In the **file server** view, select the target file server.
2. In the options bar go to **Update** > **Network Configuration**.

3. Do the following in the **Update File Server Network** window.

> **Note:** You can change the client network, storage network, DNS, and NTP IP addresses for a file server VM. This lets you move a file server from one data center to another. If you intend on changing the domain, you must un-join the domain first.

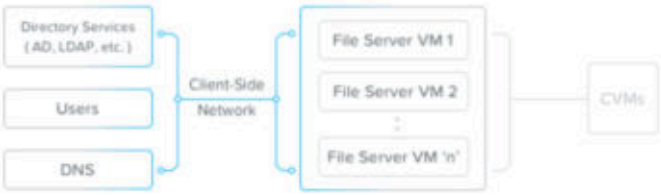
   a. In the **Client Network** tab, verify or update the client network details (VLAN, IPv6 (optional), DNS, and NTP entries) as needed and click **Next**.

   See Creating a File Server on page 36 for more information about these fields.

   > **Note:** If you change the client network configuration, click the **file server DNS entries** link and delete the existing DNS entries (see Updating Domain Name System (DNS) Entries on page 101).

Client Network Storage Network

Set up a network which communicates between the File Server VMs as well as connectivity to DNS, AD and LDAP.



Network Details

Please remove existing file server DNS entries before changing the client network.

VLAN

vlan-143unmanaged - Unmanaged

Subnet Mask

Gateway

4 IP address(es) required

FROM	TO	NO OF IPS	ACTIONS
		4	

☑ Add IPv6

Prefix Length

64

Gateway

Based on the cluster size, you need to provide 3 IPs. You can define one or multiple ranges by providing the starting IP and set the total number of IPs that will automatically be generatated from it.

4 IP address(es) required

FROM	NO OF IPS	ACTIONS
	4	

DNS and NTP

DNS Resolver IP (Comma Separated)

NTP Servers (Comma Separated)

b. In the **Storage Network** tab, verify or update as needed the storage network details and then click the **Save** button.

See Creating a File Server on page 36 for more information about these fields.

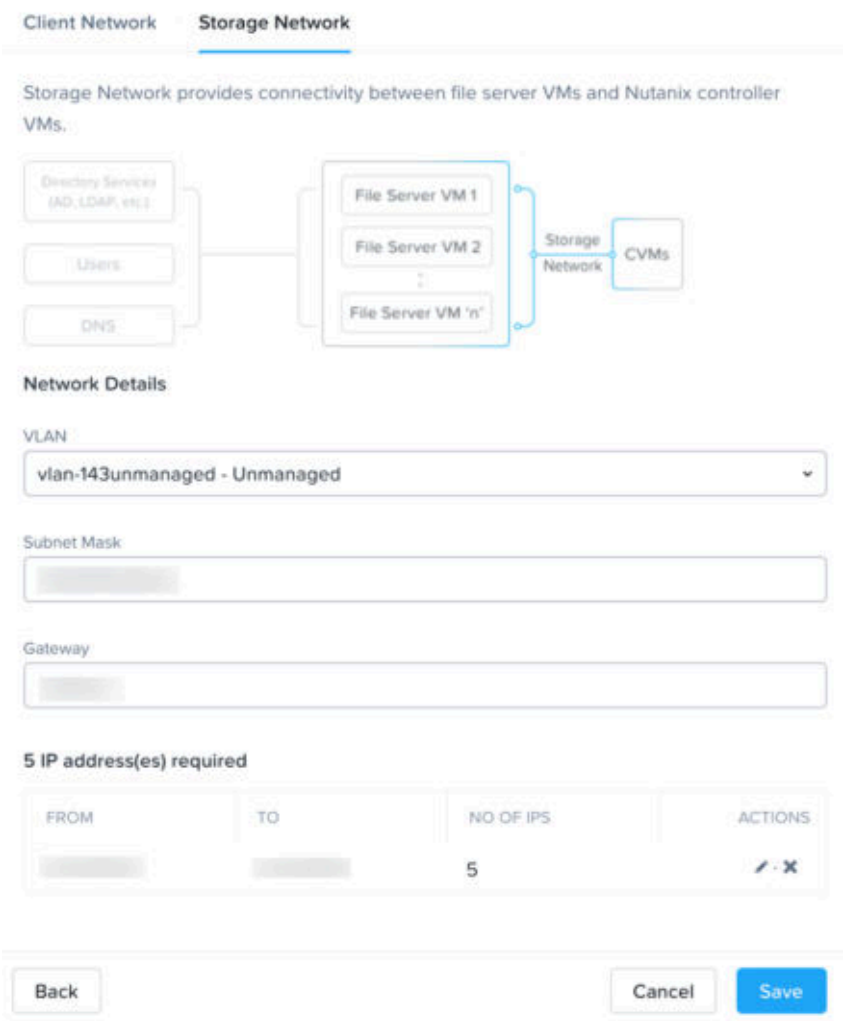


Figure 42: Update File Server Network: Storage Network tab

Note: If the network update operation is unsuccessful, use the original IP address details and try the update operation again.

Scaling FSVMs

Add or remove file server VMs (FSVMs) on your file server.

About this task

File Server Management on page 36 describes restrictions for increasing and reducing the number of FSVMs. Do the following to change the number of FSVMs.

The following limitations apply to changing the number of FSVMS:

- You cannot reduce file servers of three FSVMs.
- You cannot expand single-FSVM deployments.

- Single or dual-node AOS clusters only support single-FSVM deployments.

Procedure

1. In the Files Console, go to **Configuration** > **Platform**.
2. Click **Update** > **Scale Out** / **Scale In**.
   The **Scale Out or Scale In** window appears.
3. Under **New Capacity**, enter an integer for the new number of FSVMs.
4. Review the settings in the **Client Network** and **Storage Network** sections.

   > **Note:** For information on configuring network settings, see the client network and storage network sections in Creating a File Server on page 36.

5. Click **Update**.

Updating Memory and vCPU Resources

Change the vCPUs and memory resources per FSVM.

About this task

Follow the steps as indicated to update the memory and vCPU capacity.

> **Note:** Reducing the configuration of a filer server causes system downtime.

Procedure

1. In the Files Console, go to **Configuration** > **Platform**.
2. Click **Update** > **Scale Up** / **Scale Down**.
   The **Scale Out or Scale In** window appears.

3. Do the following in the **Updated Configuration** section.

   a. In the **vCPUs** field, select the target (total) number of vCPUs for each FSVM in the cluster.

Figure 43: Scale Up/ Scale Down File Server

   b. In the **Memory** dropdown, select the target (total) number of memory (GiB) for each FSVM in the cluster.

   c. Click **Update**.

Updating File Server Basics

Update the name, the domain, and the size of a file server.

About this task

Context for the current task

Procedure

1. In the Files Console, go to **Configuration** > **Platform**.

2. Click **Update** > **File Server Basics**.
   The **File Server Basics** window appears.

3. Do the following in the indicated fields:

> **Tip:** Before changing the file server or domain name, click the **Leave Domain** link to leave the current domain (see Leaving a Domain on page 100). Changing a file server name does not remove existing DNS entries, which you must remove manually.

   a. To change or update the file server name, enter the new name in the **File Server Name** field.

   b. To change the DNS domain name, enter the new name in the **Domain** field.

> **Tip:** After changing the file server or domain names, delete any old DNS entries from the DNS servers, add the new DNS entries (see Updating Domain Name System (DNS) Entries on page 101), and update the Directory Services configuration as needed (see Updating Directory Services on page 105).

   c. To change the size of the file server, enter a value in the **File Server Size (Logical)** field.

> **Note:** You cannot reduce the size of the file server. This operation does not impact current client connections.

4. When the entries are correct, click the **Update** button.

Logging Onto A File Server VM

The following procedures describes how to SSH onto a file server VM (FSVM).

About this task

Procedure

1. Retrieve the IP address for the FSVM.

   a. SSH onto the CVM.

   b. Enter the CVM password.

   c. List the FSVM IPs.

```
nutanix@cvm$ ncli fs ls
```

   FSVM IPs appear in the following format `Nvm IP Addresses: internal IP,external IP.`

2. SSH onto the FSVM using the internal IP.

```
nutanix@cvm$ ssh internal_ip_address
```

Changing an FSVM Password

Steps for updating the password of a file server VM (FSVM).

About this task

Follow the steps as indicated.

Procedure

1. Log on to a file server VM with SSH.

2. Change the nutanix password.

```
nutanix@fsvm$ sudo passwd nutanix
```

3. Respond to the prompts, providing the current and new nutanix user password.

```
Changing password for user nutanix.
Old Password:
New password:
Retype new password:
passwd: all authentication tokens updated successfully.
```

The password must meet the following complexity requirements:

- At least 8 characters long
- At least 1 lowercase letter
- At least 1 uppercase letter
- At least 1 number
- At least 1 special character
- At least 4 characters difference from the old password
- Should not be among the last 10 passwords

The operation updates the password of every FSVM on the file server.

Setting Timezones

About this task

Set the time to your timezone for all the FSVMs.

This time zone persists across all FSVMs when Files upgrades.

Procedure

1. SSH into any FSVM in your cluster.
2. Get the timezone for your region. The command returns your timezone into a region and city format.

```
nutanix@fsvm$ afs fs.get_timezone
```

> **Note:** To list all available timezones, see the contents of the `zoneinfo` file for the desired region using the `ls /usr/share/zoneinfo/`*`region`*command.

3. Set the timezone according to the output.

```
nutanix@fsvm$ afs fs.set_timezone "region_name/city_name"
```

Replace *`region_name`* and *`city_name`* with the region and city name specified in the timezone output. For example, `afs fs.set_timezone Asia/Kolkata`.

Starting a Files Cluster

Start a Files cluster on a CVM using the commands listed in this topic.

About this task

A deployed filer server creates multiple file server VMs (FSVMs) on the cluster. A Files cluster is a set of FSVMs. To start a Files cluster, do the following:

Procedure

1. Using SSH, log on to a Controller VM of the Nutanix cluster that deployed Files.
2. Get a list of file servers.

   ```
   nutanix@cvm$ afs info.fileservers
   ```

3. Choose to start a single file server or all file servers.

   » Start a single file server.

   ```
   nutanix@cvm$ afs infra.start fs_names
   ```

   Replace *fs_names* with the name of a single file server or with a comma-separated list of multiple file servers.

   » Start all file servers.

   ```
   nutanix@cvm$ afs infra.start *
   ```

   > **Note:** For steps on stopping and starting a Nutanix cluster, see "Node Management" in the administrative guide for the target hypervisor.

Stopping a Files Cluster

Stop a Files cluster using the listed commands.

About this task

Before you stop a Nutanix cluster that is running Files, first stop the Files cluster (set of file server VMs running on the nodes). To stop a Files cluster, do the following:

Procedure

1. Using SSH, log on to a Controller VM of the Nutanix cluster that is running Files.
2. List file server names.

   ```
   nutanix@cvm$ ncli file-server list
   ```

3. Stop a single file server.

   ```
   nutanix@cvm$ afs infra.stop fs_name
   ```

   Replace *fs_name* with the name of the file server.

SHARE AND EXPORT MANAGEMENT

Shares and exports have a variety of configurations. Review this section before creating a share or export.

You can create file shares (SMB), exports (NFS), or multi-protocol shares for a file server.

You have the option of creating distributed or standard shares and exports. A distributed share or export is the repository for the personal files of a user, and a standard share is the repository shared by a group. A distributed share or export distributes data across all file server VM (FSVM) in a cluster, while standard shares store all data on a single FSVM. You can store files in the root directory of a distributed NFS share but not in the root directory of a distributed SMB share.

To create a share or export, see one of the following.

- Creating a Share (SMB) on page 64
- Creating an Export (NFS) on page 70
- Creating a Multi-Protocol Share or Export on page 78

Limitations

The following limitations apply to shares and exports.

- Distributed shares are only available on deployments of three or more FSVMs.
- Do not use Windows Explorer to create new top-level directories, as you then cannot rename any folders created with the default `New Folder` name (see Troubleshooting on page 156).
- Windows clients do not support using UTF8 encoding for naming files and directories on NFSv3 exports.
- You cannot access top-level directories (TLDs) or shares through the FSVM short name.
- You cannot modify TLDs from Mac clients.
- Refer to "System Limits" in the *Nutanix Files Release Notes* for a list of additional limitations.
- Nutanix does not support mounting NFS shares on ESXi hypervisors.

Permissions

The following default permissions apply on distributed and standard shares.

- **Distributed shares/exports:**
  - Domain administrator: Full access
  - Creator owner: Full access (inherited only)
  - Domain user: Read only
- **Standard shares/exports:**
  - Domain administrator: Full access
  - Domain user: Full access
  - Creator owner: Full access (inherited only)

Creating a Share (SMB)

About this task

Create an SMB share.

> **Note:** Files does not support mounting SMB shares on Linux clients. Use multi-protocol shares instead, see Creating a Multi-Protocol Share or Export on page 78.

Procedure

1. Go to the **Shares** view in the Files Console and click **Create a New Share**.

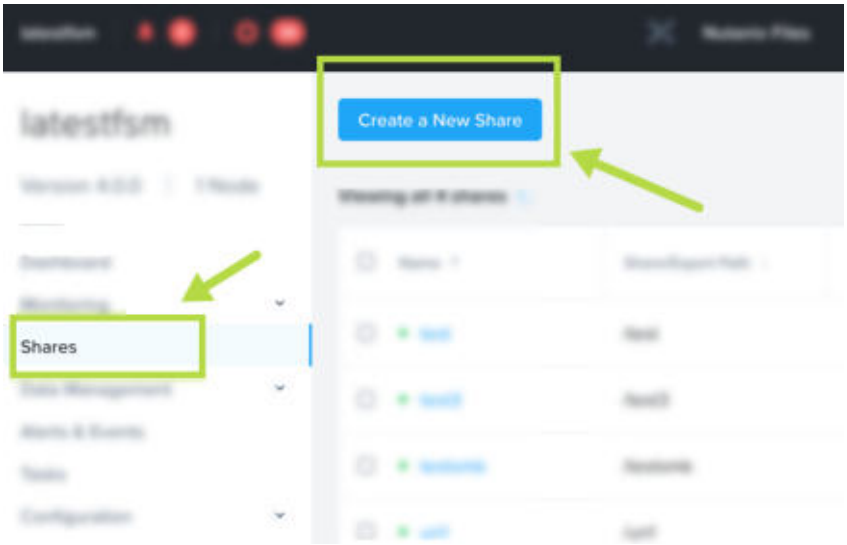


Figure 44:

2. The **Create a Share** window appears and displays the **Basics** tab. Do the following in the indicated fields:

   a. **Name**: Enter a name for the share.

      Note the following naming conventions:

      - Files does not allow the following names:
        - `Global`
        - `Printers`
        - `admin$` (reserved)
        - `ipc$` (reserved)
        - `homes` (reserved)
      - Names are not case-sensitive.
      - Each name must be unique.
      - Files allows unicode characters.
      - Maximum name length is 80 characters.
      - A blank space or space character cannot appear as the first or last character in the name.
      - Names that end with the `$` are hidden shares.

   b. **Description** (optional): Enter a description for the share.

   c. **Share path** (optional): You can create nested shares by specifying the path. Nested shares inherit some properties from the parent directory. See Nested Shares and Exports on page 92 for details.

   d. **Max Size** (optional): Enter the maximum share size in GiB, see "System Limits" in the release notes for details.

      Leaving the field blank means there is no upper limit to the size of the share. Enter a value here if you want to set an upper size limit. Setting a value changes the capacity from the client view.

   e. **Primary Protocol Access**: Select the **SMB (Ideal for Windows Clients)** option.

   f. **Enable multi-protocol access for NFS clients**: Check the box to enable the multi-protocol feature. Follow the steps in Creating a Multi-Protocol Share or Export on page 78 to configure multi-protocol settings. Otherwise, continue to the next step.

      > **Note:** For information on the multi-protocol feature, see Multi-Protocol Support for Files on page 85.

   g. Click **Next**.

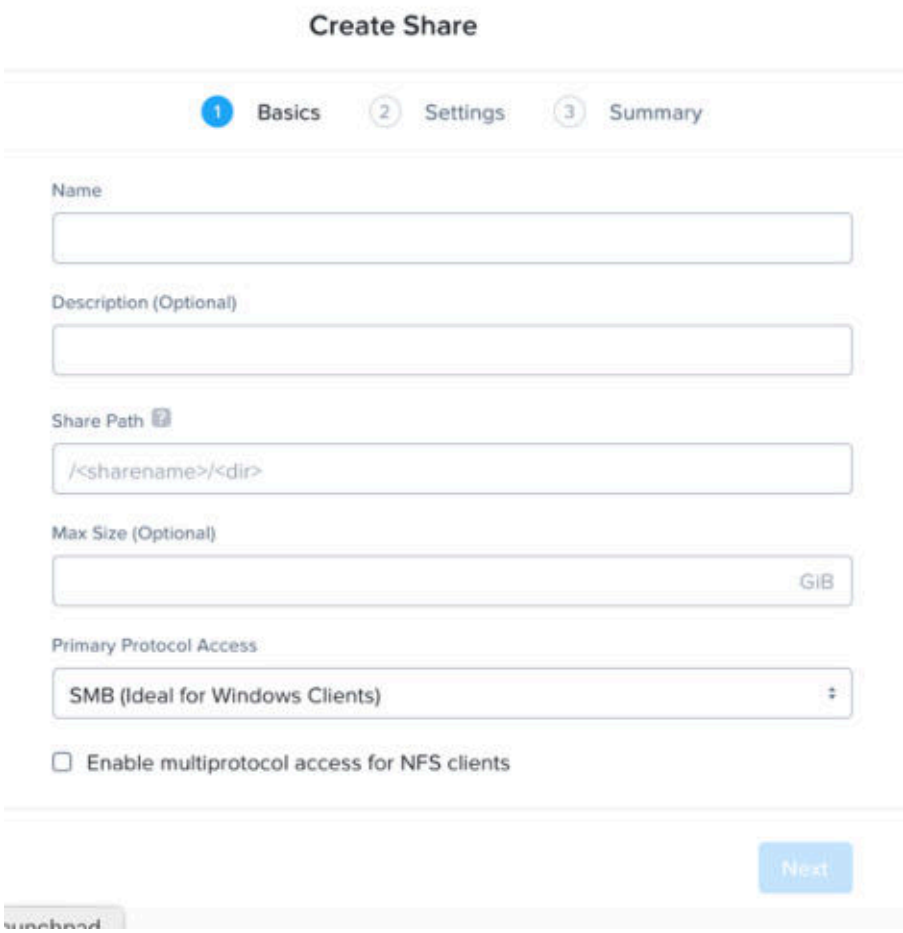


Figure 45: Create a Share: Basics tab

3. In the **Settings** tab, do the following in the indicated fields:

   a. Check the **Use "Distributed" share type instead of "Standard"** to create a distributed share:

   > **Note:** Shares on single-FSVM deployments are standard by default.

   - A *Distributed share*, also known as a home share, load-balances user data across multiple FSVMs by distributing top-level directories.
   - A *Standard share*, also known as a general-purpose share, serves data and connections from a single FSVM.

   A distributed share is frequently used as the repository for the personal files of a user, while a standard share is frequently the repository shared by a group.

   > **Tip:** When using a distributed share for user profiles, facilitate better load distribution by creating a top-level directory (TLD) for each user.

   b. **Enable Self Service Restore**: Check this box to enable snapshots of the share contents, see Self-Service Restore on page 145.

   c. **Enable File System Compression:** Check this box to save space and reduce data on the share through in-line compression of written data, see File System Compression on page 138.

   > **Note:** Enabling compression at the container level makes the **Enable File System Compression** option unavailable. The option does not appear on clusters upgraded from earlier versions that do not support file system compression.

   d. **Blocked File Types**: Check this box to block files with specific character patterns in their names, see File Blocking for more information.
   A field appears for blocked file types. Enter a comma-separated list of character patterns of directories and file names blocked from the share.

   e. **Enable Access Based Enumeration (ABE)**: Check this box to restrict user access when browsing the contents of top-level directories to only those files and folders that they have access permissions for, see Access-Based Enumeration (SMB only) on page 126.

   f. **Encrypt SMB3 Messages:** Check this box to enable message encryption between the file server and client. See Encryption on page 126.

   g. Click **Next**.

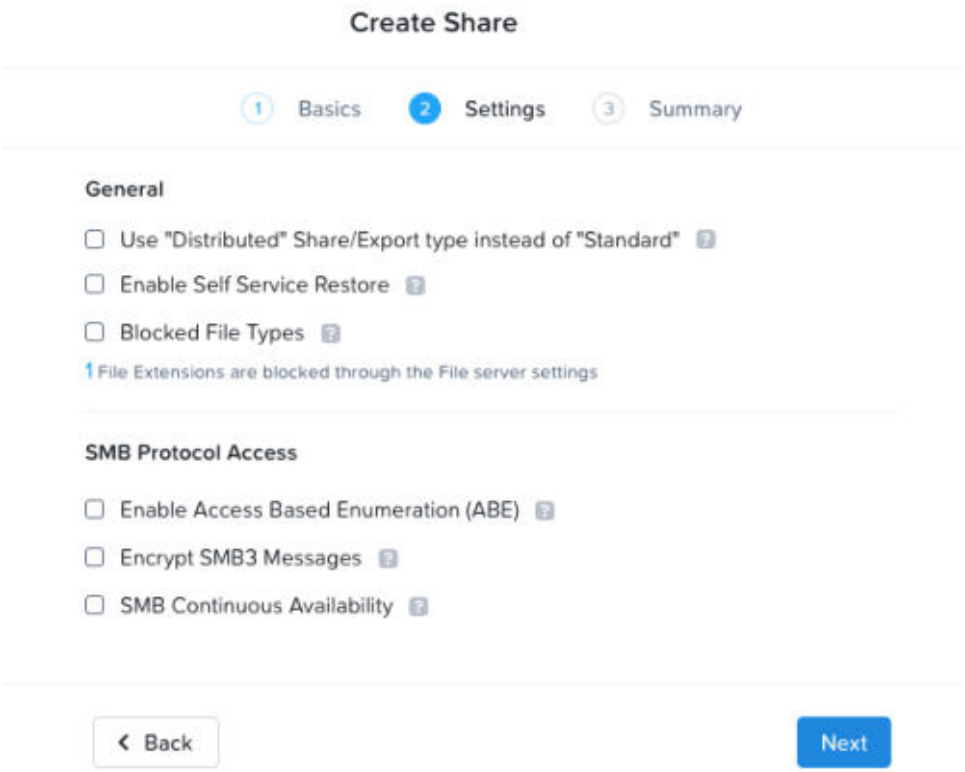


Figure 46: Create a Share: Settings tab

4. In the **Summary** tab, review the configuration of the share in the indicated sections:

   The **Basics** section includes the following details:

   - **Share Name**: The name of the new share.
   - **Share Path**: The file path to the new share.
   - **Primary Protocol**: SMB protocol.
   - **Multi-Protocol Access**: Disabled
   - **Share or Export Type**: Distributed or standard share.
   - **Self Service Restore**: Enabled or disabled.
   - **Share Compression**: Enabled or disabled.
   - **Blocked Files Types**: Enabled or disabled. When enabled, the number of blocked file types appears. Clicking the number displays the file type.

   The **Protocol Access** section includes the following details:

   - **Authentication**: Authentication type (Kerberos, AD, etc).
   - **Client Access**: Read-write.

     > **Note:** For SMB shares, SMB client access is always read-write. Access-control lists (ACLs) specify user and group access permissions.

   - **ABE**: (Access-based enumeration) enabled, disabled, or n/a.
   - **Message Encryption**: Enabled or disabled.

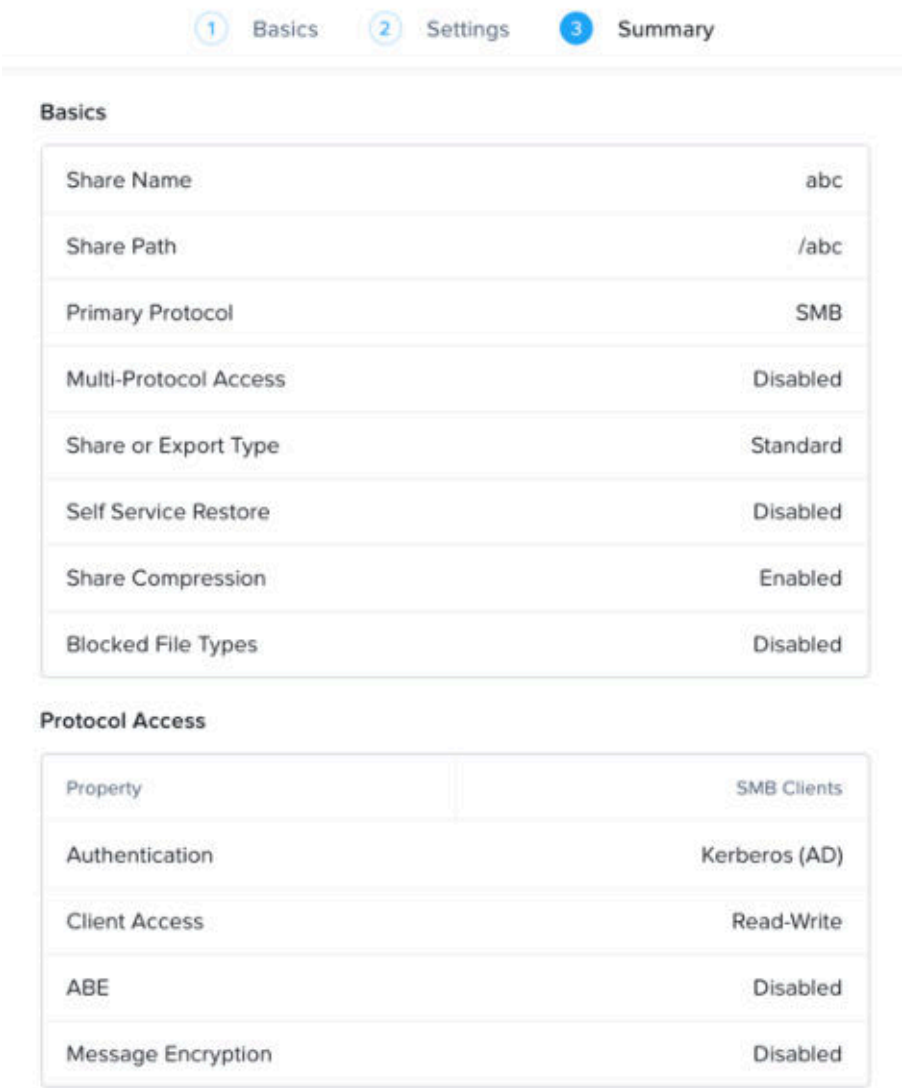


Figure 47: Create a Share: Summary tab

5. When all the information is correct, click the **Create** button.

What to do next

Map the newly created share in your name-space.

Creating an Export (NFS)

Administrators can create exports using the file server dashboard.

About this task

Follow the steps as indicated to create an export (NFS share).

> **Note:** Files does not support mounting NFSv4 exports on Windows clients. Use NFSv3 exports or multi-protocol shares.

Procedure

1. Go to the **Shares** view in the Files Console and click the **Create a New Share**.

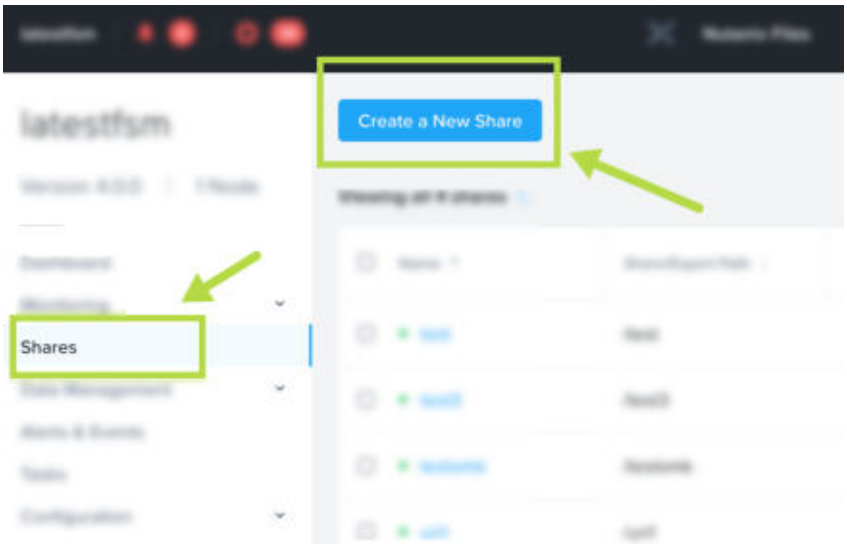


Figure 48: + File Server button

2. The **Create a Share** window appears and displays the **Basics** tab. Do the following in the indicated fields:

   a. **Name**: Enter a name for the share.

      Note the following naming conventions:

      - Files does not allow the following names:
        - `Global`
        - `Printers`
        - `admin$` (reserved)
        - `ipc$` (reserved)
        - `homes` (reserved)
      - Names are not case-sensitive.
      - Each name must be unique.
      - Files allows unicode characters.
      - Maximum name length is 80 characters.
      - A blank space or space character cannot appear as the first or last character in the name.
      - Names that end with the `$` are hidden shares.

   b. **Description** (optional): Enter a description for the share.

   c. **Share path** (optional): You can create standard nested shares by specifying the path. Nested shares inherit some properties from the parent directory. See Nested Shares and Exports on page 92 for details.

   d. **Max Size** (optional): Enter the maximum share size in GiB, see "System Limits" in the release notes for details.

      Leaving the field blank means there is no upper limit to the size of the share. Enter a value here if you want to set an upper size limit. Setting a value changes the capacity from the client view.

   e. **Primary Protocol Access**: Select the **NFS** option.

   f. **Enable multi-protocol access for SMB clients**: Check the box to enable the multi-protocol feature. Follow the steps in Creating a Multi-Protocol Share or Export on page 78 to configure multi-protocol settings. Otherwise, continue to the next step.

      > **Note:** For information on the multi-protocol feature, see Multi-Protocol Support for Files on page 85.

   g. Click **Next**.

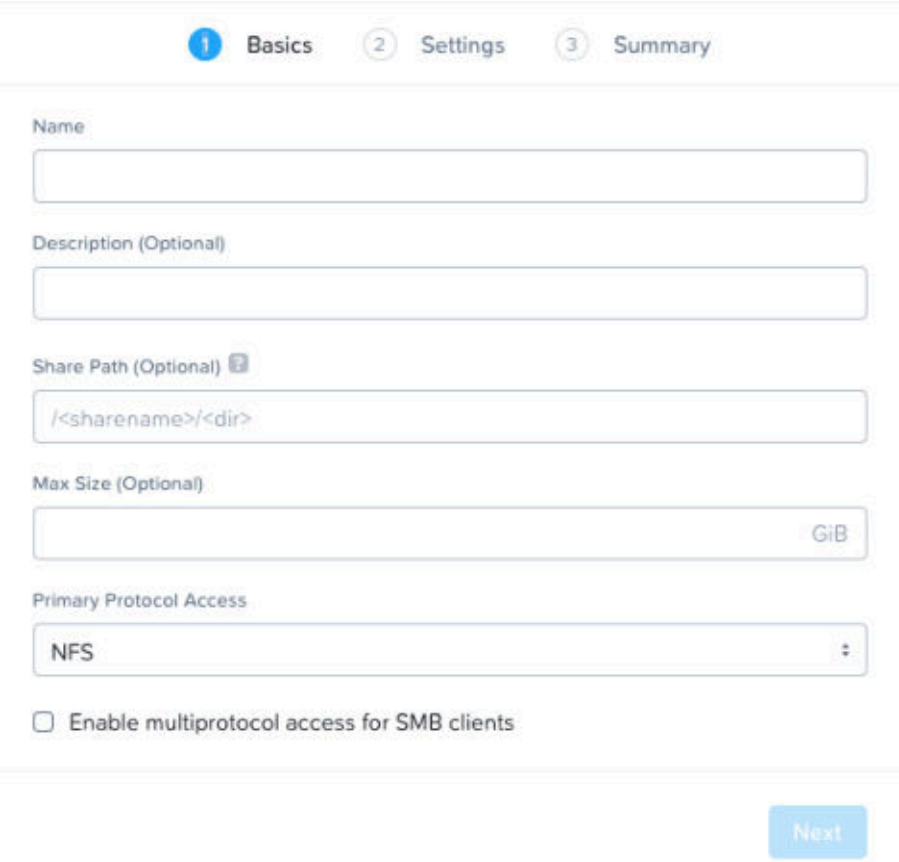


Figure 49: Create a Share: Basics tab

3. In the **Settings** tab, do the following in the indicated fields:

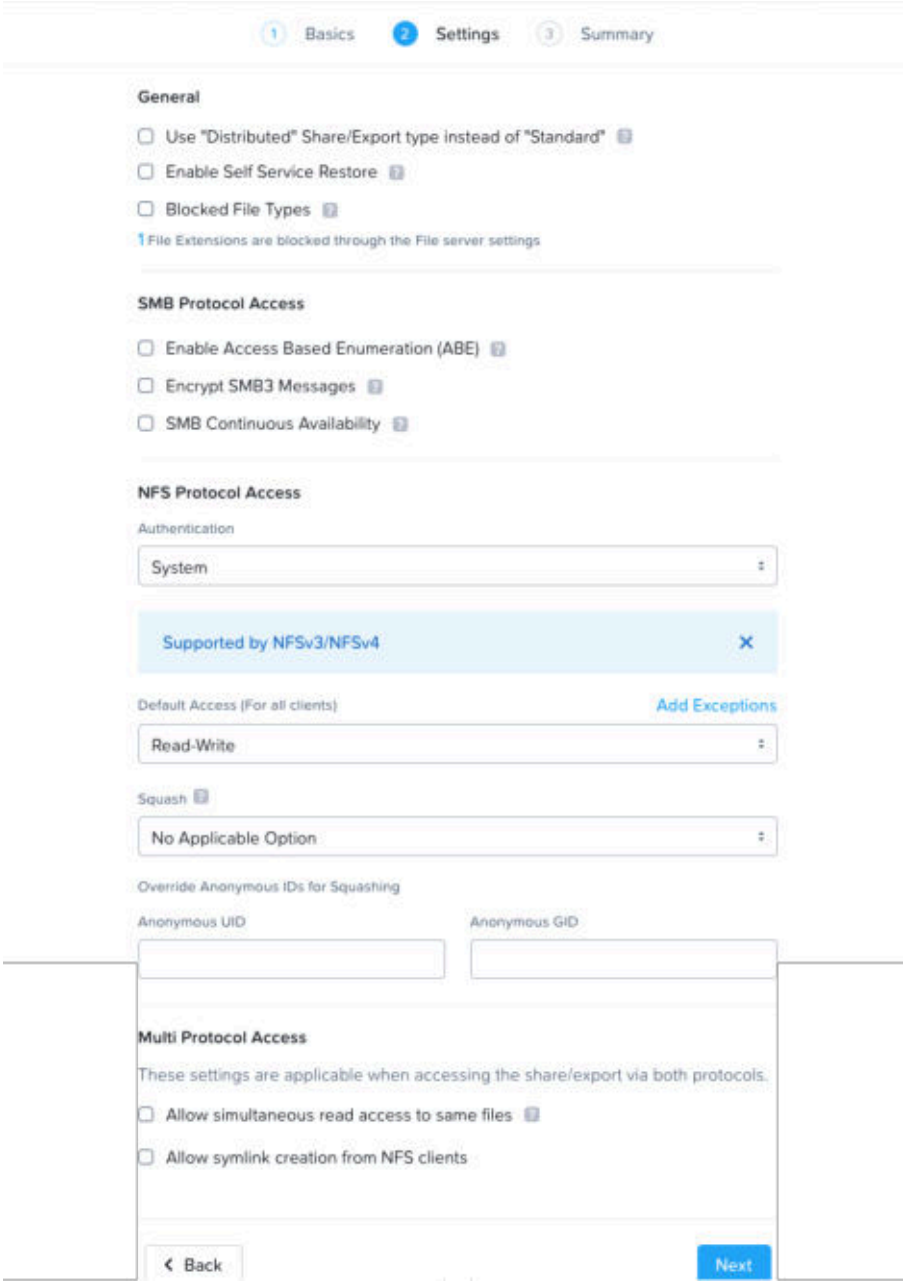


Figure 50: NFS Export Settings Tab

a. Check the **Use "Distributed" share type instead of "Standard"** to create a distributed export:

> **Note:** Exports on single-FSVM deployments are standard by default.

- A *Distributed share*, also known as a home share, load-balances user data across multiple FSVMs by distributing top-level directories.
- A *Standard export*, also known as a general purpose or non-distributed export, serves data and connections from a single FSVM

A distributed export is frequently used as the repository for the personal files of a user, while a standard export is frequently the repository shared by a group.

> **Tip:** When using a distributed export for user profiles, facilitate better load distribution by creating a top-level directory (TLD) for each user.

b. **Enable Self-Service Restore**: Self-service restore lets you restore files from previous snapshots, see Self-Service Restore on page 145.

c. **Enable Compression**: Check this box to save space and reduce data on the share through compression, see File System Compression on page 138.

> **Note:** Enabling compression at the container level makes the **Enable File System Compression** option unavailable. The option does not appear on clusters upgraded from earlier versions that do not support file system compression.

d. **Blocked File Types**: Check this box to block files with specific character patterns in their names, see File Blocking for more information. After checking the box, a field appears for blocked file name patterns. Enter a comma-separated list of character patterns blocked from the export.

> **Note:** Blocking extensions at share level overrides the blocked extensions defined at file server level.

e. **Authentication**: Select the authentication method from the pull-down list.

> **Note:** Files does not support Kerberos 5, Kerberos 5i, and Kerberos 5p with the NFSv3 protocol.

The options are **None**, **System**, **Kerberos 5**, **Kerberos 5i**, and **Kerberos 5p**.

> **Note:** Changes to the authentication method can result in invalid mounts. Resolve the issue by remounting the authentication type as a value for parameter `sec` in the mount command on the client. Example: `"-o sec=krb5"` if you change the export authentication type to Kerberos 5.

f. **Default Access (For all clients)**: Select the default access permissions from one of the following in the pull-down list:

- **Read-Write**: Client can mount the export and write to it.
- **Read-Only**: Client can mount the export and read data but not write to it.
- **No Access**: Client cannot mount the NFS export nor read or write from it.

Use one of the following formats:

- Absolute IPs (for example 129.144.0.0)
- Hostname of client (for example client.domain.com)
- CIDR format for specifying subnets (for example 129.144.0.0/24)
- Netgroups (for example @it_admin)
- Wildcards (for example clients*.domain.com)

> **Note:** Files supports Netgroups to handle client access. Netgroups can limit access to hosts, but Files does not support Netgroups that limit access to users.

g. **+ Add Exceptions** (optional): If you want to refine the read and write permissions (more than just the default), click the **+Add Exceptions** link.

Clicking the **+Add Exceptions** link displays two more fields for the two non-default permission options. For example, if you selected **Read-write** as the default access, fields for **Client with read-only access** and **Clients with no access** appear. In these fields, enter a comma separate list of clients for the specified access permission level. Exceptions take precedence over the default, so any client listed in one of the exception fields gets that level of permission instead of the default permission.

> **Note:** For IPv4 clients, exceptions can be complete IP addresses, wildcards, or subnets. For IPv6 clients, Files only accepts complete IP addresses as exceptions.

> **Note:** If you add or remove a client from a Netgroup, it reflects on the client after 30 minutes.

> **CAUTION:** If a client IP address matches more than one exception rule, the client may experience access issues.

h. **Squash**: Select the squash value from the pull-down list. The squash option controls the access privileges of root client users (users with UID 0). The **None** value gives root users super-user access privileges to the export, letting them create, edit, and delete files from any user. The **Root Squash** value disables super user access and maps all roots users to anonymous users. The **All Squash** value maps all users to anonymous ones.

i. **Override anonymous IDs for squashing**: check this box to enter anonymous user identifier (UID) and group identifier (GID) value other than the default value. Both the anonymous UID and GID default have a default value of -2.

- **Anonymous UID**: Enter the anonymous user identifier value for the export.
- **Anonymous GID**: Enter the anonymous group identifier value for the export.

You can map regular users to anonymous ones to restrict access to the NFS export. Setting the NFS authentication type to **None** maps all client users to anonymous ones.

4. Click **Next**.

5. In the **Summary** tab, you will see the following details in the indicated fields:

   The **Basics** section includes the following details:

   - **Share Name**: The name of the new export.
   - **Share Path**: The file path to the new export.
   - **Primary Protocol**: NFS protocol.
   - **Multi-Protocol Access**: Disabled
   - **Share or Export Type**: Distributed or standard share.
   - **Self Service Restore**: Enabled or disabled.
   - **Share Compression**: Enabled or disabled.
   - **Blocked Files Types**: Enabled or disabled. When enabled, the number of blocked file types appears. Clicking the number displays the file type.

   The **Protocol Access** section includes the following details:

   - **Authentication**: Authentication type (System or none).
   - **Client Access**: Read-write or read-only.
   - **Squash**: All squash, root squash, or none.

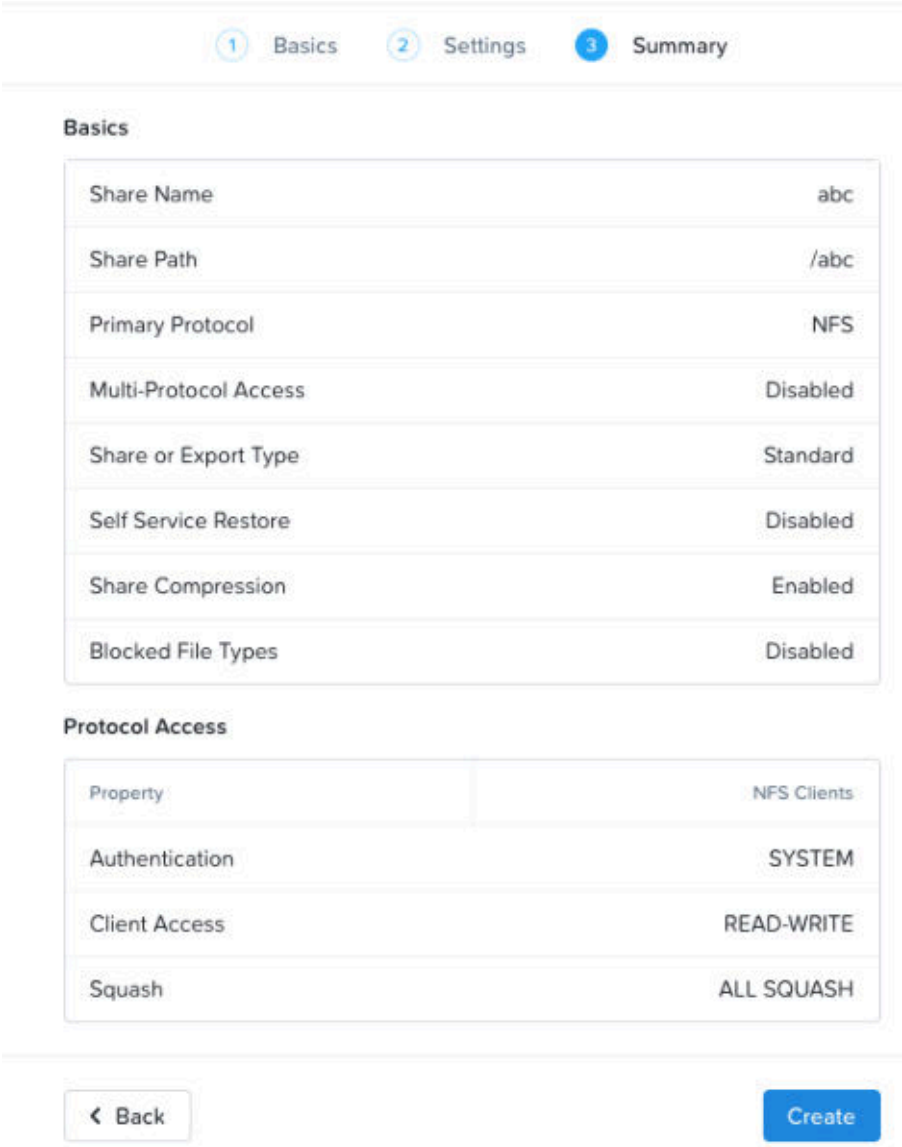


Figure 51: Create an Export: Summary tab

6. When all the information is correct, click the **Create** button.

Creating a Multi-Protocol Share or Export

Create a multi-protocol share or export.

About this task

For information on the multi-protocol feature, see Multi-Protocol Support for Files on page 85. Perform the following steps as indicated to create a multi-protocol a share or export.

Procedure

1. In the **Basics** tab, do the following in the indicated fields:

   a. **Name**: Enter a name for the share or export.

      Note the following naming conventions:

      - Files does not allow the following names:
        - `Global`
        - `Printers`
        - `admin$` (reserved)
        - `ipc$` (reserved)
        - `homes` (reserved)
      - Names are not case-sensitive.
      - Each name must be unique.
      - Files allows Unicode characters.
      - Maximum name length is 80 characters.
      - You cannot use a blank space or space character as the first or last character in the name.
      - Names that end with the `$` are hidden shares.

   b. **Description** (optional): Enter a description for the share or export.

   c. **Share path** (optional): You can create standard nested shares by specifying the path. Nested shares inherit some properties from the parent directory. See Nested Shares and Exports on page 92 for details.

   d. **Max Size** (optional): Enter the maximum size in GiB, see "System Limits" in the release notes for details.

      Leaving the field blank means that there is no upper limit to the size of the share or export. Enter a value here if you want to set an upper size limit. After setting the limit it cannot be increased.

   e. **Primary Protocol Access**: Choose **SMB (Ideal for Windows Clients)** or **NFS** as the primary protocol.

   f. **Enable multi-protocol access**: Check the box to enable the multi-protocol feature.

   g. Click the **Next** button.

2. In the **Settings** tab, do the following in the indicated fields:

   a. Check the **Use "Distributed" share type instead of "Standard"** to create a distributed export:

   > **Note:** Exports on single-FSVM deployments are standard by default.

   - A *distributed export*, also known as a sharded export or home share, load balances user data across multiple FSVMs.
   - A*standard export*, also known as a general-purpose share or export, for any other purposes.

   b. **Enable Self Service Restore**: Self-service restore lets you restore files from previous snapshots, see Self-Service Restore on page 145.

   c. **Enable File System Compression:** Check this box to save space and reduce data on the share through inline compression of written data, see File System Compression on page 138.

   > **Note:** Enabling compression at the container level makes the **Enable File System Compression** option unavailable. The option does not appear on clusters upgraded from earlier versions without file system compression support.

   d. **Blocked File Types:** Check this box to specify blocked file types, see File Blocking on page 127.

   e. Do the following in the **SMB Protocol Access** section.

   - **Enable Access Based Enumeration (ABE)** (SMB only) enable ABE by checking this box. ABE restricts access when browsing the contents of the top-level directories (TLDs) to only those files and folders that you have access permission for, see Access-Based Enumeration (SMB only) on page 126.
   - **Encrypt SMB3 Messages:** Check this box to enable message encryption between the file server and client. See Encryption on page 126.

3. In the **NFS protocol** section, do the following.

   a. **Authentication**: Select the authentication method from the pull-down list.

   The options are **None**, **System**, **Kerberos 5**, **Kerberos 5i**, and **Kerberos 5p**.

   > **Note:** Files does not support Kerberos 5, Kerberos 5i, and Kerberos 5p with the NFSv3 protocol.

   > **Note:** Changes to the authentication method can result in invalid mounts. Resolve the issue by remounting the authentication type as a value for parameter `sec` in the mount command on the client. Example: `"-o sec=krb5"` if you change the export authentication type to Kerberos 5.

   b. **Default Access (For All Clients)**: Select the default access permissions from one of the following options in the pull-down list:

   - **Read-Write**: Clients can mount the export and write to it.
   - **Read-Only**: Clients can mount the export and read data but not write to it.
   - **No Access**: Clients cannot mount the NFS export nor read or write from it.

   > **Note:** Files supports Netgroups to handle client access. Netgroups can limit access to hosts, but Files does not support Netgroups that limit access to users.

   c. If you want to refine the permissions (more than just the default), add exceptions in the provided fields.

   For example, if you selected **Read-Write** as the default, lines for **Read-Only** and **No Access** appear. In these fields, enter a comma separate list of clients for that access permission level. Exceptions take precedence

over the default, so any client listed in one of the exception fields gets that level of permission instead of the default permission.

Use one of the following formats:

- Absolute IPs (for example 129.144.0.0)
- Hostname of client (for example client.domain.com)
- CIDR format for specifying subnets (for example 129.144.0.0/24)
- Netgroups (for example @it_admin)
- Wildcards (for example clients*.domain.com)

> **Note:** If you add or remove a computer from a Netgroup, it reflects on the client after 30 minutes.

d. **Squash**: Select the squash value from the pull-down list. The squash option controls the access privileges of root client users (users with UID 0). The **None** value gives root users super-user access privileges to the export, letting them create, edit, and delete files from any user. The **Root Squash** value disables super user

access and maps all roots users to anonymous users. The **All Squash** value maps all users to anonymous ones.

e. **Override anonymous IDs for squashing**: check this box to enter anonymous user identifier (UID) and group identifier (GID) value other than the default value. Both the anonymous UID and GID default have a default value of -2.

   - **Anonymous UID**: Enter the anonymous user identifier value for the export.
   - **Anonymous GID**: Enter the anonymous group identifier value for the export.

   You can map regular users to anonymous ones to restrict access to the NFS export. Setting the NFS authentication type to **None** maps all client users to anonymous ones.

f. In the **Multi Protocol Access** section, check the box for the settings that you would like to enable:

   - **Allow simultaneous read access to the same files**: Clients from either protocol can perform simultaneous reads.
   - **Allow symlink creation from NFS clients**: This option only appears if SMB is the primary protocol.

     Note: On SMB clients, NFS client-created symlinks appear as regular objects (files or directories).

g. Click **Next**.

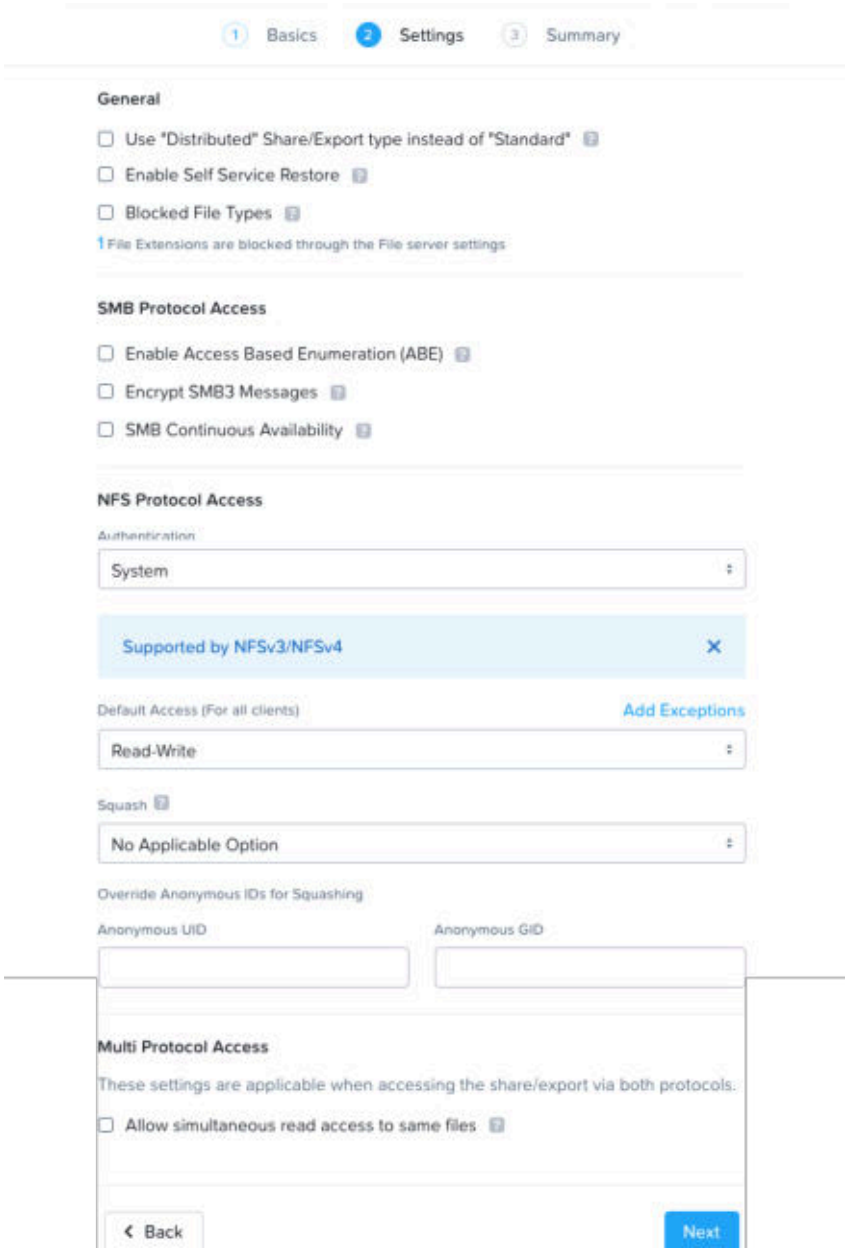


Figure 52: Settings Tab

4. In the **Summary** tab, review the configuration of the share in the indicated sections:
   The following details appear in the **Basics** section:

   - **Share Name**: The name of the new export.
   - **Share Path**: The file path to the new export.
   - **Primary Protocol**: NFS protocol.
   - **Multi-Protocol Access**: Enabled.
   - **Share or Export Type**: Distributed or standard share.
   - **Self Service Restore**: Enabled or disabled.
   - **Share Compression**: Enabled or disabled.
   - **Blocked Files Types**: Enabled or disabled.

   The **Protocol Access** section includes the following details:

   - **Authentication**: Authentication type (Kerberos, AD, and so on).
   - **Client Access**: Read-write.

     > **Note:** For SMB shares, SMB client access is always read-write. Access-control lists (ACLs) specify user and group access permissions.

   - **ABE**: (Access-based enumeration) enabled, disabled, or n/a.
   - **Message Encryption**: Enabled or disabled.

   The **Multi Protocol Access** section includes the following details:

   - **Allow simultaneous read access to same files**: Disabled or enabled.

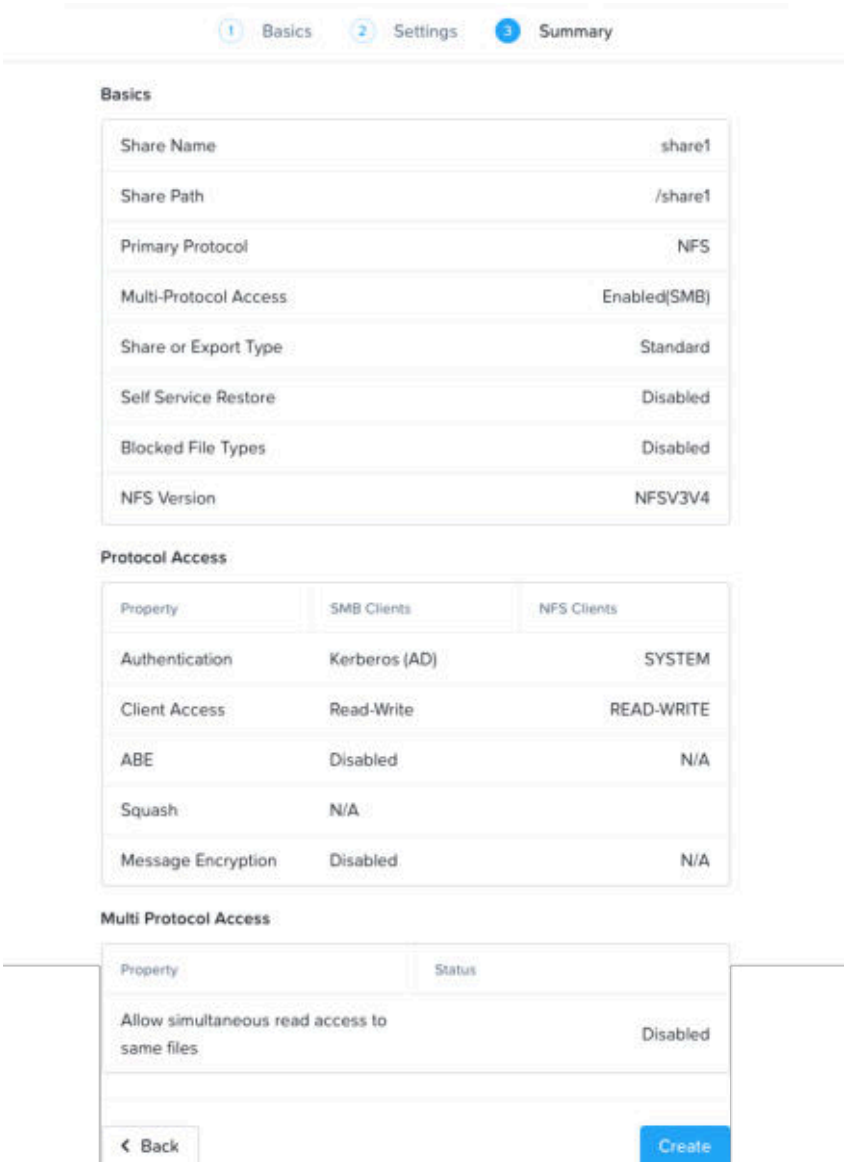


Figure 53: Create an Export: Summary tab

5. When all the information is correct, click the **Create** button.

What to do next

To configure user mapping between NFS and SMB users, see User Mapping on page 108.

Multi-Protocol Support for Files

Multi-protocol support for Files provides SMB and NFS clients read-and-write access to the same share or export.

When you choose a native protocol for the share, SMB or NFS, the protocol determines access-control options for that share or export. You can then manage all access-control options using the native protocol. Once you create a share or export, you cannot change the native protocol.

Consider the following protocol-specific options and caveats before deciding on a primary protocol.

Authentication

SMB clients have only one authentication type, Active Directory (AD), that they can use to access exports. NFS clients can access SMB shares using AD, LDAP, system, or unmanaged protocols. Files uses the native protocol to authenticate all non-native shares.

Authorization

Files uses user mapping to map IDs and user or group names, which authorizes access to shares and exports from native and non-native protocols. Configure user mapping when the NFS authentication of a multi-protocol share is not Kerberos (when NFS uses Kerberos, the users and groups for NFS and SMB are the same).

An IT administrator must configure user mapping, see User Mapping on page 108.

Access-Based Enumeration

Files supports access based enumeration (ABE) for native and non-native SMB shares.

Access Control Lists (ACLs)

ACLs are stored in the format of the native protocol. Modify ACLs using the native protocol. NFS clients do not honor share-level ACLs. When you create an object using a non-native protocol client, the object inherits access control details from the parent folder.

Alternative Data Stream (ADS)

Files supports alternative data streams (ADS) for native and non-native SMB shares.

Antivirus

Files supports antivirus protection on native and non-native SMB shares.

If you access a file that has a virus through an NFS client-created symlink or hard-link, only the path of the first access appears in the quarantine table.

Audit

Files supports audits for native and non-native shares.

Backup and Restore

Files supports back-up and restore only using the native protocol.

> **CAUTION:** Backup using a non-native client can result in metadata and ADS loss.

Change Notifications

There is no directory change notification support across protocols. Namespace changes made to a share from one protocol take up to a minute to reflect for a client of the second protocol.

Concurrent Access

Files supports concurrent read-access but not concurrent write-access for multi-protocol shares.

High-Availability

Files equips multi-protocol enabled shares and exports with high-availability. Files does not honor the NFS grace period on SMB clients.

Namespace Compatibility

Since all SMB lookups are case insensitive, and all NFS lookups are case-sensitive, SMB clients can encounter multiple objects with the same name string but different cases.

NFS clients can create multiple files on SMB shares that use the same name-string but different cases. This leads to issues for Windows clients that find files with the same name but in different cases. Files maps NFS objects with illegal or special characters in the namespace to a new SMB name.

Quota

Files supports quota, see Quotas on page 117.

Stats

Files combines stats for native and non-native shares.

Symlinks

NFS clients can create symlinks on native and non-native NFS shares and follow them using a native or non-native protocol.

Limitations

- Files silently ignores permission change requests for non-native protocol clients.
- NFS file or directory names that include trailing spaces are treated as illegal characters on SMB namespaces.
- Native SMB shares cannot enable the NFS squash option.
- NFS users might not see ownership details on native SMB shares. For NFS users, default owner and group on native SMB shares appears as `root/root`. However, SMB users see the default owner and group as `BUILTIN administrators`.
- Template mapping ignores the domain of users and groups.
- Files does not support multi-protocol compatibility for legacy NFS exports.
- When SMB clients create share objects on non-native shares, the objects derive the mode information from their parent share.
- If SMB is the primary protocol, you can only apply a quota to SMB users.

Modifying a Share or Export

Modify or delete a share or an export.

About this task

You can update or delete a share from the **Shares** tab in the Files Console. Note the following limitations for modifying a share or export:

- Clusters created using earlier versions of Files might not support some new features.
- When renaming a share or export, access to the old name remains if you maintain the existing connection. Close and reopen the browser or CLI to no longer see the old name.

> **Note:** See Explicit Paths for Shares and Exports for a list of modifiable features on shares and exports with explicit paths.

Procedure

- In the Files Console, select the target entry in the share table of the **Shares** tab.

- In the row for the selected entry, click **three dots menu** following.
  - » Click **Delete**, to delete the share (you do not need to proceed to the next steps).
  - » Click **Update** to modify the settings of the share or export. Proceed to the next steps for directions.
  - » Click **Add Quota Policy** to add a user or group quota to the share, see Quotas on page 117.
- To update the share, follow the steps as indicated.

  a. In the **Basics** or **Settings** tab, update one or more values accordingly.

  > **Note:** When renaming a share or export, access to the old name remains as long as the existing connection is maintained. Close and reopen the browser or CLI to no longer see the old name.

  See the Creating a Share (SMB) on page 64, Creating an Export (NFS) on page 70, or Creating a Multi-Protocol Share or Export on page 78 for details about each field.

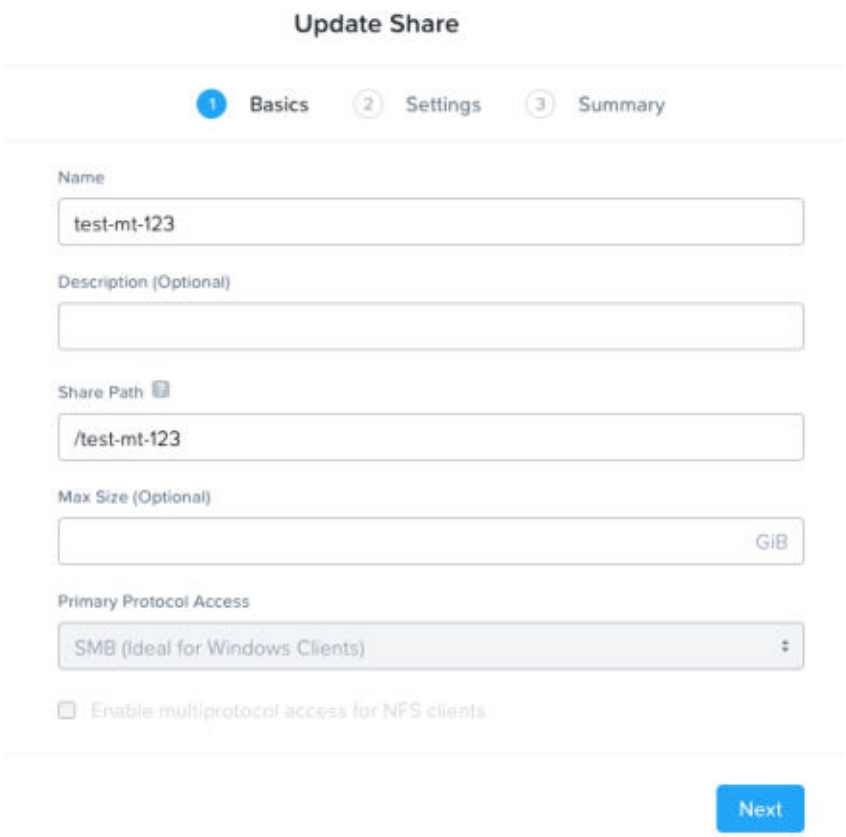


Figure 54: Update Share Window

  b. When all the values are correct, click **Update** (**Summary** tab).

Deleting a Share or Export

Delete a share or an export through the Files Console.

About this task

> **Warning:** Deleting a share or export is permanent, and the deleted share or export cannot be recovered.

To delete a share or export, do the following:

Procedure

1. Remove the share or export contents and disconnect all clients. (All directories and files within the share/export should be deleted.)
2. In the Files Console, select the target entry in the share table of the **Share** tab.
3. In the row for the selected entry, click **three dots menu** > **Delete**.

> **CAUTION:** When deleting a share or export with an explicit path, the data and storage is not removed, see Nested Shares and Exports on page 92.

> **Note:** All shares and exports with explicit paths must be deleted before deleting the parent share.

Accessing Home Shares

Access home directories without requiring the physical distributed share or export in the profile location.

Before you begin

Confirm the following:

- Home directories are top-level directories in the distributed share
- Home directories have the required permissions

About this task

Home directories are a type of distributed share. Files supports the following CLI commands for FSVMs at the distributed share location. Files automatically enables distributed share support. If the user profile exists in one of the distributed shares, access the distributed share directly using the universal naming convention (UNC) path.

> **Note:** If your home directory exists in multiple distributed shares, Files matches to the first share created chronologically.

Distributed shares appear with other shares in Files when enumerating shares on the file server. Files enables distributed home share support by default.

Access home directories using the UNC path `\\Files_server\SamAccountName` (instead of `\\Files_server\home_share\SamAccountName`).

Procedure

1. To enable distributed home share support:

```
nutanix@fsvm$ afs smb.set_conf “enable user homes” yes section=global
```

2. To enable dollar user distributed share access.

```
nutanix@fsvm$ afs smb.set_conf “template user homeshare” “%U$” section=global
```

Accessing User Home Shares (Advanced)

This topic outlines advanced options for accessing home directories without requiring the physical distributed share with the profile.

About this task

Before you begin

Ensure that you have met the requirements, see Accessing Home Shares on page 89.

Procedure

To configure preferred home shares.

```
nutanix@fsvm$ afs smb.set_conf “preferred home shares” “home-share1, home-share2”
 section=global
```

Replace `home-share1` and `home-share2` with the names of the homes shares.

Continuously Available Shares (SMB Only)

Persistent file handles facilitate continuous availability (CA) for SMB shares.

During a disruption of service, SMB shares remain continuously available by using the persistent file handles and the high-availability features. Persistent file handles reduce the period of data unavailability by automatically reconnecting users to the file service in-use before the disruption.

Files stores open lock information on persistent volume groups available of every FSVM in the cluster. During a fail-over, the fail-over FSVM can access the volume group. Persistent file handles let clients wait to reconnect without notifying the user about the connection failure. As a result, disruptions are seamless to users and applications.

> **Note:** During a lock protection interval, Files does not prevent NFS clients from accessing the file. Files only prevents SMB clients from accessing files during the disconnected state.

> **Note:** The periodic handler cleans up stale locks for the disconnected persistent file handles every 24 hours.

Synchronous writes on CA shares may impact performance. Nutanix recommends continuous availability for shares with less metadata and more I/O intense workloads, such as VHD or VHDx based profile disks, where performance impact is minimal.

Enabling Continuous Availability

Enable continuous availability (CA) on an SMB share.

About this task

Follow the steps as indicated.

Procedure

1. To confirm the status of continuous availability, replace the `share-name` and check the share profile.

   ```
   nutanix@fsvm$ afs share.list sharename=share-name
   ```

2. Replace the `share-name` and enable continuous availability on the share.

   ```
   nutanix@fsvm$ afs share.edit share-name continuous_availability=true
   ```

What to do next

Check the status of continuous availability by repeating the command in step 1.

Connected Shares

Connect standard or distributed shares in the namespace of another standard or distributed share.

Connecting shares creates a unique, continuous namespace. Use connected shares to distribute data across multiple FSVMs from a specified directory.

Distributed shares let you only shard data on the top-level directory (TLD) level. To shard data from a lower-level directory, connect shares by submounting a distributed share onto the path of another share or export. Files distributes that data from the mount point across multiple FSVMs. The distribution of data achieves load balancing.

The connected, continuous namespace lets you have different settings within the namespace, such as directory-level quotas. For more details on directory-level quotas, see Setting Directory-Level Quotas on page 96. Rather than connecting existing shares, you can also create a new share with an explicit path, see Nested Shares and Exports on page 92.

> **Note:** For multi-protocol enabled shares, a submount point can apply to multiple directories due to the case-sensitive NFS behavior and case-insensitive SMB behavior. As a result, Files uses all directories with the same name but different cases as submount points.

Connected Share Limitations

The following limitations apply to custom-level sharding:

- The submount path cannot point to the root of a parent share.
- The submount path cannot contain snapshots.
- The submount path on a parent share must be empty before configuring submounts.
- Both parent and child shares must use the same type of NFS authentication.
- Nutanix does not allow nesting submounts. For example, if you have a child share `child1` submounted on the submount path `/parent/dir1/dir2/child1`, you cannot submount `child2` anywhere on the path.
- Nested shares and submounted child shares cannot have the same directory hierarchy.
- Stale locks for disconnected persistent file handles get cleaned up every 24 hours.

Connecting a Share

Shard a standard share or export.

About this task

Submount a distributed share onto a standard share to connect shares.

Procedure

1. Replace *`submount-path`* with the file path to the mount point from the standard share. Replace *`child-share-name`*.

   ```
   nutanix@fsvm$ afs share.edit child-share-name submount_path=submount-path
   ```

2. List the connected shares.

   ```
   nutanix@fsvm$ afs share.list_all_submounted_shares
   ```

Disconnecting Shares

Disconnect connected shares or exports.

About this task

To disconnect a child share from a parent share, update the mount point of the child share.

Procedure

1. Replace *child-share-name* with the name of the child share.

```
nutanix@fsvm$ afs share.edit child-share-name submount_path=''
```

2. List all child shares connected to a parent share.

```
nutanix@fsvm$ afs share.list_submounted_shares parent-share-name
```

> **Note:** Due to caching issues, child shares and exports can still appear as mounted.

What to do next

For NFS, remount the parent share on the client.

Nested Shares and Exports

Use nested shares and exports to hide share paths or restrict access to parts of a tree.

Unless specified, Files creates shares and exports with an implicit path (the `/share-name`). To create a share or export with a different path, use the **Share Path** field to specify an explicit path for the share. Specifying an explicit path creates a nested share or export. The explicit path for the nested share must begin with the parent share or export (for example, to create an explicit path for `share1` you can specify the explicit path `/share1/dir1/dir2`). Child shares and exports are nested under the parent shares.

To connect existing shares, see Connected Shares on page 90.

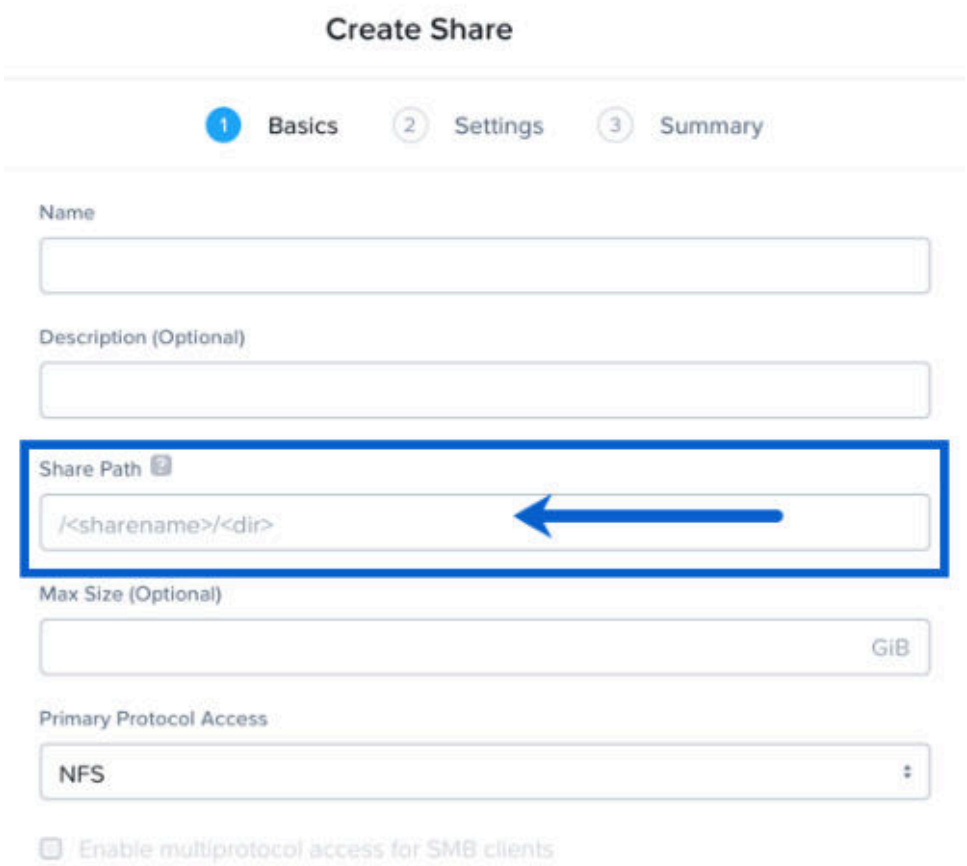


Figure 55: Shares and Exports with Explicit Paths

The following rules and limitations apply for creating, deleting, or modifying shares and exports with explicit paths:

- New nested shares and exports cannot start with the name of another share or export created using an explicit path.
- Explicit paths must start with the forward-slash (/) character.
- When you create a share or export, an explicit path must contain a parent share or parent export name and the directory path (do not use the root directory).
- Files does not support changed file tracking (CFT) backup for nested shares and exports.
- When you enable or disable multi-protocol on a parent share, the action reflects for all nested shares or exports for that parent.
- Nested shares and exports inherit both the primary and secondary protocols, which you cannot modify at the nested-share level.
- The share path is case-sensitive for the NFS protocol.
- When you delete any of the directories in an explicit path, the path becomes inaccessible.
- Namespace changes under the root of the nested share do not appear immediately when accessed through the parent share path.

Consider the following recommendations:

- Delete all nested shares and exports before deleting the parent share.
- When a path becomes inaccessible after the deletion of a directory in an explicit path, you can recover the path by creating a directory with the same name.

Feature Modification

You can modify some of the properties of nested shares. Most of the time you cannot modify inherited properties. Refer to the following table for modification options.

Table 18: Nested Share Modification Options

Feature	Default State	Modification
SMB access based enumeration (ABE)	Disabled	Enabled
NFS advanced settings	Enabled	Enabled
Add quota policy	Disabled	Disabled
AV settings	Inherited from parent	Disabled

Durable SMB File Handles

Durable handles facilitate seamless SMB client reconnection.

Files enables durable handles for SMB shares automatically. Durable handles let SMB clients survive a temporary client-side connection loss after opening a file, allowing transparent client reconnection within a timeout. Files does not support durable file handles for directories or alternate data stream (ADS) files.

Durable handle reconnection times increase when a previous session, with multiple durable handles, must flush data and metadata before re-establishing the connection.

Enabling SMB Symlinks

Enable symlink support over the SMB protocol.

About this task

A symbolic link (also known as a symlink or a soft link) is a file type that contains a path to a target file or directory.

Although symlink support is enabled by default, Files requires enabling the following link configurations on Windows clients to use symlinks on SMB shares:

- Remote-to-remote
- Remote-to-local

> **Note:** By default, remote to local and remote to remote symlinks are disabled on Windows clients. To enable, symlink access on Windows clients, use the following command.
>
> ```
> > fsutil behavior set SymlinkEvaluation  R2R:1  R2L:1
> ```
>
> SMB clients must have symlink access enabled to access symlinks and follow the targets.

Files does not support the following functionalities for symlinks:

- Directory symlinks at the root of distributed shares.
- Alternate data streams (ADS).
- Durable or persistent file handles.
- Opportunistic locks (oplocks) and leases.
- There is no interoperability support between SMB and NFS on Multi-protocol shares for symlinks with absolute targets.

> **Note:** Hycu does not support symlinks. Commvault does not support cross-share symlinks.

Procedure

1. Disable symlink support on the file server.

   ```
   nutanix@fsvm$ afs smb.set_conf "enable smb symlinks" "False" section=global
   ```

   Disabling SMB links only prevents the creation of new symlinks. Existing symlinks remain accessible.

2. Enable symlink support on the file server.

   ```
   nutanix@fsvm$ afs smb.set_conf "enable smb symlinks" "True" section=global
   ```

   Disabling SMB links only prevents the creation of new symlinks. Existing symlinks remain accessible.

Managing Limited Local Users (SMB Only)

Add limited local users to your file server.

About this task

You can add limited local SMB users to a file server for SMB share access using the command line or using the Microsoft Management Console (MMC) Local Users and Groups snap-in.

If you add a limited local user using the MMC Local Users and Groups snap-in, MMC adds the user to the built-in Local Users group. Users in the built-in Local Users group inherit default permissions from the group, refer to Microsoft documentation for steps on adding local users using the MMC.

Manually add users to a group with the desired permissions, or browse and assign permissions using the Local Users and Groups snap-in.

Limited local user management includes the following limitations:

- Files permits a maximum of 100 limited local users per file server.
- Limited local users cannot access non-native shares.
- Files does not support user-mapping for limited local users.
- Files does not support quota for limited local users.
- Files does not support renaming local users.
- Files does not support limited local users not joined to a domain or standalone fileservers. The file server must join an AD domain.

> **Attention:** Limited local users is an SMB-only feature.

The the commands as indicated to manage limited local users.

Procedure

- To add a limited local user, use the following command:

```
nutanix@fsvm$ afs user.add limited-localuser-name
```

- To add a limited local user to a limited local group, use the following command:

```
nutanix@fsvm$ sudo net sam addmem 'group-name' 'file-server-name\user-name'
```

> **Note:** You can add local users to local groups, which are in `BUILTIN\Users` and `BUILTIN\Administrators`.

- To list limited local users, use the following command:

```
nutanix@fsvm$ afs user.list
```

- To edit a limited local user, use the following command:

```
nutanix@fsvm$ afs user.edit limited-localuser-name
```

- To delete a limited local user, use the following command:

```
nutanix@fsvm$ afs user.delete limited-localuser-name
```

Configuring Backup for Distributed Shares

Specify a backup server for distributed shares and exports.

About this task

Third-party solutions do not automatically back up distributed shares. Follow the steps as indicate to update the backup configuration syntax and allow the backup of distributed shares and exports.

Procedure

1. Check the global backup configuration.

```
nutanix@fsvm$ afs smb.get_conf "backup hosts"
```

2. Update the backup configuration.

   » Add a backup server by replacing *backup-server-ip* with the IP address for the backup server.

   ```
   nutanix@fsvm$ afs smb.set_conf "backup hosts" backup-server-ip section=global
   ```

   » Remove a backup server.

   ```
   nutanix@fsvm$ afs smb.del_conf "backup hosts" section=global
   ```

3. Validate changes to the backup configuration.

   ```
   nutanix@fsvm$ afs smb.get_conf "backup hosts"
   ```

Setting Directory-Level Quotas

Set a directory-level quota by connecting shares.

About this task

Files supports size limits for shares and exports. Set a directory quota by submounting a share or export onto the path of a standard share.

Procedure

1. Create a share or export and specify the max size, see Creating a Share (SMB) on page 64.

   Alternatively, you can apply a share size limit to an existing share or export, see Modifying a Share or Export on page 87.

2. Connect the share or export to a directory of a standard share or export, see Connecting a Share on page 91. Files applies the quota (max size) for the connected share or export to the directory.

Share Migration

Share migration lets you copy data from a share on a third-party file server source to a share on the target Files file server.

> **Note:** Nutanix recommends running share migration from a static share.

Each time you initiate share migration, Files starts an iteration of the migration copy process. If some files fail to migrate, you can copy the failed files to the target using the migration retry procedure, see Re-attempt Share Migration (SMB Only) on page 98.

> **Note:** Once the number of failed files exceeds a threshold, the migration retry procedure cannot succeed, and you must run another iteration of the migration.

If the source share is not static and the data changes either during or after the migration, run another iteration to copy the changed data from the source to the target. If you run the iteration on a static share, Files shows that zero entities (files, folders, streams) were copied. To ensure the migration of all data, run multiple iterations.

Limitations

The following limitations apply to share migration.

- Files does not migrate hard-links from the source.

Migrating Shares (SMB Only)

Migrate shares from a third-party vendor to Nutanix Files.

Procedure

Follow the steps as indicated to migrate shares.

1. Provide source file server details.

```
nutanix@fsvm$ afs migration.source_add source_alias=source_server_alias /
 source_fs_fqdn=source_server_IP_address_or_FQDN source_user='domain\\user_name'
```

> **Note:** The file server must be able to access the source user domain.

Files asks you to provide the source user password.

2. If you have not already, create a target share.

> **Note:** You can create a target share using the Files UI.

```
nutanix@fsvm$ afs share.add share_name protocol=SMB share_type=Distributed\Standard
```

3. Check that Files registered the source.

```
nutanix@fsvm$ afs migration.source_list
```

4. Create the migration plan.

```
nutanix@fsvm$ afs migration.plan_create name=plan_name source_alias=source_server_alias /
source_share=share/path target_share=target_share  target_share_path=path/within/share /
migrate=true|false
```

> **Note:** The default state of the migrate attribute is false. Setting the attribute to true automatically triggers the migration.

> **Note:** Share migration requires `target_share_path` when the target is a directory within the target share.

(optional) Add more arguments.

```
json_output=         Render output in json format


total_bytes=         The total number of Bytes to be migrated.

total_files=         The total number of files to be migrated.
```

> **Tip:** Setting the `total_bytes` or `total_files` arguments lets Files display the progress of the migration.

5. Start the migration.

```
nutanix@fsvm$ afs migration.plan_migrate name=plan_name
```

```
allow_unknown_sids=          Copy entities with an unknown owner or group SID.
                               When set to false, Filesdoes not copy entities with an
                               unknown user or group SID.
ignore_files_at_root=         Ignore files at the root of the source. If set to
                               false, Files fails the migration when the root of the
                               source share path has files.(For distributed shares
                               only.
ignore_long_tld_names=        Ignore TLD names with more than 255 bytes on source. If
                               set to false, Files fails migration when the name of a
                               TLD at the root of the source share is longer than 255
```

```
                                bytes long. (For distributed shares only.)
```

> **Note:** You can rerun the migration multiple times after completion.

6. Check the status of the migration.

   Setting verbose to true provides extra migration status details for every FSVM on the file server.

```
nutanix@fsvm$ afs migration.plan_status name=plan_name verbose=true|false
```

> **Note:** Checking the status of a migration plan returns a granular status view of all nodes on the file server. For any node excluded from the migration plan, Files reports a `succeeded` status.

7. (optional) Use one or more of the following commands as needed.

```
migration.plan_summary                  Display a summary of each iteration of the plan.
migration.plan_pause                    Pause the migration.
migration.plan_resume                   Resume the migration.
migration.plan_cancel                   Cancel the migration.
migration.source_update                 Update the source.
migration.source_delete                 Delete the source.
migration.source_list                   List configured sources.
```

Re-attempt Share Migration (SMB Only)

Retry share migration for files that failed to migrate.

About this task

Before retrying share migration for failed files, Nutanix recommends running the share migration operation multiple times to ensure the migration of all source data.

Follow the steps as indicated.

Procedure

1. (optional) Get a list of files that failed to migrate in the last iteration of the plan.

```
nutanix@fsvm$ afs migration.plan_get_failed_files_list name=plan_name /
out_file=output_directory_path [iter_index=-1]
```

2. Retry the migration for the files that failed to migrate in the last iteration of the plan.

```
migration.plan_retry_failed_files name=plan_name
```

DIRECTORY SERVICE AND DOMAIN MANAGEMENT

Administration of directory services and file server domains.

The domain and directory service configuration is an essential step of file server management. Join a domain and update domain name system (DNS) entries during or after file server creation. The this chapter describes domain and directory service management options and operations.

The directory services configuration specifies the primary and, if applicable, secondary protocol of a file server. If you intend to create multi-protocol shares and exports on file servers with secondary protocols, configure user mapping, see User Mapping on page 108 and Multi-Protocol Support for Files on page 85.

Joining a Domain

Join a file server to a domain.

About this task

When joining to an Active Directory domain, Files adds the following entries to the Active Directory.

- Machine account with the Files cluster name
- SPN entries

Join a domain either during or after file server creation. To join a domain during file server creation, see Creating a File Server on page 36. To join to a domain after file server creation, follow the instructions:

> **Note:**
>
> - If the file server time and the Network Time Protocol (NTP) time are not in sync, Prism Element (PE) raises an alert. Be sure to sync the file server time and NTP time before joining to a domain.
> - The join domain process can be unresponsive if the Active Directory computer objects exist from previous unresponsive join attempts. Before attempting to join to the domain, remove the previous Files computer objects from the Active Directory. Prism Element (PE) also prompts you to overwrite existing computer accounts.
> - Do not remove the file server computer object in AD, as that can cause disruptions to file-server services.

Procedure

1. In the Files Console, go to **Configuration** > **Authentication**.
2. In the **Directory Services** tab, select one or more protocols you want to use and complete the fields for the selected protocols.

   You can select SMB, NFS, or both. Each protocol contains more fields to complete, including fields about Active Directory. See the **Directory Services** tab step in Creating a File Server on page 36 for detailed instructions.

   You must have an administrator account or the following required permissions for the relevant AD organizational unit (OU) realm:

   - Create computer objects
   - Read/Write service principal names

3. When all the fields are correct, click **Update**.

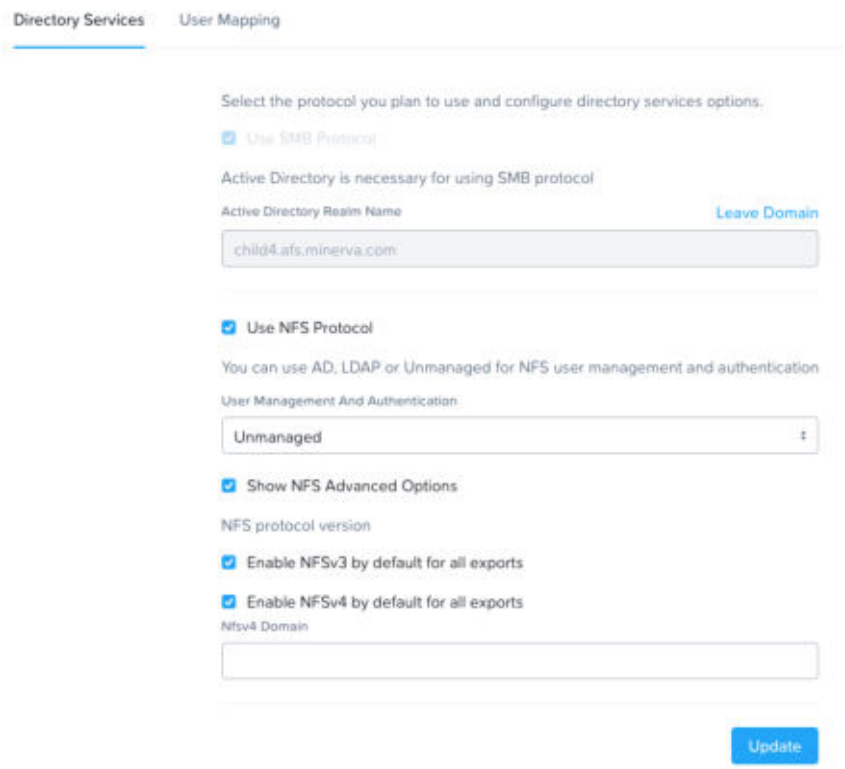


Figure 56: Directory Services Window

Leaving a Domain

Disconnect your file server from a domain by following the steps described.

About this task

The following tasks require that you first remove (leave) the connection between the file server and its current domain:

- Changing the file server name
- Joining the file server to a different domain
- Deleting the file server

Leaving a domain deletes service principal name (SPN) entries and Files computer accounts created by the file server on the Active Directory. To leave a domain, do the following:

> **CAUTION:** Leaving a domain causes Files to stop services until you join the file server to a domain. During this time, existing and future clients cannot use shares, exports, or permissions.

Procedure

1. In the Files Console, go to **Configuration** > **Authentication**.

2. In the **Directory Services** tab, click the **Leave Domain** link.

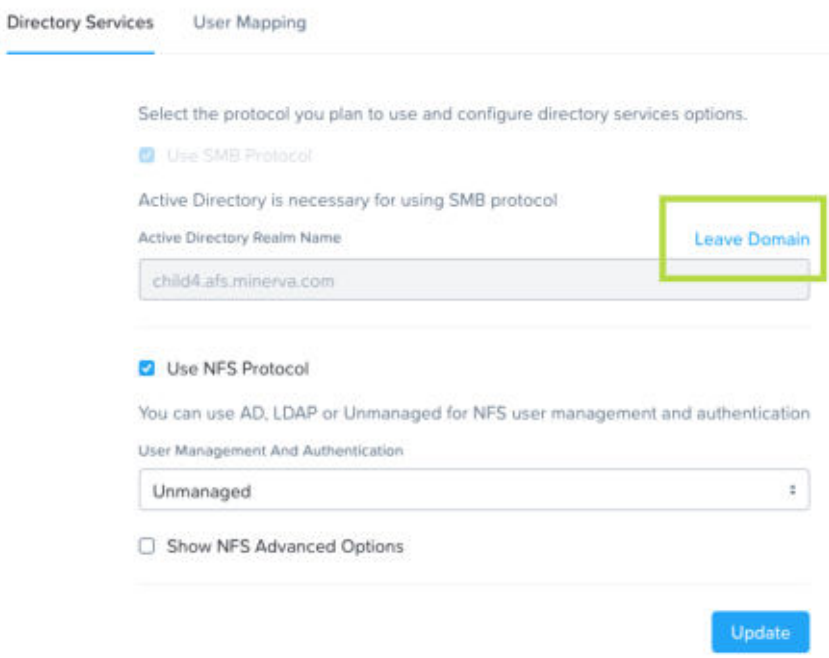


Figure 57: File Server Basics Window

Clicking the **Leave Domain** link removes the file server connection to the domain. After changing the file server name or domain, do the following to re-establish the file server connection to a new domain:

- Remove any old DNS entries from one or more DNS servers.
- Add the new DNS entries (see Updating Domain Name System (DNS) Entries on page 101).
- Update the directory services (see Updating Directory Services on page 105).

Updating Domain Name System (DNS) Entries

Modify or add to existing domain name system (DNS) entries.

About this task

Administrators can choose to update or add more DNS entries for the following scenarios:

- Creating a file server
- Changing the filer server name or target domain
- Removing a node
- Adding a node

Files adds entries to the DNS server to ensure the availability of file services to clients using a Files namespace. Files uses DNS entries to resolve DFS referrals.

The total number of DNS entries added to the specified DNS server is three times the number of nodes in the Files cluster. Use the following format for DNS entries:

- **File server name for each FSVM IP address**: `file_server_name.domain_fqdn`

- **FSVM name for its IP address** : `fsvm_name.domain_fqdn`.

To update the DNS configuration, do the following:

Procedure

1. In the Files Console (see Files Console on page 17), go to **Configuration** > **Update DNS Entries**.
2. To automatically configure DNS entries, select the **Automatic (MS-DNS only)** option and do the following in the indicated fields:

> **Note:**
>
> - The automatic option is available only when using Microsoft DNS. If you use a different DNS server or you see a message that Files did not find the entries on the DNS server, use the manual method.
> - Files can only add PTR entries for the /24 zone. Check the subnet of the external network to confirm the zone. Files does not add PTR records automatically for external networks that are part of a large DNS zone. For these networks, follow the manual process instead.

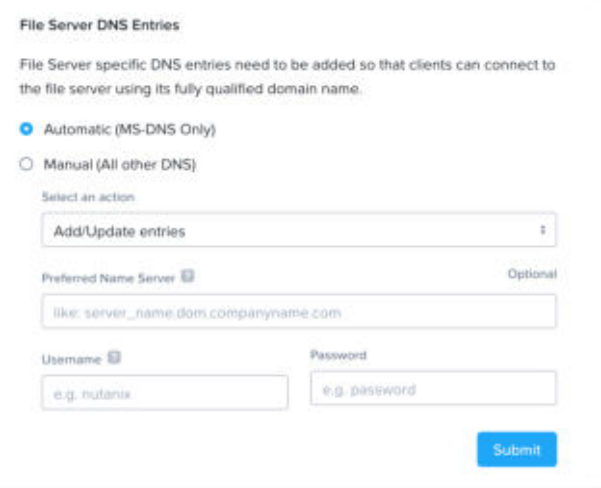


Figure 58: DNS Window (automatic view)

a. **Select an action**: Select **Add/Update entries** or **Delete Entries** from the pull-down menu.

   Selecting **Add/Update entries** adds (or updates as needed) the required file server entries to the DNS server automatically. (You do not need to enter them yourself.) Selecting **Delete Entries** automatically deletes the current file server entries from the DNS server.

b. **Preferred Name Server** (optional): Enter a preferred name server.

c. **Username**: Enter the user name of the DNS server administrator.

d. **Password**: Enter the administrator password.

e. Click **Submit**.

3. To manually configure DNS entries, click the **Manual (All other DNS)** option and do the following:

   A table appears that lists entries consisting of a fully qualified host name and the corresponding IP address/PTR record.

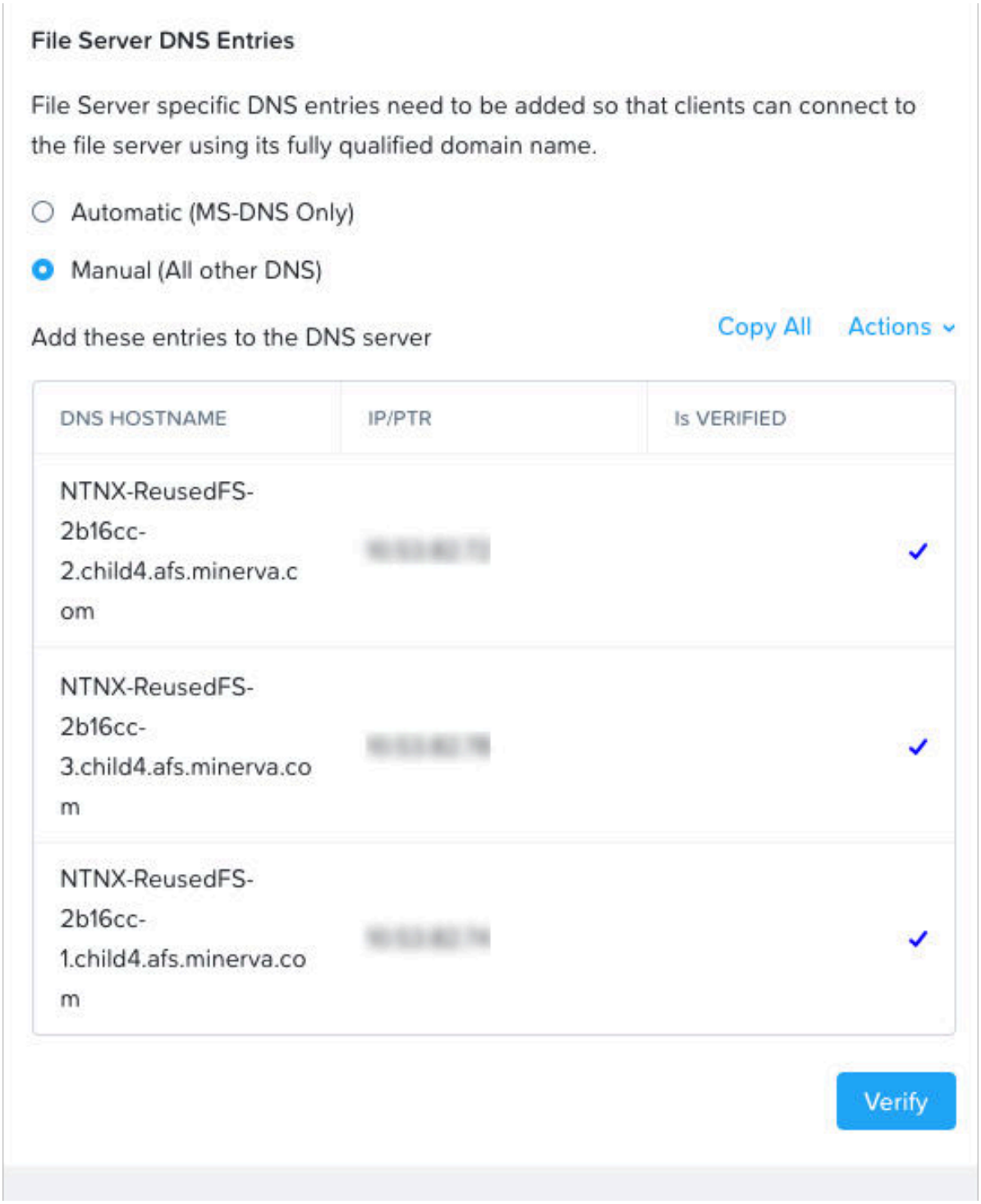


Figure 59: DNS Window (manual view)

a. **Note:** For file server name, add A records only. For FSVM names, add both the A records and PTR records.

Copy the entries from the table and add them to the DNS server.

You can either copy the entries manually or click the **Copy All** link to copy the entire table. Click **Actions** to download that table in the CSV or JSON format. (To delete the current file servers entries on the DNS server, use this list to determine which entries to delete.)

b. Click the **Verify** button.

After copying the entries to the DNS server, verify that they are correct. Clicking the **Verify** button checks each entry in the table against the DNS server entries. A check mark in the **Is Verified** column indicates that entry is present on the DNS server. There should be a check mark for every entry.

c. Repeat step (a) for any entries that do not have a check mark in the **Is Verified** column.

The most likely reason (other than a network issue) for a missing check mark is a copy-and-paste error when transferring that entry to the DNS server. Check your work and repeat this process until Files verifies all entries.

Disjoint Domains

Configure different DNS domain and active directory (AD) realm values through the disjoint domains feature.

Your deployment must have the following requirements prior to configuring disjoint domains:

- A running AOS cluster
- A DNS domain name
- An AD realm name

Configuring Disjoint Domains

To have different DNS domain and AD realms, configure disjoint domains.

About this task

Follow the steps as indicated after creating a file server.

Procedure

1. After creating the file server configure the protocols and join the domain.

```
nutanix@fsvm$ afs fs.configure_name_services AD-realm-name username protocol
```

"SMB", "NFS", "NFS,SMB", "SMB,NFS" and "None" are valid arguments for *Protocol*.

(Optional) use one of following instead:

```
nutanix@fsvm$ afs fs.configure_name_services AD-realm-name username protocol
 organizational_unit=organizational unit
```

```
nutanix@fsvm$ afs fs.configure_name_services AD-realm-name username protocol overwrite=true/
false
```

2. Configure the Microsoft DNS, see Updating Domain Name System Entries. If the DNS domain name and AD realm names are not the same, provide the full user principal name (UPN) of an administrator or user.

Updating Directory Services

Update the authentication and user management configuration of a file server.

About this task

To update the directory services configuration, do the following:

Procedure

1. In the Files Console (see Files Console on page 17), go to **Configuration** > **Authentication**.
2. In the **Directory Services** window, select the protocols you want to use and complete the fields for the selected protocols.

   You can select SMB, NFS, or both. Each protocol contains more fields to complete. See the **Directory Services** tab step in Creating a File Server on page 36 for detailed instructions.
3. When all the fields are correct, click the **Update** button.

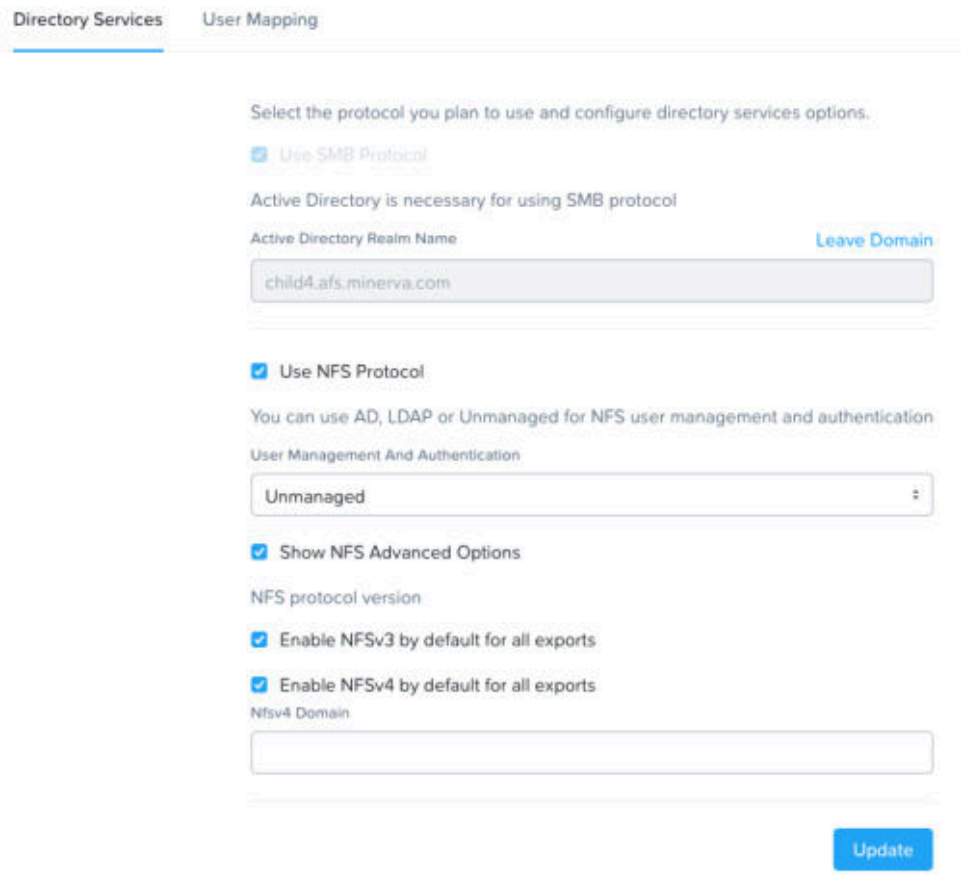


Figure 60: Directory Services Window

Setting AD Machine Account Password Expiry

Set a 30-day expiry period for the Active Directory (AD) machine account password.

About this task

The Files AD machine account password expiry is 0 days by default. Follow the steps as indicated to change the password expiry to 30 days.

Procedure

1. Get the AD machine account password.

```
nutanix@FSVM:~$ afs smb.get_conf "machine password timeout"
[global]
```

2. Update the password expiry period to 30 days.

```
nutanix@FSVM:~$ afs smb.set_conf "machine password timeout" 30 section=global
```

Authentication

This topic provides an overview of unmanaged, LDAP, and Active Directory (AD) authentication for file servers.

Depending on the primary and secondary protocols configured for your file server, access and authentication configuration vary. File servers using SMB protocol must configure Active Directory (AD). File servers using the NFS protocol can use Lightweight Directory Access Protocol (LDAP), AD, or unmanaged authentication services.

LDAP

When you configure LDAP as the authentication service, Files uses Netgroups to control client access. The Files LDAP setup is based on the RFC 2307 standard.

To configure LDAP, assign an LDAP server and a base distinguished name (DN) in the **Directory Services** tab during file server creation. The base DN specifies where the search for users begins. The file server binds to the LDAP server. For added security, set up a bind DN for the LDAP server. Without the bind DN, Files binds to the LDAP server as anonymous.

Active Directory

Files supports AD with NFS and SMB. To set up Active Directory, specify the AD realm name, username, and password. If you have multiple domain controllers, you can configure a preferred domain controller (PDC). If you do not configure a PDC, Files uses any available domain controller.

Using the RFC 2307 standard is an optional setting in AD.

Authorization

The authorization setup on native shares and exports determines user and group read-and-write access permissions.

By default, all domain users and domain admin have full access to standard shares and exports. For distributed shares and exports, domain admins have full access but domain users have read-only access.

> **Note:** Files does not support namespace operations on top-level directories (TLDs) from Mac clients. You can manage TLDs from Windows clients using the Microsoft Management Console (MMC) Snap-In for Nutanix Files.

Native SMB Shares

You can only make access permissions changes using the native protocol. New Technology File System (NTFS) permissions indicate user and group authorization on native SMB shares. Share access-control lists (ACLs) indicate SMB permissions at the root of the share. Share ACLs on SMB shares only apply to SMB clients.

To access multi-protocol enabled shares using a non-native protocol, configure user mapping, see User Mapping on page 108.

You can use the Windows Microsoft Management Console (MMC) Shared Folders plug-in to manage share ACLs on SMB shares. Use the MMC Snap-In for Nutanix Files to modify NTFS ACLs for native SMB shares. On multi-protocol enabled shares, Files performs permission checks based on the native ACLs.

Native NFS Exports

Unix mode bits, set from NFS clients, enforce permissions on NFS exports. By default, the NFS export root directory sets sticky bits, which are permission bits that allow only object owners and root users to remove objects. As an object owner, you can change the permissions on the files and directories that you own. By default, super users can only change export permissions after disabling root squashing.

Files does not support NFS ACLs.

USER MANAGEMENT

User and group administration.

This chapter describes options and steps for managing users and groups on the file server, including their roles, quotas, and mapping across protocols.

The directory services configuration specifies the primary and, if applicable, secondary protocol of a file server. If you intend to create multi-protocol shares and exports on file servers with secondary protocols, configure user mapping, see User Mapping on page 108 and Multi-Protocol Support for Files on page 85.

User Mapping

Configure user mapping on file servers that have multi-protocol shares or exports.

User mapping lets you access the same share or export using native and non-native protocols. You can retain your user-mapping configuration while configuring the directory services. User-mapping configurations are on a file-server level, which extend to all shares and exports on the file server.

> **Note:** If the NFS security type is not Kerberos, you must configure user mapping for multi-protocol shares and exports.

The following restrictions apply to directory service use across protocols:

- NFS clients accessing SMB shares can use IDs from AD, LDAP, or client-generated unmanaged IDs .
- If the NFS security type is not Kerberos, you must configure user mapping for multi-protocol shares and exports
- LDAP usernames cannot be numeric.

> **Note:** Group identifiers (GID) and user identifiers (UID) can appear mismatched because of the access point. The first part of the UID/GID is a config-based range. The last part of the UID/GID is the relative ID (RID) of the user, which is based on the SID. Clients and file servers use different config ranges, so the first part of the GID/UID can appear mismatched.

Mapping Behavior

Files user the following mapping behaviors:

- *UserGroup*: Default mapping for new shares and exports created with Files 3.6.1 or later. Files maps a non-native user to a native user. Files ignores all groups of the non-native user and uses groups of the native user for authorization. Files does not use the groups of the non-native user in the access token.
- *MappedGroups*: Files maps the non-native user and the associated groups to a native user and the respective native groups. The native groups can also be groups of the native user. The access token has a set of user and group identities. Use MappedGroups for the following use cases:
  - Deny access to a specific group.
  - Give access to users based on groups in the ACLs of a file.
- *Legacy mapping*: Mapping for legacy multi-protocol shares and exports (shares and exports on file servers created with versions before Files 3.6.1). Legacy non-native SMB shares use the MappedGroups mapping behavior, and legacy non-native NFS exports use the UserGroup mapping mechanism. You cannot modify the legacy-mapping behavior.

> **Note:** With legacy mapping, Files requires group mapping for the primary group of SMB users to access native NFS exports.

Mapping Configurations

User mapping includes the following mapping configurations:

- **Search**: Use the **Search** tab to search for mapping rules of a user or a group.
- **Rule-Based Mapping**: Use the **Rule-Based Mapping** tab to configure a mapping rule for AD and LDAP users. The following options apply:
  - SMB name matches NFS user name.
  - No template mapping.
- **Explicit Mapping**: Explicit mapping overrides rule-based mapping. The rule entered first takes precedence for users and groups that have multiple mappings configured. Explicit mapping consists of two mapping subcategories: a **one-to-one mapping** list and **wildcard based mapping**.

  You can use the one-to-one mapping list to manually enter or upload a `csv` file that maps users across protocols. Use wildcards for many-to-one mapping. Do not use wildcards on both ends of a user-mapping entry. You can also deny share and export access for a specific user or group.
- **Default Mapping**: Default mapping is the simplest mapping method to configure. You can map all non-native SMB users and groups to a specific native NFS user or group, and conversely. Default mapping also specifies how to handle users that have not had user mapping configured.

The **Summary** tab shows the configured mapping rules and the order in which the rules are prioritized. User-mapping rules take priority in the following order:

1. Deny access rules
2. One-to-one mapping
3. Wildcard mapping
4. Template mapping
5. Default mapping

Configuring User Mapping

Administer non-native user access for multi-protocol shares and exports.

About this task

For information on multi-protocol support, see Multi-Protocol Support for Files on page 85.

> **Note:** User mapping does not support the user name format (UPN).

Procedure

1. In the **Files Console**, go to **Configuration** > **Authentication** > **User Mapping**.
2. (optional) search for mapping rules for a specific user or group.
   a. Select from one of the following search options:
      » **SMB to NFS mapping**
      » **NFS to SMB mapping**
   b. From the dropdown, choose to search by **User** or **Group**.
   c. In the search bar, enter the search target.

3. (optional) In the **Explicit Mapping** section, click **Configure** to set up explicit mapping rules.
   If you have already configured rules, click **Edit**.

   a. Configure explicit mapping by specifying one-to-one mapping, wildcard mapping, and deny access rules.

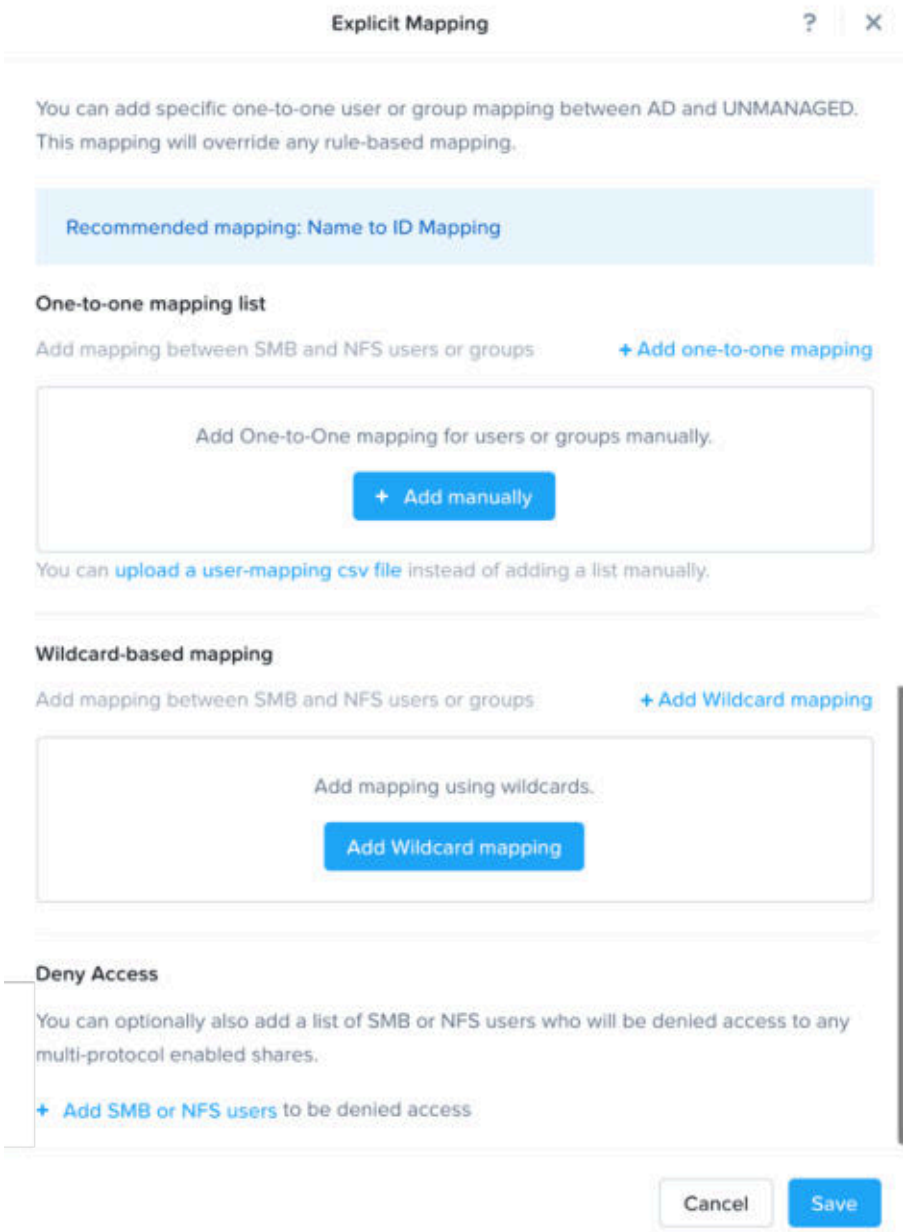


Figure 61: Explicit Mapping

   b. (optional) Configure one-to-one mapping to map single users or single groups by doing one of the following:

   - To map users or groups manually, click **Add one-to-one mapping**. Add the following information in the indicated fields:

     1. In the **SMB Name** field, enter the name of an SMB user or group.
     2. In the **NFS ID**, enter the name of an NFS user or group.
     3. In the **User/Group** field, indicate if the mapping is between users or groups.

4. To add the add one-to-one mapping rule, click the **check icon**.

- Click **upload a user-mapping csv file** to upload a file that specifies one-to-one mapping rules. Format the CSV file to consist of three columns with the indicated information, in the following order:

  - Name of an SMB user or group
  - Name of an NFS user or group
  - Indication of whether the mapping is for a user or group

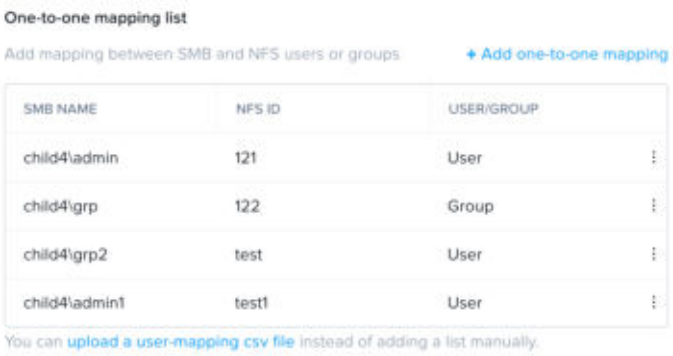


Figure 62: One-to-One Mapping List

c. (optional) Add wildcard mapping to map multiple users or groups to one.

> **Note:**
>
> - Files does not support user mapping entries that have wildcards on both ends.
> - Wildcards do not support the use of numeric NFS IDs.

- 1. Click **Add wildcard mapping**.
  2. In the **Priority** field, choose the priority for the rule. (The lower the number, the higher the piority).
  3. In the **SMB Name** field, enter the SMB user or group name.
  4. In the **NFS Name** field, enter the name of the NFS user or group.
  5. In the **User/Group**field, indicate if the mapping is between users or groups.
  6. To add the wildcard mapping rule, click the **check icon**.

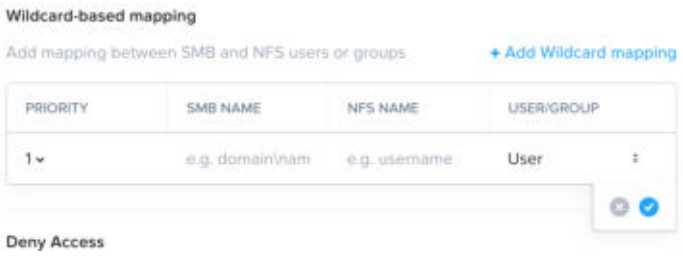


Figure 63: Wildcard Mapping

d. (optional) Add a list of users or groups to be denied access.

- 1. In the **Deny Access** section, click **+Add SMB or NFS users**.

> **Note:** Deny access rules take the highest priority.

2. To deny access, add comma-separated users and groups in one or more of the following fields:

   - **SMB users to be denied access to NFS exports**
   - **SMB groups to be denied access to NFS exports**
   - **NFS users to be denied Access to SMB shares**
   - **NFS groups to be denied access to SMB shares**

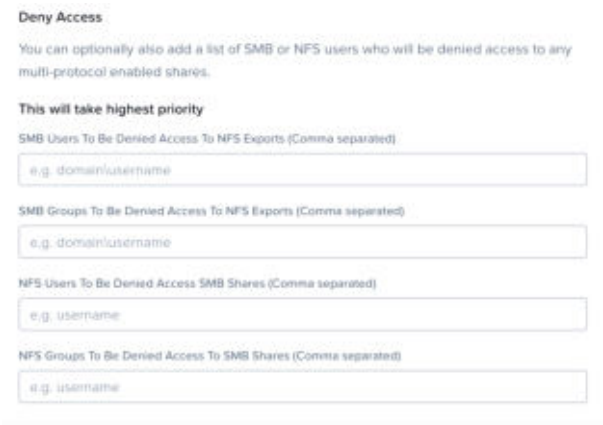


Figure 64: Deny Access Rules

e. Click **Save**.

4. (optional) Configure rule-based mapping.

   a. To map SMB and NFS users and groups, choose one of the following default rules:

      » SMB name matches NFS name.

      » No template mapping.

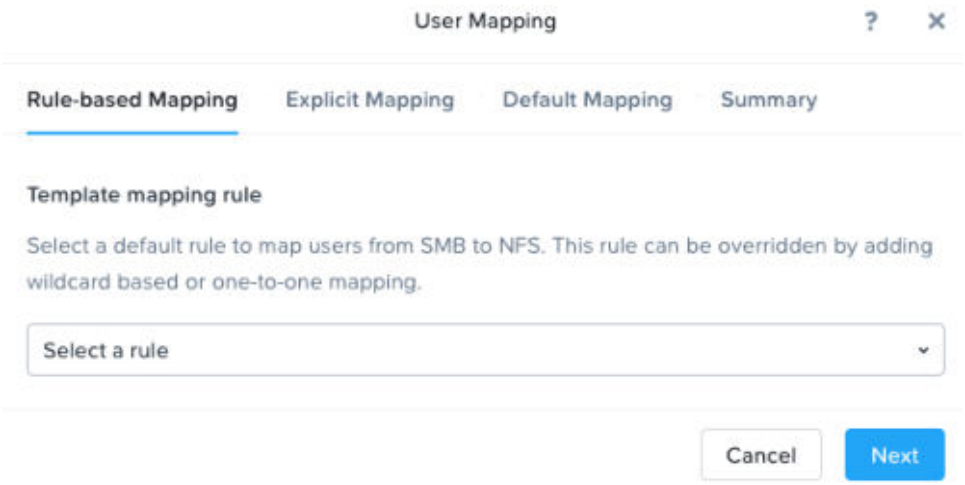


Figure 65: Rule-Based Mapping

   b. Click **Save**.

5. (optional) Set up default-mapping rules for Files to use when no applicable rule-based or explicit mapping rules exist for the user or group.

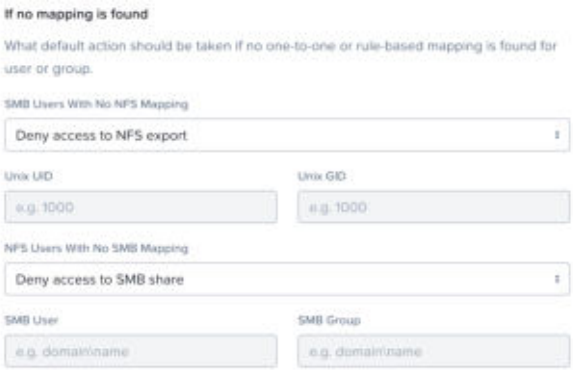


Figure 66: Default Mapping Rules

a. Choose one of the following options from the **SMB Users With No NFS Mapping** dropdown:

   » Deny access to NFS export.

   » Map to specific Unix user and group.

b. If you selected **Map to specific Unix user and group**, do the following (otherwise move on to the next step):

   - Enter a value in the **Unix UID** field.

     Note: If NFS has LDAP, use a user or group name. Otherwise, use a user or group ID.

   - Enter a value in the **Unix GID** field.

c. In the **NFS Users With No SMB Mapping** field choose from one of the following options:

   » Deny access to SMB share.

   » Map to specific AD user and group.

d. If you have selected **Map to specific AD user and group**, fill out the following fields (otherwise move on to the next step):

   - In the **SMB User** field, enter a default SMB user target for NFS users without mapping.

   - In the **SMB Group** field, enter a default SMB group target for NFS groups without mapping.

e. Click **Save**.

6. (optional) To delete all mapping rules for all users and groups, click the **Purge All Mapping** button

   CAUTION: Clicking **Purge All Mapping** permanently removes all existing mapping rules.

Managing Roles

Manage roles by adding, removing, or modifying administrator privileges.

About this task

You can create two types of Files administrators:

- File server admin. The file server admin can manage all file server operations, modify the access permissions for all users in all the shares/exports, and back up and restore data on the file server.
- Backup admin. The backup admin can back up and restore data on the file server (but does not have other administrative permissions).

> **Note:** Assign a backup service account (AD user or group) the backup admin role to prevent insufficient access issues.

To add or modify a Files administrator, do the following:

Procedure

Creating a New Role

1. In the Files Console (see Files Console on page 17), go to **Configuration** > **Manage Roles**.

   The **Manage roles** view displays.

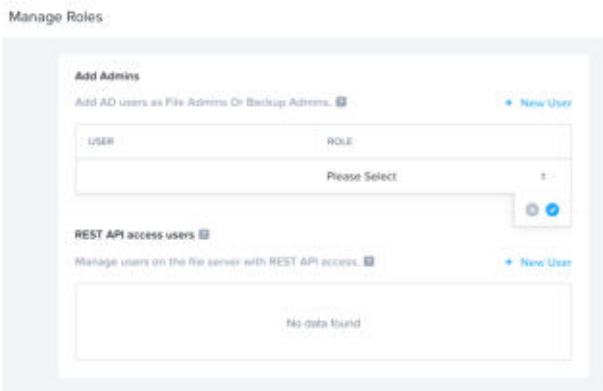


Figure 67: Manage roles window

2. To add an administrator, click **+ New user** in the **Add Admins** section.
   A line for new credentials appears at the bottom of the list.

3. Do the following :

   a. In the **User** field, enter the Active Directory user or group name.

      Enter user or group names in the `samAcctName` or NETBIOS\`samAcctName` format. Replace `samAcctName` with the SAM-account-name.

   b. In the **Role** field, select **File Server Admin: Full access** or **Backup Admin: Backup access only** from the pull-down list.

   c. To add the user, click the **check mark** icon.

   d. To add more administrators, repeat these steps.

Modifying Roles

4. To modify an administrator, click the pencil icon in the **three dot menu** > **edit** for that user and update the name or role as desired.

5. To delete an administrator, click the **three dot menu** > **edit** for that user.

Managing REST API Roles

Manage REST API access for all users on a file server.

About this task

To add, modify, or remove REST API access, follow these steps.

Procedure

1. In the Files Console, go to **Configuration** > **Manage roles**.
2. To add a RESP API user, click **+ New User** in the **REST API access users** section.
   A new line for the new user appears at the bottom of the list.

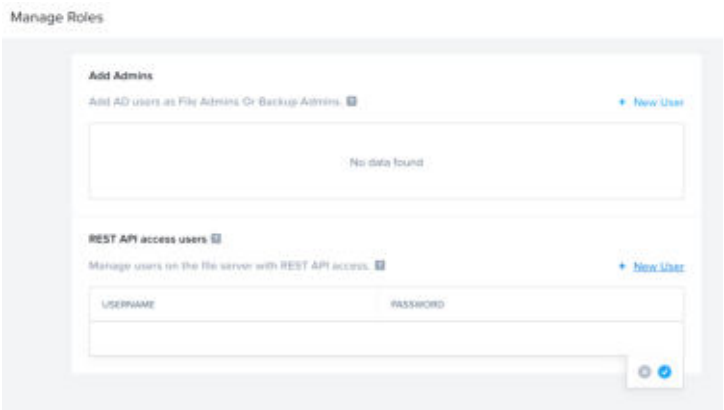


Figure 68: Manage roles window

3. To add a user, follow these steps:
   a. In the **Username** field, enter the username requiring REST API access.

      Note: You cannot add the "admin" username.

   b. In the **Password** field, type in a password for that user.
   c. In the same row, click the **check mark icon** to save the configuration.
   d. To add more users, repeat this step.
4. To modify a username or password, click **three dots menu** > **Edit** in the row for the target entry.
5. To delete a user, click **three dots menu** > **Delete** in the row for the target entry.

Authorizing a REST API User

To use the Files APIs, authorize a user in the REST API explorer.

About this task

Follow the steps as indicated from the Files Console.

Note: To create a REST API user, see Managing REST API Roles on page 116.

Procedure

1. Go to **Admin** > **REST API explorer**.

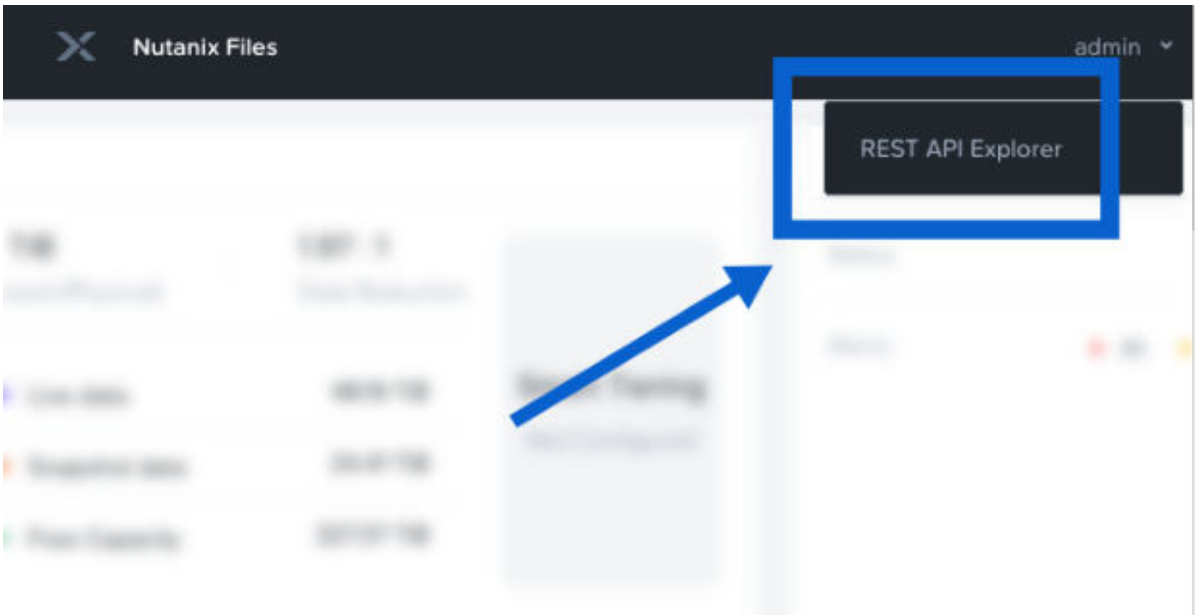


Figure 69: REST API explorer

2. Click **Authorize**.
   A dialog box for REST API user credentials appears.
3. Enter the REST API user credentials in the **username** and **password** fields.
4. Click **Authorize**.

Quotas

This topic describes the Files quota types, notifications, and policies.

Set quotas to allot the storage space a user or group can use.

There are two quota levels:

- *User*: Sets a specific amount of storage for a single user. For example, if an administrator allots only 1 GB, then you cannot use more than 1 GB – the total storage capacity for you is limited to 1 GB.
- *Group*: Sets the amount of space for each user in that group. For example, a group with a policy of 10 GB and 10 users can potentially use 100 GB of data (10 x 10 = 100 GB) under that quota policy.

Alternatively, rather than configuring quotas for specifics users, groups, or directories, you can configure a maximum share size to restrict the amount of storage space used by a share. To limit the space in a directory, see Setting Directory-Level Quotas on page 96.

Notifications

You can configure email alert notifications that Files sends when user or group quotas are near the maximum threshold. Files can send the alerts to you and other recipients. Emails notifications alert the recipient when the quota is near maximum and when it is near full consumption. When the quota reaches 90 percent consumption, Files sends warning emails to the recipients. When the quota reaches 100 percent consumption, Files sends alert emails to the recipients. If the quota has a soft limit, you can continue to consume over 100 percent of the storage and Files will send an email notification to the recipients every 24 hours.

Policies

A quota policy specifies the consumption limit and enforcement type for all quota levels as configured by the administrator. Enforcement types determine if a user or group can continue to use the quota once they consume their share. See the enforcement types descriptions in the following table.

> **CAUTION:** Quota policy enforcement begins several minutes after policy creation. Therefore, if you reach the quota limit before the interval is complete, Files raises the alert but does not enforce the quota.

> **Note:** Beginning with AOS 5.15.1 and AOS 5.17.1 you can set decimal quota values, earlier AOS versions only permit integer quota values. During a disaster recovery (DR) event to a container with a version earlier than AOS 5.15.1 and AOS 5.17.1, Files rounds the decimal quota value down to an integer.

Quota Configurations	Description
User or Group	The designated name for a specific user or group.
Quota	The limit of quota space (in GB).
Enforcement Type	• **Hard Limit**: Prevents further writes once quota limit is reached. • **Soft Limit**: Does not prevent writes. Sends email notifications to email recipients. **Note:** You cannot set both a soft and hard limit for the same user or group.
Email Recipients	Enable the email recipients box and enter the email addresses for recipients Files should notify about hard and soft quota limits.

Applying Quota Policies

Files resolves quotas policies per the following:

- If you have defined a Files user-level quota, then recipients receive the quota from this user-level policy.
- If you have not defined a user level, but you have defined multiple group-level policies, then Files applies the policy with the most space.
- If you have not defined a user or group policy for any given user, Files chooses the quota default policy.
- For distributed shares and exports, each user has one home directory. Therefore, quota applies only to the first user directory at the root of a distributed share.

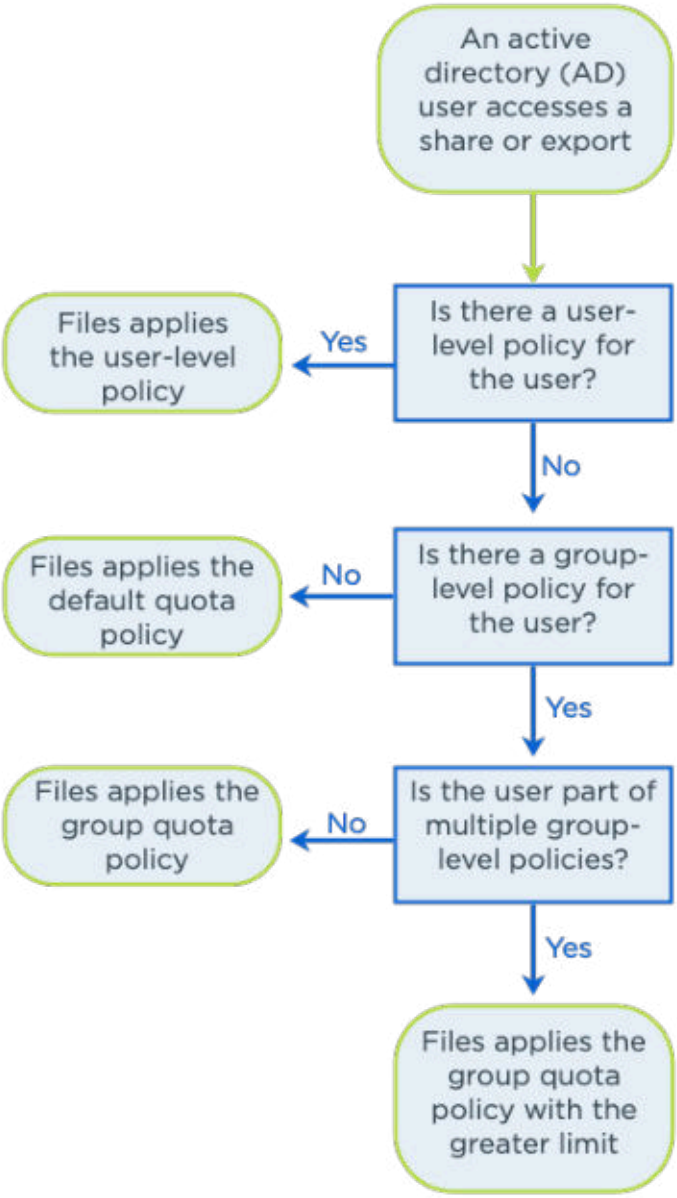


Figure 70: Quota Policy Hierarchy

> **Note:** If you add a new AD group and want to add a quota policy for that group, contact Nutanix Support to refresh the quota cache.

Multi-Protocol Limitations

The following limitations apply when you enable both SMB and NFS read-and-write access on a share or export.

- You can only apply a quota to users and groups of the primary protocol. For example, if SMB is the primary protocol, you can only apply a quota to SMB users.
- When you map multiple non-native users or groups to a single native user or group, Files only applies a quota to the first non-native user or group.

- Quota applies to non-native users mapped to native users who belong to a group quota.

Managing Quotas

Add or edit user or group quotas in Files.

About this task

Files implements user and group quota types that balance storage per a user of a share or export. To configure the quota levels for user, group, or default levels, perform one of the following steps.

Procedure

Creating a new quota policy

1. In the Files Console (see Files Console on page 17), go to **Shares**.
2. Click the name of the target share.

3. Click **Actions** > **Add Quota Policy**.

   a. Under **Add Users**, select to add the quota policy for a **Individual User** or **User Group**.

   b. In the **Username** field, enter the target user-name or group-name.

   c. In the **Quota Limit** field, enter the space for the quota limit (in GiB).

   d. Select the enforcement type.

      - **Hard limit**: Prevents further writes after reaching quota limit and puts the user or group into read-only mode.
      - **Soft limit**: Does not prevent writes. Sends email notifications to email recipients.

      > **Note:** You cannot set both a soft and hard limit for the same user or group.

   e. Check the **Send email notification to the Files administrator** box to enable email notifications and add email recipients in the **Email Recipients** field.

   f. To add the quota policy after entering the required information, click **Add**.

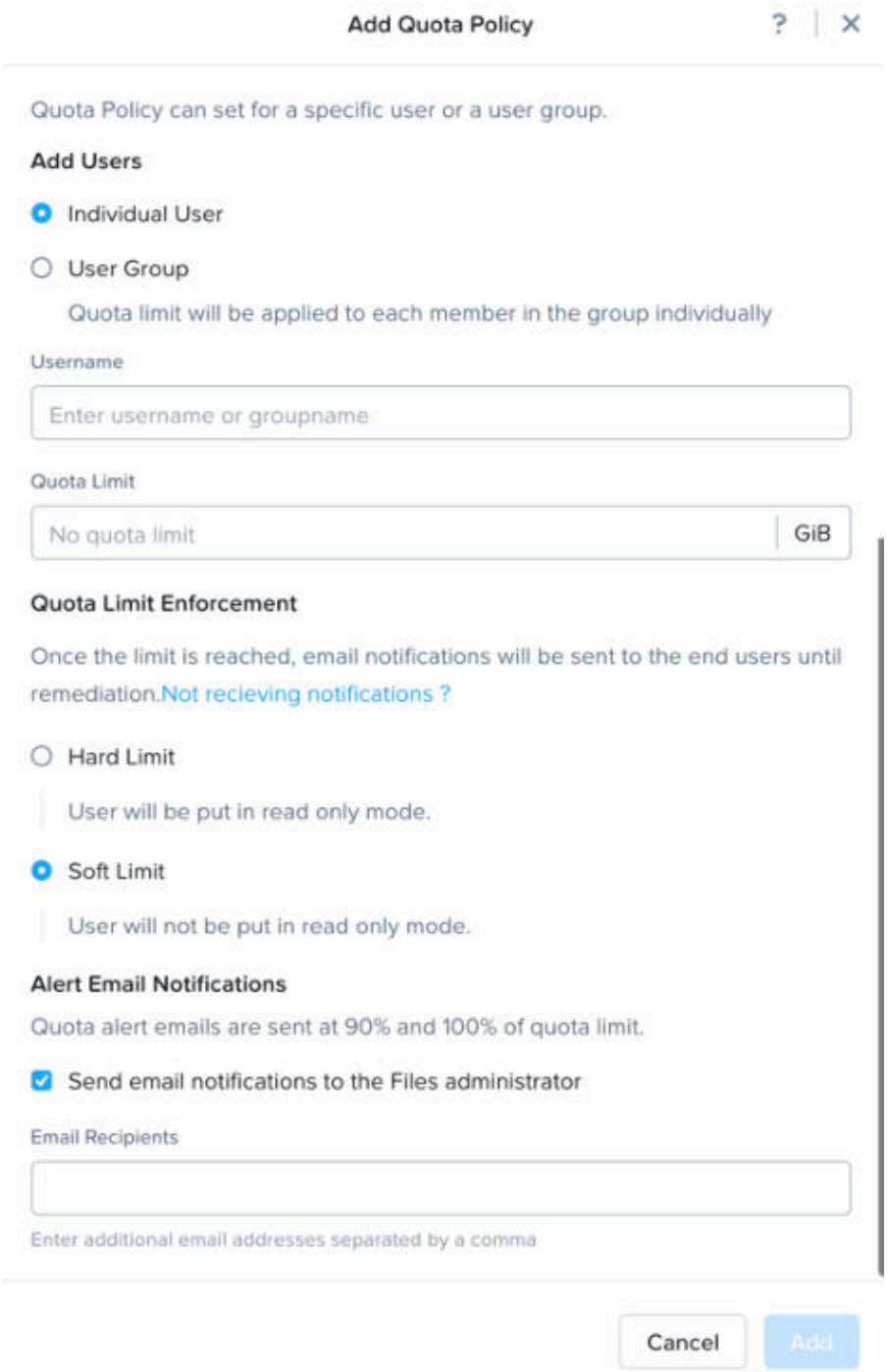


Figure 71: Add a Quota Policy

Editing a quota policy

4. Edit an existing quota policy.

   a. Click the share or export in the **Shares** tab.
      The **Quota Policies** tab displays all of the quota policies on the share.

   b. In the row for the target policy, click the **three dot menu** > **edit**.

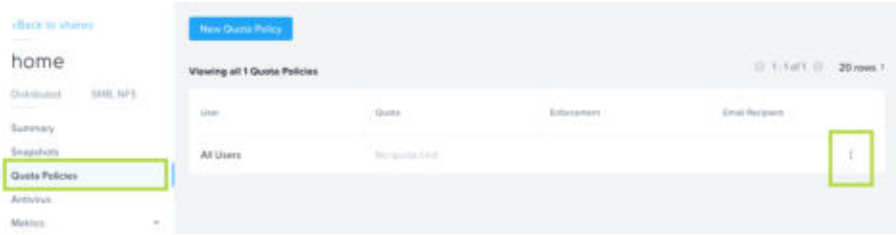


Figure 72: Edit an Existing Quota

   The **Quotas** window displays.

   a. You can edit the existing policy by updating the amount of space (GiB), the enforcement type (hard or soft limit), and the email notifications and recipients. Click **Add**.

> **Note:** To remove a quota, change the share size to 0.

Remove a quota policy

5. In the row for the target policy, click the **three dot menu** > **delete**.

FILES OPTIONS

Files provides a number of options you can employ to accommodate your file server implementation.

Cloning

Clone any file server protection domain snapshot at the local or remote site.

The file server clone is not protected by default. Be sure to enable the protection domain if you want the file server protected. Files cannot clone snapshots taken in earlier releases. Also, file server clones cannot be replicated or migrated to clusters that use earlier AOS and Files versions.

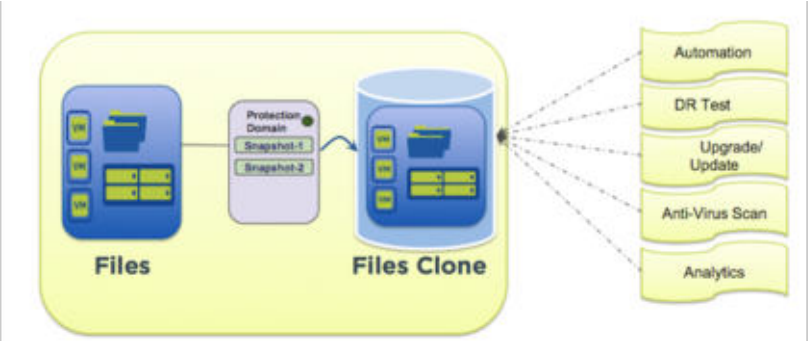


Figure 73: Cloning capability for Files

Cloning helps with the following without impacting the original Files cluster:

- Create backups at the primary and secondary sites
- Undertake DR test at secondary site
- Recover a file server from a specific point in time
- Spin-up a file server at the primary or remote site for testing or development purposes

Cloning a File Server

Follow this procedure to clone a file server from a specific snapshot.

About this task

The file-server clone is a thin copy that consumes minimal storage space. You cannot clone a file server to a storage container that is different from the original container.

Procedure

1. In the **File Server** view in PE (see File Server View in Prism on page 15), select the target file server and then click the **Clone** button.

   The **Clone File Server** window displays.

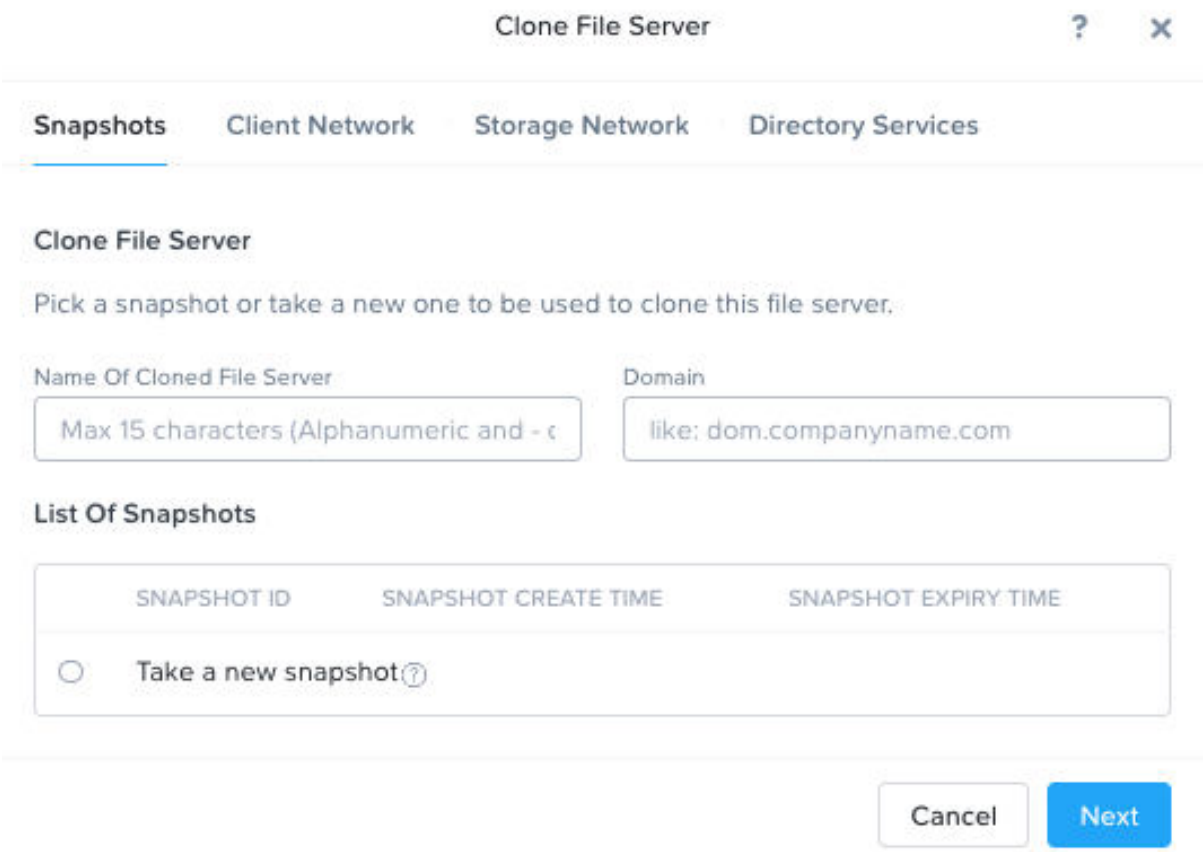


Figure 74: Clone File Server Window (Snapshots tab)

2. In the **Snapshots** tab, do the following in the indicated fields:

   a. **Name of Cloned File Server**: Enter a name for the new (cloned) file server.

      The clone name must be different that the original file server name.

   b. **Domain**: Enter a fully qualified domain.

   c. **List of Snapshots**: Click the option of the snapshot to use for the clone.

      A list of available snapshots (if any) appears in this field. Select one of the existing snapshots or select **Take a new snapshot**, which takes a new snapshot of the file server (after you complete this form) and then use that snapshot to create the clone.

   d. Click the **Next** button.

3. In the **Client Network** tab, enter the required information to configure the client network for the clone and then click the **Next** button.

   Note: See Creating a File Server on page 36 for details about configuring the client network, storage network, and user management.

4. In the **Storage Network** tab, enter the required information to configure the storage network for the clone and then click the **Next** button.

5. In the **Directory Services** tab, select one or more protocols to use (SMB, NFS, or both) and enter the specified configuration information. When all the information is complete, click the **Create** button.

Encryption

Encryption options for Files.

Files supports AOS software encryption and in-flight message encryption for SMB3 shares.

You can apply AOS software encryption to Files by activating it through Prism, see *Configuring Data-At-Rest Encryption (Software Only)* in the *AOS Security Guide*. Refer to the *Files Release Notes* to ensure that you are running a compatible version of AOS.

SMB3 Message Encryption

To enable SMB3 message encryption, see Modifying a Share or Export on page 87. After enabling message encryption, Files encrypts messages on the file server side and decrypts them on the client side (only on new connections for the share). Clients that do not support encryption (Linux, Mac, Windows 7) cannot access a share with encryption enabled.

Files Data Collection

Files data collection with Pulse.

The feature known as Pulse collects Files diagnostic system data and sends it to Nutanix Support. After you enable Pulse, Files synchronizes Pulse configurations from the Controller VM to the file server VMs. Synchronized data includes the Pulse enablement status and the mechanism chosen for streaming data to Nutanix Support. If Pulse data cannot reach Nutanix Support, administrators receive alerts in Prism. To enable or disable Pulse, see "Configuring Pulse" in the *Prism Web Console Guide*.

Access-Based Enumeration (SMB only)

Access-based enumeration (ABE) restricts user-access by only letting you view the files and folders you have read access to when browsing content on a file server.

About this task

ABE is a Microsoft Windows (SMB protocol) feature that filters the list of available files and folders on the file server to only include files and folders that the requesting user has read-access to. The filtering ensures that Files enforces read-and-write privileges for all users and that information can remain confidential. ABE controls the user visibility of shared folders on mounted file system shares based on the user permissions.

Enable ABE during or after share creation.

> **Note:** To activate ABE after a group membership of a user changes, remove all previous share sessions, remount the share, and reconnect existing client connections.

Procedure

1. To enable ABE during share creation, see Creating a Share (SMB) on page 64.

2. After creating a file share, you can modify ABE settings. To enable ABE after share creation, do the following:

   a. In the Files Console, go to the **Shares** tab.

   b. In the row for the target share, click the **three dots menu** > **Edit**.
   The **Update Share** window displays.

   c. In the **Settings** tab, check (enable) or clear (disable) the **Enable Access Based Enumeration (ABE)** box and click **Save**.

File Blocking

Restrict specific files or file types from appearing on a file server or share.

Specify a character pattern of file names or extensions to block files. Use an asterisk (*) as a wildcard for multiple characters or a question mark (?) as a wildcard for a single character.

> **Note:** The question mark character (?) only matches UTF-8 single byte ASCII characters. The question mark character (?) does not apply to multibyte unicode characters.

Files applies the file blocking policy to all levels of a share or export, which disables the ability to create files with the specified character pattern in the name. An attempt to create blocked files results in an error. Share-level file blocking overrides the files blocked on the file-server level.

Files allows a maximum of 300 file blocking patterns on a file server.

After enabling file blocking, Files does not permit the following operations:

- Creating a file with the blocked character pattern.
- Renaming an existing file to one with the blocked character pattern.
- Duplicating a file with the blocked character pattern.
- Moving a file with a blocked character pattern.

You can still perform read-and-write operations on existing blocked files.

To enable file blocking on a file server, see File Server Updates on page 53. To enable file blocking on a share or export, see Modifying a Share or Export on page 87.

Files blocked on a share level appear in the **Summary** tab of the **Create share/export** and **Update share/export** windows. Files blocked on a file server level appear in the **Blocked File Types** tab.

Blocking Files on a File Server

Block the creation and modification of files on a share, export, of file server.

About this task

Refer to File Blocking on page 127 for information about the file blocking feature. To block files on the share or export level, refer to Modifying a Share or Export on page 87.

To block files with specific character patterns in their names, do the following.

Procedure

1. In the **File Server** view, select the target file server.

2. Click **Update** > **Blocked File Types**.

3. In the **Blocked File Types** field, enter (or modify) a comma-separated list of file extensions for blocked file types.

Figure 75: Blocked File Types

4. Click **Save**.

Antivirus (AV) Scanning (SMB Only)

Third-party antivirus software for SMB shares.

Files supports the Internet Content Adaptation Protocol (ICAP) to enable communication with external servers hosting third-party antivirus software. This software scans files stored on file shares to help provide protection against viruses. This software scans files in real time when you open, close, read from, or write to files.

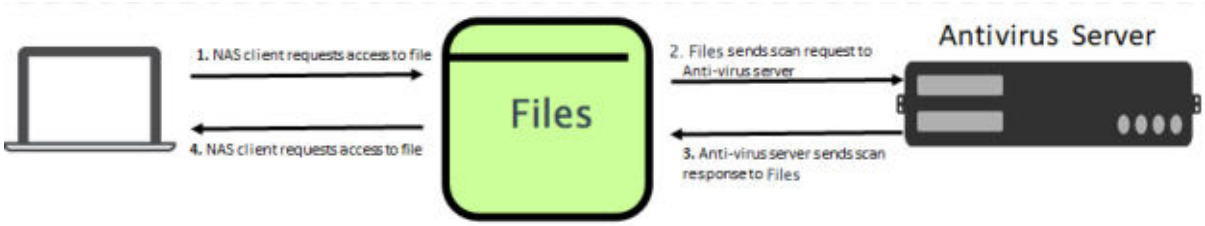


Figure 76: Files AV Concept

> **Note:** Refer to the Compatibility Matrix for a list of compatible security software. Filter by **Solution Type** > **Security** and **Additional Component** > **Nutanix Files**. Files requires two or more ICAP servers. Nutanix recommends having a minimum number of scanning threads that is 11 times the number of FSVM nodes or (11 * number of FSVM nodes).

Overview

Files performs the following tasks with ICAP servers when a client requests to read, write, open, or close a file.

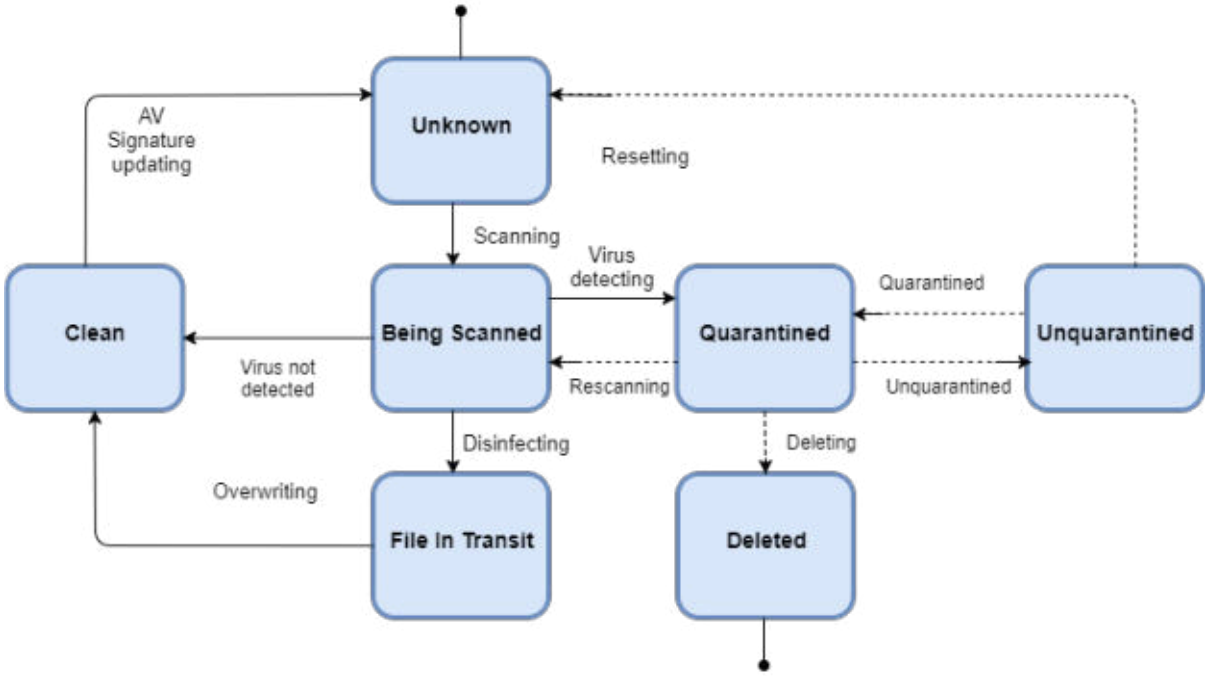
1. Files determines that the file requires scanning.
2. Files sends files that require scans to the ICAP server with a scan request.
3. The ICAP server scans the file and reports the scan results to Files.
4. Files quarantines and denies access to unsafe files.
5. If the file is clean or disinfected, then Files allows the client access to the file.

Note: By default, shares have antivirus scan disabled.

Antivirus File State

This diagram shows the process flow for file scans.

Administrator actions are denoted with dotted



lines.

Figure 77: Antivirus File State Diagram

Glossary

Files and Prism Element use the following terms to show file status applied by the antivirus scanning feature.

Table 19: State

State	Definition
Quarantined	A scanned file that the antivirus scan qualifies as unsafe. Files blocks access to the file until the administrator manually changes the file state.
Unquarantined	The administrator moves the file from the quarantined state to allow client access. Files does not rescan unquarantined files.

Table 20: Events

Event	Definition
Cleaned	The antivirus scanner has scanned and cleaned the file. This process overwrites the original file. Using this feature requires the disinfected virus file function on the ICAP server.
Quarantined	A scanned file qualified as unsafe. Files blocks all access to the file until the administrator manually changes the file state.
Unquarantined	The administrator manually moves the file from the quarantined state to allow client access. Files does not rescan unquarantined files.
Deleted	File removed from the file system.

Configuring Antivirus Scanning (SMB Only)

Configure and enable antivirus scanning for SMB shares.

About this task

After configuring the antivirus scan, enable the scan for each share that you want scanned.

Procedure

1. In the Files Console, go to **Configuration** > **Antivirus**.
2. Connect the ICAP server.
   a. Click **+ Connect ICAP Server**.
   A new row appears for new ICAP server details.
   b. Enter the following information in the corresponding fields:
      - IP address or hostname
      - Port (the default port number is 1344)
      - Description
   c. To save the configuration, lick the **check mark icon**.
   For a detected antivirus server, the software tests the validity of the configured server and updates the status to `OK`.
   d. Ensure the connection status automatically updates to **OK**.
   e. Click **Next**.

3. Complete the **Scan Settings**.
   You can override settings through the share-level antivirus settings.

   > **Note:** Nutanix recommends two or more ICAP servers.

   a. **Scan on Write**: Scans saved and updated files (a write operation).

   b. **Scan on Read**: Scans opened files (read operation).
      Nutanix recommends to always enable **Scan on Read**.

   c. **File extensions to be excluded**: Add one or more file extensions to exclude from the scan.

      > **Note:**
      >
      > - Ensure these settings match the file type configuration of your ICAP servers.
      > - Nutanix recommends adding the following file extensions for user profiles when using the Files antivirus scanning:
      > - `.dat`
      > - `.ini`
      > - `.pol`
      > - When Files with a specific extension type are quarantined incorrectly by the ICAP server, adding this file type extension to the ignore list only prevents future file quarantines. Remove the quarantine for the incorrectly quarantined files to access them.

   d. **File Size**: Limit the size scanned of files.

   e. **Advanced Settings**:

      - **Scan Timeout**: Set the maximum amount of time that a scan can take before timing out.
      - **Block access to files if scan cannot be completed (recommended)**: Block access if the ICAP servers are unavailable or cannot scan the file for any reason.

   f. Click **Save**.

4. Enable the antivirus scanning on each share.

   a. Go to the **Shares** tab and click on the target share.

   b. In the **Share Details**, go to **Actions** > **Configure Antivirus**.
      The **Configure Antivirus** setup window displays.

   c. > **Note:** By default, antivirus scan is disabled on all shares.

      Check the **Enable antivirus scan** box.

   d. (optional) Change the settings for the share (see Step 3 for details).

   e. Click **Save**.

Antivirus Tab

The layout and elements of the **Antivirus** tab in the Files Console.

The **Antivirus** tab displays dynamic information about scanned files. The tab includes the following sub-tabs:

- ICAP servers
- Reports
- Quarantined Files
- Unquarantined Files

To view this information, in the Files Console, go to **Monitoring** > **Antivirus**.

ICAP Servers

The ICAP servers tab displays the scanned files information for each ICAP server.

- **ICAP Server Statistics**: The table displays information such as port number, description, files scanned, disconnect count, average latency, connection status, and actions available.
- **Average Latency**: This graph displays the latency times for the scans (in milliseconds).
- **Files Processed** or **Data Processed**: Click the **Files Processed** drop-down arrow to select the files processed or data processed graph. The processed files graph displays the number of scanned files. The data processed graphs display the amount of processed data (in GiB).
- **Queue Length**: The number of files in the scan queue.

Reports

The **Reports** tab displays the information about the scanning period and share status.

- **Scan Period**: This information displays the files scanned, threats detected, number of files cleaned, and number of files quarantined during each scanning period.
- **Share Status**: Displays the state of the scanned share. The parameters includes: file path, threat description, ICAP server, time, action taken on share.

Quarantined Files

The **Quarantined Files** tab displays the files that contain a virus. The antivirus software places virus-infected files into quarantine where clients cannot read or write the files. An administrator can perform the following actions on the quarantined files.

- **Rescan**: Rescan the files that have been quarantined.
- **Unquarantine**: Move the files out of quarantine. The selected file is then available for use.
- **Delete**: Delete the quarantined file permanently.

Unquarantined Files

The **Unquarantined Files** tab displays files manually released from quarantine. You can use unquarantined files. Files does not rescan unquarantined files again until the administrator resets the unquarantine state. Perform the following actions on unquarantined files.

- **Reset**: Move the files to a normal state that is not quarantined or unquarantined. In this state, the next access to the file triggers the scan.
- **Quarantine**: Move the files to quarantine to block read and write access.

> **Note:** Reaching the limit of the number of files in both the **Quarantined** and **Unquarantined** tables impacts scanning through the UI. The web consoles alerts you when the number of files has reached 80 percent of the number of files supported.

Files REST APIs

An introduction to Files REST APIs and the Files REST API explorer.

> **CAUTION:** Alpha APIs are intended for use in testing clusters only and are meant for early feedback from customers. Do not use the alpha APIs in a production environment. Support for alpha API-based features may not appear in future releases. Revisions of multiple v4 API versions may not be compatible. Also note that the APIs could be incomplete, the object schema and semantics may change drastically. There is no commitment on support for alpha APIs from Nutanix Support.

The Files REST API Explorer offers developers tools to customize the Files experience using Files v4 alpha REST APIs. You can access the Files REST API Explorer through the Files Console or through an FSVM. The Files service v4 APIs are independent from Prism Element (PE) and Prism Central (PC) APIs. However, the platform APIs, for operations such as create, clone, update or delete a file server, remain in Prism Element (PE).

To access the Files API Explorer in the Files Console (see Files Console on page 17), go to **admin** > **REST API Explorer**.

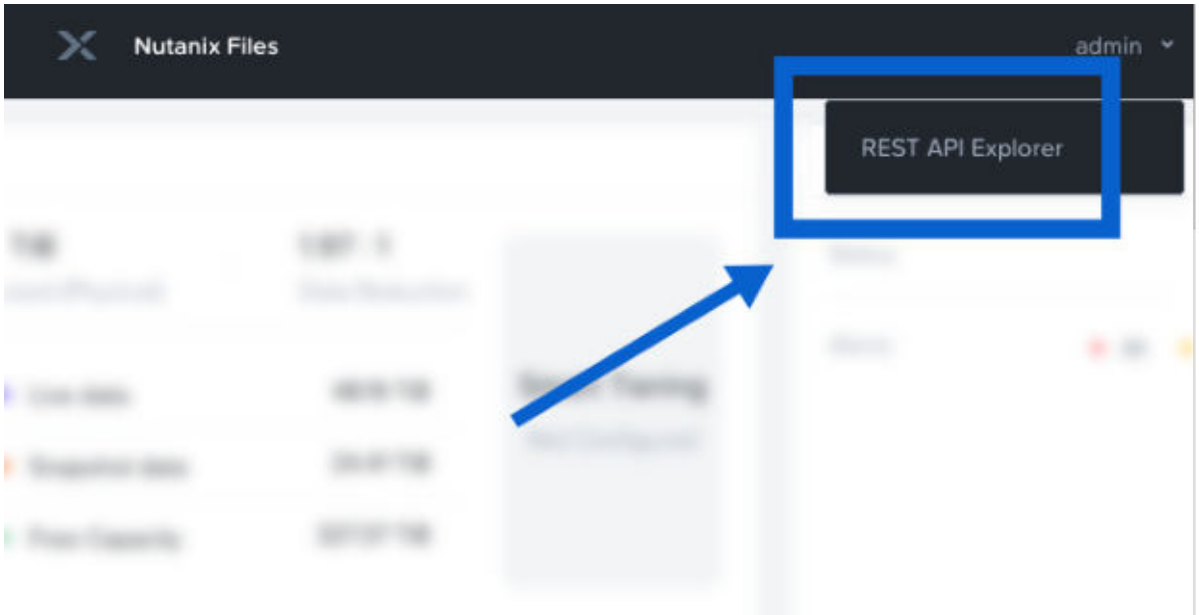


Figure 78: REST API Explorer

PERFORMANCE OPTIMIZATION

Files performance optimization notifies you when the file server is under load and needs a change to improve optimization.

Performance optimization includes scale up and scale out options. The **Dashboard** tab in the Files Console includes a **Recommendations** widget that indicates performance optimization options.

Scale-up recommendations occur when an FSVM reaches 95% client connection utilization over a two hour time window. If performance is slow due to storage group disruption, the recommendations for optimal performance options appear in the **Recommendations** widget. When possible, perform scale-out and rebalancing operations during maintenance windows, as scaleout and rebalancing disrupts existing connections.

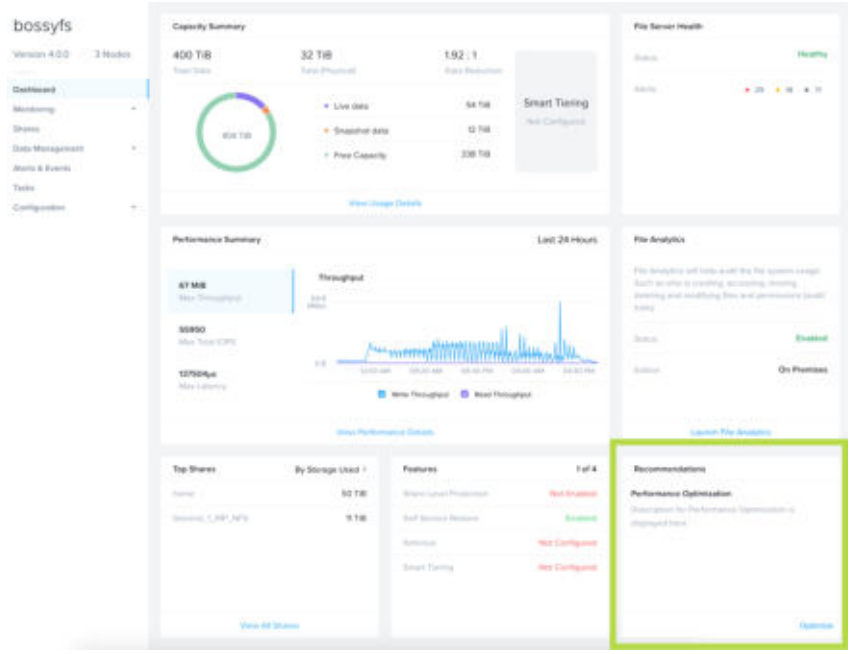


Figure 79: Performance Optimization Recommendations

Note: Files options.

Note:

Managing Performance Optimization

Performance optimization moves consistently used storage between available storage groups within the file server.

About this task

CAUTION: This operation can cause a momentary connection drop for end users accessing files on the file server. When possible, perform scale-out and rebalancing operations during maintenance windows, or off-peak hours.

Procedure

1. In the Files Console, go to the **Dashboard** tab.
   The file server displays a warning to recommend performance optimization.

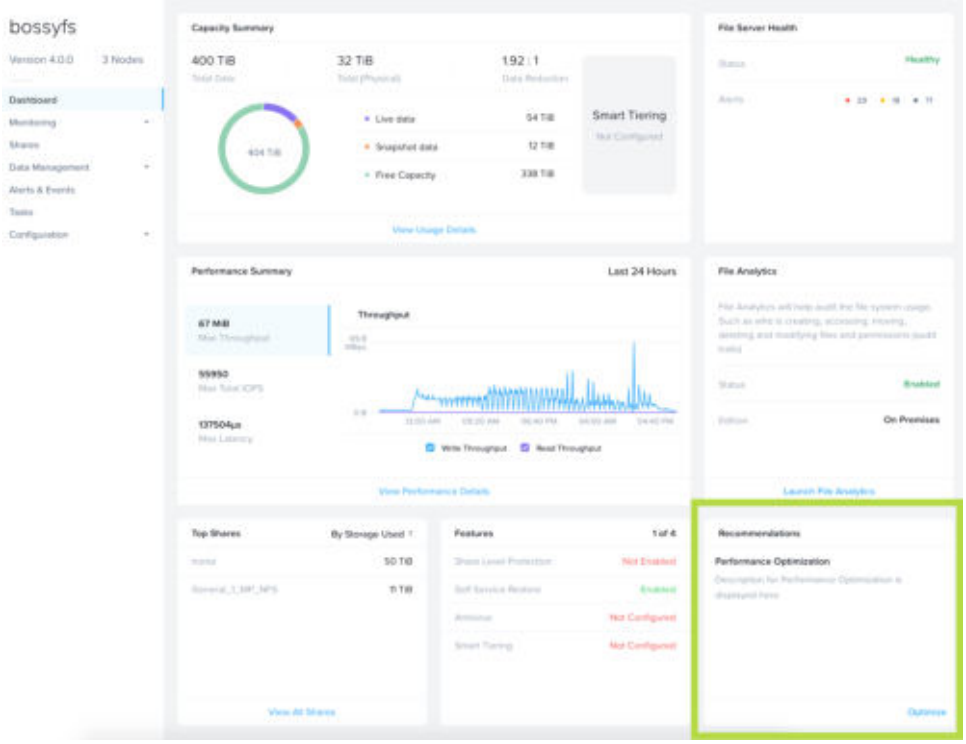


Figure 80: Performance Optimization Warning

2. Under **Recommendations**, click **Optimize**.
   The **Recommendations: Performance Optimization** window appears.

3. Select one of the following options:

   » Scale up

   » Rebalance

   » Scaleout

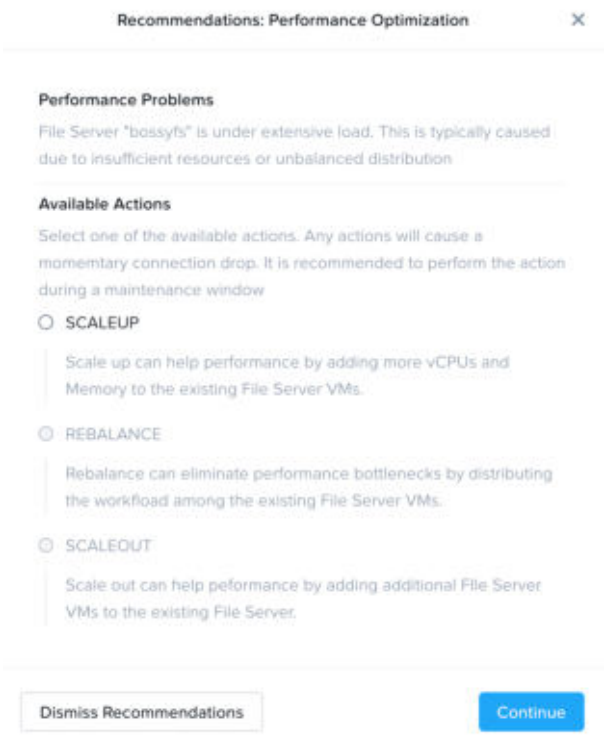


Figure 81: Performance Optimization Recommended Options

4. Click **Continue**.

   » For the rebalancing, confirm that you are performing the operation during off-peak hours.

   > **Note:** Perform rebalance operations during off-peak hours. For earlier AOS versions, you must manually unblock the rebalancing operation, see Unblocking Rebalancing on page 137.

   » For scale up or scale out, continue to the next step to update the file server capacity configuration.

5. (Scale up and scale out only) update the file server capacity configuration.

   a. Review or modify the recommended value for **Number of VCPUs Per File Server VM** as it appears in the drop-down menu.

   b. Review or modify the recommended value for **Memory Per File Server VM** as it appears in the drop-down menu.

   c. To complete the configuration, click **Save**.

Dismissing recommendations

6. In the **Recommendations: Performance Optimization** window, click **Dismiss Recommendations**.

   New recommendations appear in the **Recommendations** widget once the system identifies new instances of high performance load.

Unblocking Rebalancing

Unblock rebalancing guardrails.

About this task

Files guardrails prevent initiating rebalancing operations during high-volume periods.

Procedure

Unblock load-balancing guardrails on the file server.

```
nutanix@fsvm$ afs lb.unblock_rebalancing
```

Workload Optimization

This chapter describes options for optimizing the performance of your Files cluster.

There are three types of share workload types: default, random, and sequential. Designating a workload type determines the file system characteristics (including block size) used for the share, which optimizes the resource usage and performance of certain workloads. For example, workloads with small I/O on large files perform more efficiently with small block sizes.

The different workload types have the following specifications.

- Default: Uses 64 KB per block. Does not provide specified optimization. The share can perform all workloads varying performance
- Random: Uses 16 KB per block. Optimized for small I/O workloads.
- Sequential: Uses 1 MB per block. Optimized for large I/O workloads. Requires a minimum of 24 GB memory per FSVM.

The **Shares** view in the Files Console includes a **Metrics** tab and a **Performance** subtab, which displays write, read, and metadata I/O per second in the **IOPS** graph. Use the data from the graph to configure the workload type for the share, see Modifying the Workload Type on page 137.

> **CAUTION:** If you modify the share type without following the workload optimization guidance as specified, share performance can degrade.

- If the I/O sizes for read and write operations are less than or equal to 16 KB and the file sizes equal to 10 MB or more, use the **Random** workload type.
- If the I/O sizes for read and write operations are less than or equal to 1 MB and the file sizes equal to 10 MB or more, use the **Sequential** workload type.
- If the I/O sizes for read and write operations do not match the criteria for **Random** or **Sequential** workload types, use the **Default** workload type.

Modifying the Workload Type

To optimize performance, modify the workload type of a share or export.

About this task

Perform the following steps to modify the workload size of a share or export, see Workload Optimization on page 137 for optimization guidelines.

> **Note:** Changing the workload type of a share changes the performance characteristics only for the files created after the change.

Procedure

Replace *share-name* with the name of the share or export. Replace *workload* with one of the following workload types: `default` , `sequential`, or `random`.

```
nutanix@fsvm$ afs share.edit share-name share_workload_type=workload
```

File System Compression

File system compression reduces the input and output (I/O) load, iSCSI traffic, space usage, and the amount of data on a share or export.

You can enable file system compression at the share or export level. Earlier versions of Files included compression at the container level. Files applies share-level compression during ingest operations, compressing incoming data in-line prior to writing it to storage. As a result, share-level compression reduces storage traffic between Files and AOS.

> **Note:** Only clusters created with later versions of Files and AOS support file system compression. The option to enable file system compression does not appear on clusters created with earlier versions of Files and AOS.

To enable or disable compression on the share-level, see the following:

- Creating an Export (NFS) on page 70
- Creating a Share (SMB) on page 64
- Creating a Multi-Protocol Share or Export on page 78
- Modifying a Share or Export on page 87

DATA MANAGEMENT

File server data recovery and management options.

Files provides several features to manage how you access and recover data on your file server. Manage the availability of data using data recovery features, which include self-service restore (SSR) and several types of disaster recovery (DR). Use the tiering feature to maximize space on the file server by moving stale data to an object store. You must configure tiering policies for the file server through File Analytics, see Smart Tiering on page 145 and the *File Analytics Guide* for more details.

Async and NearSync DR replicates data to a protection domain at the granularity of a file server, while Smart DR replicates data with share-level granularity to a recovery file server (see "Smart DR" in the *Files Manager User Guide* for information on Smart DR). If there is a disaster, you can restore your data from snapshots on the protection domain or on the recovery file server.

SSR does not copy file data but instead takes read-only snapshots at the share/export-level. You can recover data from a deleted file or an earlier version of a file based on your snapshot retention configuration. SSR for SMB lets you restore and manage previous versions of Files at the system-level, while SSR for NFS lets you restore files by manually copying read-only snapshot versions.

Both DR and SSR let you configure a desired snapshot interval. With DR, you can set a specific time to take the snapshot, which is not possible with SSR.

Refer to Self-Service Restore on page 145 and Data Protection and Recovery on page 139 sections for more details on configuring SSR and DR.

Data Protection and Recovery

Files supports disaster recovery (DR) through customizable protection domains and protection policies.

> **Tip:** For information on disaster recovery with share-level replication, see "Smart DR" in the *Files Manager User Guide*.

As part of DR, Files automatically creates a protection domain for a file server and the entities within the file server (such as VMs and volume groups) during file server creation. By default, Files adds all entities on the file server to the protection domain.

To activate DR, enable and set up the schedule for snapshots and replication for the protection domain. Files Async and NearSync DR take snapshots when the preceding snapshot is complete. Async has a 60-minute recovery point objective (RPO), and NearSync has a 1-minute RPO.

Files creates a dedicated container for each file server instance, which cannot be used by another file server, VM, or for any other purpose. This requirement also applies to remote containers used for replication. If you want to replicate a file server container to a remote site, make sure that the remote container, like the local container, is not used for any other purpose. The remote site must have at least the same number of nodes as the number of FSVMs in the Files instance. To ensure feature parity after activating the file server on the remote site, both sites must have the same AOS version.

You can provide custom names for the file server protection domains at the time of setting up the file server (see Creating a File Server on page 36). If a file server does not have a specified protection domain name, the default protection domain name is `NTNX-`*`file_server_name`*.

Files supports cross-hypervisor DR. Configuration steps are the same.

> **Note:** You can restore all self-service restore (SSR) and Windows Previous Version (WPV) snapshots that exist at the time of the protection domain (PD) snapshot locally or remotely.

Configuring Disaster Recovery

This task describes how to set up disaster recovery for file server clusters for planned or unplanned (disaster) migration.

About this task

To set up file server Async disaster recovery (protection domain based), follow the steps in this procedure.

> **Note:** Refer to "Smart DR" in the *Files Manager Guide*, for steps on configuring Smart DR.

Procedure

1. If you have not done so already, configure a remote backup site to the local cluster.

   See the "Configuring a Remote Site (Physical Cluster)" topic in the *Data Protection and Recovery with Prism Element* guide.

   > **Note:** The remote site must have at least the same number of nodes as the number of FSVMs in the Files instance.

2. In the **Files Console** (see Files Console on page 17), go to **Data Management** > **Protection**.

   The **Disaster Recovery** and **Self Service Restore** tabs appear. In the **Disaster Recovery** tab, the **Protection Domain (PD) Based** and **Smart DR** sections specify if the indicated DR type is enabled.

   After creating a file server, Files automatically adds the file server to a newly created protection domain. However, when the file server does not have a protection domain, the **Configure** action link displays in the **Protection Domain (PD) Based** section with the **not enabled** status.

3. To add the file server to a protection domain, do the following:

   a. Click **Configure**.
      Files redirects to the **File Server** view in Prism Element.

4. In the file server table, select the file server by clicking the row it appears in.

5. Under the file server table, click the **Protect** action link.
   The **Protection configuration: [file-server-name]** window appears.

6. In the **Disaster Recovery** section do the following:

   a. (Optional) in the **Protection Domain Name** field, update the name of the protection domain.

   b. Click **Protect File Server**.

7. (Optional) configure a schedule for disaster recovery.

   > **Note:** Prism creates a default schedule for every protection domain.

   a. In the Files Console, go to **Data Management** > **Protection**.

   b. Click **Manage on Prism Element**.
      Files redirects you to Prism Element.

   c. See Creating a Protection Domain Schedule on page 141 to add or modify the protections schedule.

8. Configure the local and remote container mapping.

   a.

   > **Note:** In Metro Availability enabled environments, avoid using identical container names. Ensure that the name of the remote container is unique, then map the containers to each other.

   If you did not map the local and remote containers when configuring a remote site (vStore name mapping), create a new remote container.

   > **Tip:** See the "Creating a Storage Container" section in the *Prism Web Console Guide* for this procedure.

   Changing a vStore mapping causes associated protection domains to initiate full replication of protected entities to the newly specified target container as if it were initial replication. A mapping change therefore results in the overconsumption of storage resources at the remote site. Contact Nutanix Support for help with cleaning up snapshots in the previously specified container.

   b. Ensure that the remote and the local containers have symmetric configurations and that both containers map to each other. On the remote VStore site, configure explicit mapping between the source and the destination container.

Creating a Protection Domain Schedule

This task describes how to create a snapshot schedule for Files protection domains.

About this task

Create a snapshot schedule for the protection domain to use Files disaster recovery.

> **Note:** Ensure to also save the snapshot schedule on the remote site.

Procedure

1. In Prism Element, go to the **Data Protection** tab in the pull-down menu.
2. In the **Table** view and **Async DR** tab, select the protection domain from the table.
3. Click **Update**.
   The "Update Protection Domain" window appears.
4. In the **Schedule** tab, click **New Schedule** (or, to update an existing schedule, click the **pencil icon**).

5. Complete the indicated fields:

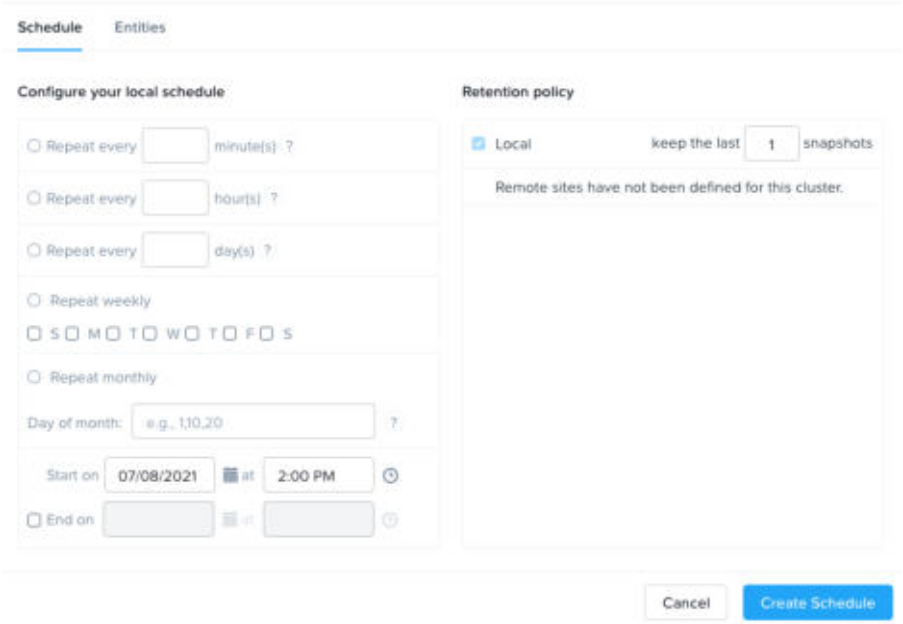


Figure 82: Create the File Server Protection Domain Schedule

a. **Repeat every [minutes|hours|days]**: Click the appropriate circle for minutes, hours, or days and then enter the desired number in the box for the scheduled time interval.

   The interval cannot be less than 1 minute.

   > **Note:** Intervals of less than 60 minutes use NearSync disaster recovery. NearSync schedules inherit requirements and limitations of AOS NearSync, see "Requirements of Data Protection with NearSync Replication" in the *Data Protection and Recovery with Prism Element* guide.

b. **Repeat [weekly|monthly]**: Select which days to run the schedule.

   - If you select weekly, select the boxes for the days of the week the schedule should run.
   - If you select monthly, enter one or more integers (in a comma-separated list) to indicate which days in the month to run the schedule. For example, to run the schedule on the 1st, 10th, and 20th days, enter "1,10, 20".

c. **Start on**: Enter the start date and time in the indicated fields.

   The default value is the current date and time. Enter a new date if you want to delay the schedule from starting immediately.

d. **End on**: To specify an end date, check the box and then enter the end date and time in the indicated fields.

   The schedule does not have an end date by default, and the schedule runs indefinitely unless you enter an end date.

e. **Retention Policy**: Enter the number of snapshots to save locally and at the remote sites.

   - Enter a number in the **Local** line "**keep the last ## snapshots**" field. The default is 1.
   - Enter the number of snapshots to save on the **Remote Site** in the "**keep the last ## snapshots**" field. This number can be different from the number that you have entered in the **Local** line. This replication is

an async replication. After the replication completes, the protection domain is going to be available in the **Async DR** tab of the remote site.

- The saved snapshots equal to the value entered in the **keep the last ## snapshots** field + 1. For example, if you entered 20 as the value for **keep the last ## snapshots** field, Files saves 21 snapshots. When Files takes the next (22nd) snapshot, Files deletes the oldest snapshot and replaces it with the new snapshot.

> **Note:** If too many schedules have the same start time, replications can fall behind. To avoid this issue, stagger start times across schedules.

Activating Disaster Recovery

This topic describes how to recover a file server cluster after a planned or unplanned (disaster) migration.

About this task

> **Note:** The name for the automatically created protection domain contains `NTNX` as a prefix followed by the file server name.

Procedure

1. Fail over the protection domain.

   Unactivated protection domains display shaded indicators next to the protection domain name. Activated protection domains display green indicators next to the name. For AOS 5.15 in later, see "Failover and Failback Operations for Asynchronous and NearSync DR" in the *Data Protection and Recovery with Prism Element* guide. For earlier AOS versions, see the "Failing Over a Protection Domain" topic in the *Prism Web Console Guide* for either a planned (migration) or unplanned (disaster recovery) activation procedure.

2. Fail back a protection domain.

   See the "Failing Back a Protection Domain" topic in the *Prism Web Console Guide* for this procedure.

What to do next

Activate the file server for planned or unplanned migration (see Activating a File Server on page 143).

Activating a File Server

Activate an inactive file server.

About this task

Follow the steps as indicated to activate a file server after disaster recovery.

Procedure

1. In the Prism Element **File Server** view, select the target file server.

   When a file server is inactive, `(Needs activation)` appears next to the file server name and an **Activate** button appears in the action button list (just below the file server table).

2. Click the **Activate** button.

   The **Activate File Server** window appears.

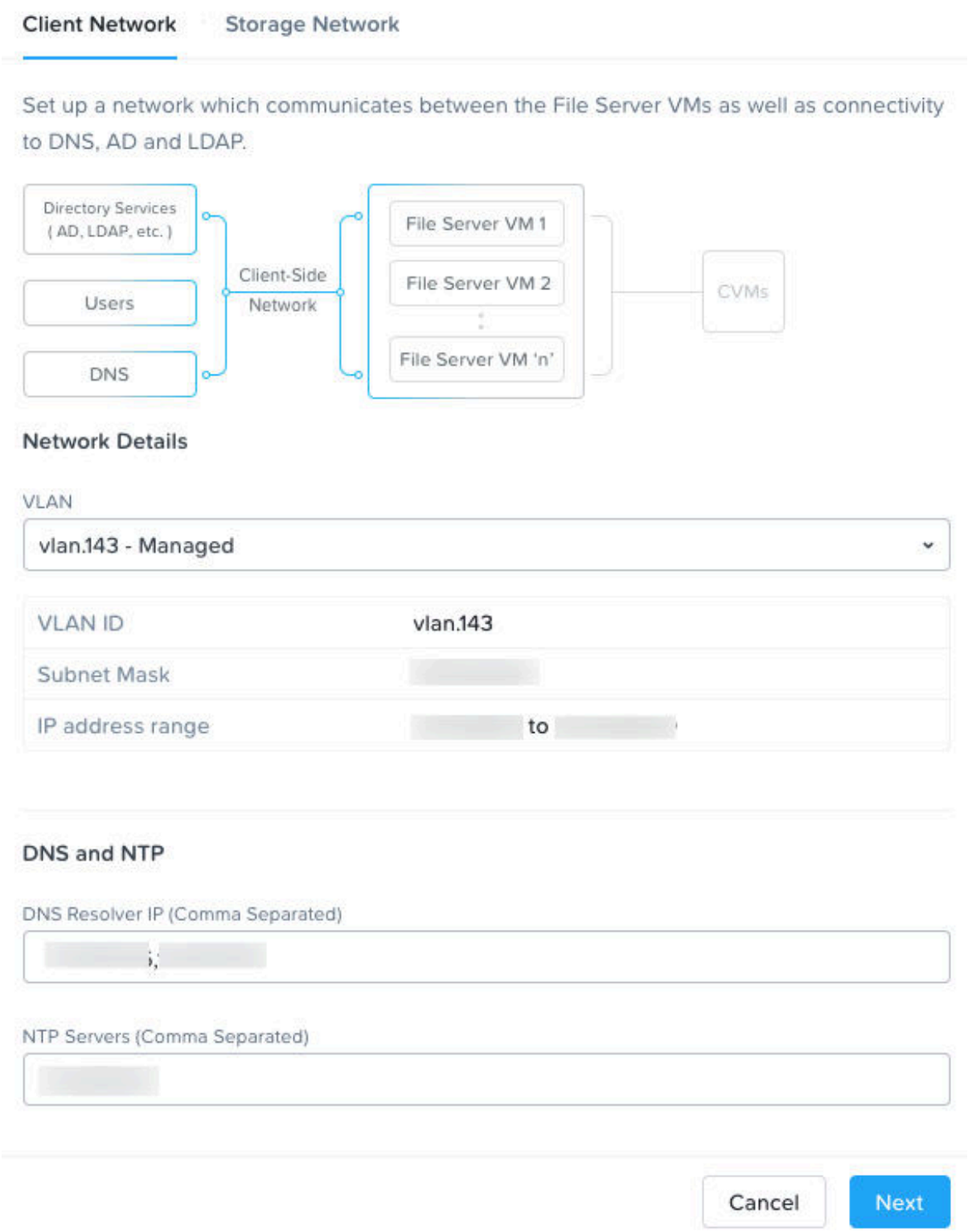


Figure 83: Activate File Server

3. Complete the indicated fields in the **Client Network** and **Storage Network** tabs.

   Some fields populate from information provided when the file server was created. See Creating a File Server on page 36 for more information about the fields.

4. When all the information is complete, click the **Save** button.

   The file server configuration updates, and the `(Needs activation)` message disappears indicating the file server is now active.

High Availability

Fail over for file server VMs (FSVMs).

High Availability (HA) for Files insures that during a disruption of service a file server VM (FSVM), on clusters of two or more FSVMs, can fail over to another FSVM. High Availability is enabled by default on all clusters of two or more FSVMs.

When an FSVM experiences an issue, Files reassigns the IP of the FSVM to another FSVM in the cluster. The IP of the out-of-service FSVM remains available. However, the shares and exports on the impacted FSVM are unavailable for several minutes during a failover.

Affinity rules do not affect HA; multiple FSVMs can share a single host during a HA event.

Smart Tiering

Tier data to an object store.

You can free up space on your file server by tiering data to an object store. You must configure tiering through Data Lens. However, you can also access the Tiering Dashboard on Data Lens from the Files Console. In the Files Console, go to **Data Management** > **Smart Tiering** and click **Manage on Data Lens.** For more details, see "Smart Tiering" in the *Data Lens User Guide*.

Self-Service Restore

Self-service restore (SSR) lets you open and copy a previous version of a file. For SMB you can use SSR to restore files.

With Self-Service Restore (SSR), Files takes snapshots of the stored cluster data at the share level. SSR exposes these snapshots to the share or export and lets you view or restore a file from any of the previous snapshots without an administrator. The snapshots are read-only and point-in-time (snapshots taken at a certain time) copies.

SSR is disabled by default, but you can enable it during or after share creation.

Files supports 24 hour (every hour), daily, weekly, and monthly snapshots on a fixed schedule. By default, SST takes a snapshot every hour, retains the most recent 24 snapshots, and deletes the oldest SSR snapshot after exceeding the retention count for the snapshot type. Schedule snapshots for regular or frequent intervals to provide same-day protection against accidental deletions.

The snapshot retention count corresponds to the retention period, which are as follows:

- 24 hours for hourly snapshots
- 7 days for daily snapshots
- 4 weeks for weekly snapshots
- 3 months for monthly snapshots

For example, when the snapshot count for daily snapshots is 7, Files deletes the oldest snapshot and creates a new one every day.

You can view removed or overwritten files and choose a snapshot from the history of a share or export. For SMB, you can restore files through the system. For NFS, Files only provides read-access and you must manually perform restoration. Admins can configure snapshots schedules at a file-server-level that are applicable to all shares and exports in the file server. Currently, it is not possible to configure unique SSR schedules for shares.

Files supports share updates for both standard and distributed shares. To enable SSR during share creation, see Creating a Share (SMB) on page 64 or Creating an Export (NFS) on page 70. To enable SSR after share creation, see Enabling Self-Service Restore on page 146.

Limitations

Consider the following limitations before enabling SSR.

- SSR for SMB does not restore streams or attributes in directories.
- Files does not support SSR at the root of distributed shares or exports.

Enabling Self-Service Restore

Enable self-service restore on a share or export.

About this task

Follow the steps as indicated.

Procedure

1. In the Files Console, go to the **Shares** tab.
2. In the row for the target share, click **three dot menu** > **edit**.
3. In the **Update Share** window, click **Next** to go to the **Settings** tab.

4. For the **Enable Self-Service Restore** box, do one of the following:

   » To enable self-service restore, check the box.

   » Otherwise, to disable self-service restore, clear the box.

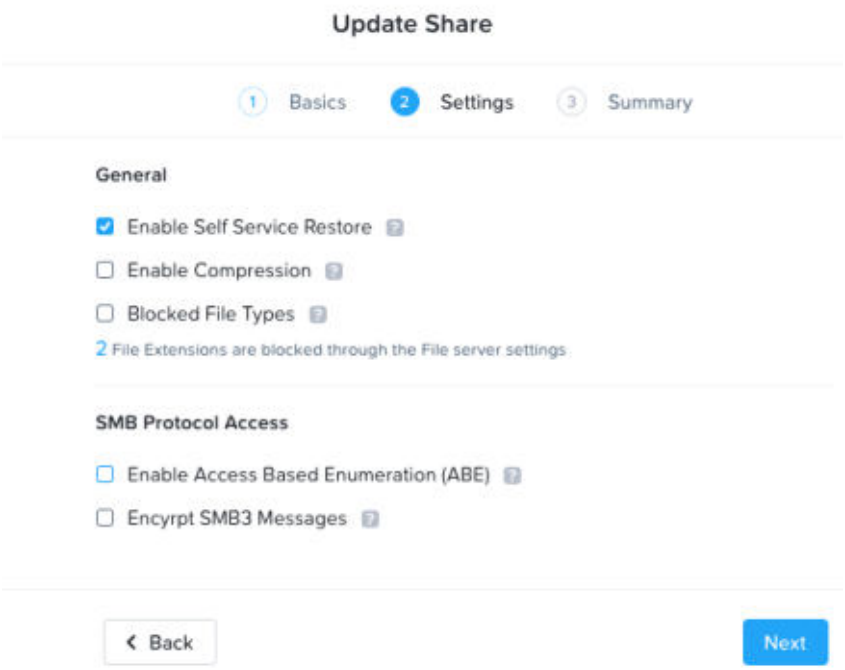


Figure 84: Enabled Self-Service Restore

5. Click **Next** > **Save**

What to do next

Add a snapshot schedule for SSR, see Adding Snapshot Schedules on page 147.

Adding Snapshot Schedules

Add file server protection by adding a snapshot schedule.

About this task

Use self-service restore (SSR) to create snapshot schedules to protect the file server. You can change the snapshot intervals but cannot change the specific time when Files takes the snapshot. Files takes snapshots at 00:00 UTC (midnight) for daily snapshots and at 0 minutes for hourly snapshots.

Before you begin

Ensure that the file server shares have SSR enabled.

> **Note:** The limit for configured snapshots is 50 or all schedule types.

Procedure

1. In the **Files Console**, go to **Data Management** > **Protection** > **Self Service Restore**.
   The **Self Service Restore** window displays.

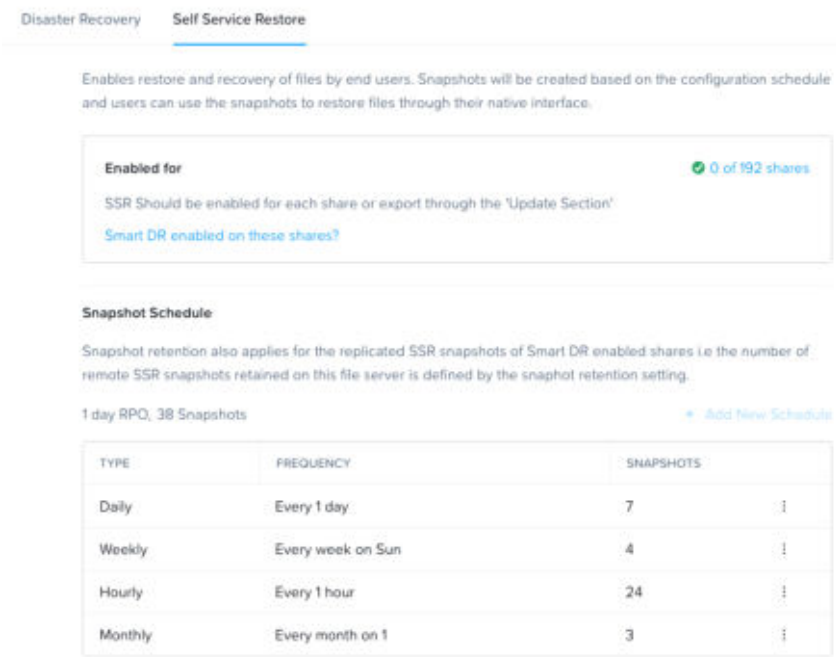


Figure 85: Add snapshot schedule

2. Click **+ Add New Schedule** and enter schedule details in the indicated fields.

   a. **Type**: Set the schedule interval. The snapshot types include hourly, daily, weekly, and monthly.

   > **Note:** You can only have one schedule type per file server. For example, a single file server cannot have two hourly schedules.

   The schedule does not have an end date by default, and the schedule runs indefinitely unless you enter an end date.

   - If you select weekly, select boxes for the days of the week to run the schedule. Select the boxes in **Pick days of week**.
   - If you select monthly, enter one or more integers (in a comma-separated list) to indicate which days in the month to run the schedule. For example, to run the schedule on the 1st, 10th, and 20th days, enter "1,10,20".

   b. **Frequency**: Enter the number of snapshots to occur within that type of schedule in the box. Enter the value in numerical format for minutes, hours, or days.
   The interval cannot be less than an hour, so the minutes value must be at least 60.

   > **Note:** The frequency field only supports hourly and daily schedules.

   c. **Snapshots**: Enter the number of snapshots to retain for this schedule. Enter the value in numerical format.

3. Click the **blue check mark icon** to add the schedule.
   The new schedule appears in the **Snapshot Schedule** table.

4. To edit an existing snapshot schedule, click the **three dots menu** > **edit** icon.

   a. To delete an existing schedule, click **three dots menu** > **delete**. Deleting schedules ages out the snapshots based on the schedule type. For example, Files deletes hourly schedules every hour until complete.

Retrieving Files (SMB Only)

This task describes how to retrieve files from file shares using self-service restore.

About this task

Procedure

1. To access the file share, go to the target directory and select **Properties** > **Previous Versions**.

   Previous versions of files display in order of date modified. Therefore, unmodified files do not appear. Previous versions of folders display every available snapshot.

   To see deleted files or directories, access a previous version of the parent folder and restore the file or directory.

2. Open and manage the previous versions according to your vendor documentation.

Retrieving Files (NFS Only)

Follow this procedure to retrieve snapshots of earlier file versions using Self-Service Restore (SSR).

About this task

Enabling SSR on NFS exports lets you access snapshot versions of a directory using a hidden `.snapshot` subdirectory. Browse the directory for the intended snapshot and restore it by manually copying the file or its content.

> **Note:** When the absolute path to the directory of the snapshot is longer than 3922 characters, the attempt to browse into the snapshot can fail with a `file name too long` error message.

Follow these steps to restore a snapshot version from an NFS client.

Procedure

1. Go to the snapshots of the target directory by replacing `/dir1/dir2` with the directory path.

   ```
   $ cd /dir1/dir2/.snapshot
   ```

   > **Note:** You can only get to the `.snapshot` directory using the `cd` command. The `.snapshot` directory is not visible otherwise. Using the `ls -a` command does not show the `.snapshot` directory.

2. List the snapshots for the target directory.

   ```
   $ ls
   ```

3. Browse through the snapshots and copy the desired data.

Deleting SSR Snapshots

Manually delete SSR snapshots to reclaim disk space.

About this task

By default, Files deletes the oldest SSR snapshot after exceeding the retention count for that snapshot type. To manually delete SSR snapshots, follow the steps as indicated.

Procedure

1. List the snapshots by creation time.

```
nutanix@fsvm$ afs snapshot.list share_name=share_name
```

2. List the space occupied by a single snapshot or by a range.

```
nutanix@fsvm$ afs snapshot.reclaimable_space uuid_start:uuid_end
```

3. Remove snapshots.

   » Remove a single snapshot using the universally unique identifier (UUID).

   > **Note:** You can delete multiple snapshots by specifying a comma-separated list of UUIDs.

   ```
   nutanix@fsvm$ afs snapshot.remove share_name=share_name snapshot_uuid_list=snapshot_uuid
   ```

   » Remove multiple snapshots using labels by defining the label. Labels define if the snapshot schedule is hourly, daily, weekly, or monthly. Deleting snapshots using labels deletes all snapshots for the specified label on the share.

   ```
   nutanix@fsvm$ afs snapshot.remove share_name=share_name label=label
   ```

Setting Custom Snapshot Times

Set a custom hour for SSR snapshots.

About this task

Files takes SSR snapshots at 00:00 UTC (midnight) for daily snapshots and at 0 minutes for hourly snapshots, taken 24 times per day. To take snapshots at a custom time, indicate the time by specifying the number of hours offset from midnight UTC time.

Procedure

Replace `1 - 23` with an integer to specify the time offset from UTC. For example, replace `1 - 23` with `2` to offset the time by 2 hours.

```
nutanix@fsvm$ afs snapshot.set_ssr_hourly_offset 1 - 23
```

SECURITY HARDENING

Use the Nutanix command-line interface (nCLI) or Files CLI to customize your Files security configuration.

> **Note:** Hardening Files with the settings described in this section requires AOS 5.19.2 (or later) and Files 3.8.1 (or later).

Table 21: General Files Hardening Configuration

Description	Command or settings
Support file server configuration of the SCMA policy.	`ncli file-server get-security-config fs-name=`*`file-server-name`
Schedule weekly execution of advanced intrusion detection environment (AIDE)	`ncli file-server edit-security-params fs-name=`*`file-server-name` ` enable-aide=true`
Enable the strong password policy.	`cli file-server edit-security-params fs-name=`*`file-server-name` ` enable-high-strength-password=true`
Enable the Department of Defense knowledge consent banner of the US department.	`ncli file-server edit-security-params fs-name=`*`file-server-name` ` enable-banner=true`
Change the default schedule of running the SCMA. The schedule can be hourly, daily, weekly, and monthly.	`ncli file-server edit-security-params fs-name=`*`file-server-name` ` schedule=hourly`
Disable the core-dump settings to let the file server VM generate stack traces for any cluster issue. **Note:** On a file server, this parameter turns both the `core` and the `kerneldump` salt status on or off.	`ncli file-server edit-security-params fs-name=`*`file-server-name` ` enable-core=false`

When a high governance official must run the hardened configuration.	Have the following settings.
	`Enable Aide : true`
	`Enable Core : false`
	`Enable High Strength Password : true`
	`Enable Banner : false`
	`Schedule : HOURLY`

When a federal official must run the hardened configuration.	Have the following settings.
	`Enable Aide : true`
	`Enable Core : false`
	`Enable High Strength Password : true`
	`Enable Banner : true`
	`Schedule : HOURLY`

Description	Command or settings
Back up the DoD banner file.	Run the following command, and repeat on all FSVMs. `nutanix@FSVM$ sudo cp -a /srv/salt/security/AFS/sshd/DODbanner /srv/salt/security/AFS/sshd/DODbannerbak`
Modify DoD banner file.	Run the following command, and repeat on all FSVMs. `nutanix@FSVM$ sudo vi /srv/salt/security/AFS/sshd/DODbanner`

Create a separate mount for audit logs on new file servers.	`nutanix@cvm$ allssh ‘echo “--audit_logs_separate_fs=true” >> ~/config/minerva_cvm.gflags’`
	`nutanix@cvm$ allssh ‘genesis stop minerva_cvm’`
	`nutanix@cvm$ cluster start`

Manage sudo authentication settings.	**Note:** Running the commands from a single FSVM applies the changes to all FSVMs on the cluster.
	Enable sudo authentication settings.
	`nutanix@fsvm$ afs infra.set_sudo_auth state=on`
	Disable sudo authentication settings before performing Files upgrades, expanding a files cluster, and before using the following commands: `genesis restart`, `genesis stop`, `cluster stop`, and `cluster restart_genesis`.
	Disable sudo authentication settings.
	`nutanix@fsvm$ afs infra.set_sudo_auth state=off`
	After performing file server activate, clone, or migrate operations, re-enable sudo authentication.
	`nutanix@fsvm$ afs infra.set_sudo_auth state=on`

Enable SMB signing.	Run the following command on any FSVM in the cluster. `nutanix@fsvm$ afs smb.set_conf "server signing" "mandatory" section=global`

Secure Sockets Layer (SSL) Certificate Settings

Files supports installing a self-signed or custom SSL certificate for certificate-based authentication. For file servers, these operations are only available through the nCLI and afs-CLI commands.

Table 22: SSL Certificate Settings on a File Servers

Description	Command or Settings
Generate a new self-signed SSL certificate.	Run the following command. **Note:** Replace *fs_uuid* with the universally unique identifier of the file server. `nutanix@CVM$ ncli file-server ssl-certificate-generate uuid=fs_uuid`
Replace the existing self-signed SSL certificate with a new one.	Run the following command. `nutanix@CVM$ ncli file-server ssl-certificate-change-pfx-file-password uuid=fs_uuid`
Get the current SSL certificate.	Run the following command. `nutanix@CVM$ file-server get-ssl-certificate`
Apply a custom SSL certificate on an FSVM.	Run the following command. • Replace *ca-path* with the CA certificate or chain file path. • Replace *cert-path* with the SSL certificate file path. • Replace *key-path* with the private key path. • Replace *value* with the type of private key (`ECDSA_256`, `ECDSA_384`, `ECDSA_521`, `RSA_2048`). `nutanix@FSVM$ afs misc.import_ssl ca_chain_file=ca_path cert_file=cert-path key_file=key-path key_type=value`

Quality of Service (QoS) Setup

The table following details QoS setup on a file server using differentiate services code point (DSCP) values.

Note: QoS requires the `--qos_enabled=True` gflag .

Table 23: QoS Settings

Description	Command or settings
Enable QoS. **Note:** The default management traffic value is 16, and the default data traffic value is 10.	`nutanix@FSVM$ afs net.enable_qos [management_dscp_value=value] [data_dscp_value=value]`
Get the currently configured DSCP values.	`nutanix@FSVM$ afs net.get_qos`
Modify the DSCP value of one or more traffic types.	`nutanix@FSVM$ afs net.edit_qos [management_dscp_value=value] [data_dscp_value=value]`
Disable QoS on all traffic types and delete the existing configuration.	`nutanix@FSVM$ afs net.disable_qos`

Rsyslog

The table following provides details on rsyslog daemon configuration for log forwarding.

Table 24: Rsyslog

Description	Command or settings
Add a rsyslog server.	`nutanix@FSVM$ afs rsyslog.add_server server_name=server_name ip_address=server_ip port=server_port network_protocol=TCP/UDP relp_enabled=true/false`
Modify the properties of the rsyslog server.	`nutanix@FSVM$ afs rsyslog.update_server server_name=configured_server_name ip_address=server_ip port=server_port network_protocol=TCP/UDP relp_enabled=true/false`
Remove the configured rsyslog server.	`nutanix@FSVM$ afs rsyslog.remove_server server_name=configured_server_name`
Add forwarded modules and their debug-level to the rsyslog server. **Note:** Files only supports the syslog module.	`nutanix@FSVM$ afs rsyslog.add_server_module server_name=configured_server_name module_name=SYSLOG_MODULE level=ALERT/ CRITICAL/ DEBUG/ EMERGENCY/ ERROR/ INFO/ NOTICE/ WARNING`
Remove the module and forwarding level form the rsyslog server.	`nutanix@FSVM$ afs rsyslog.remove_server_module server_name=configured_server_name module_name=SYSLOG_MODULE`
Set the status for rsyslog forwarding. (You can disable rsyslog forwarding completely without removing the configured server details.) **Note:** Setting the *false* status ends forwarding and removes the configuration.	`nutanix@FSVM$ afs rsyslog.set_status enable=true/false`
Get the status of rsyslog forwarding.	`nutanix@FSVM$ afs rsyslog.get_status`
Get the current rsyslog configuration.	`nutanix@FSVM$ afs rsyslog.list`
Get the configured modules and their forwarding level for a configured rsyslog server.	`nutanix@FSVM$ afs rsyslog.list_modules server_name=server_name`
Set up transport layer security (TLS) parameters of the rsyslog server. **Note:** Not setting the auth-mode to *anon*, requires specifying permitted peers in a comma-separated list.	`nutanix@FSVM$ afs rsyslog.set_tls auth_mode=anon, x509/certvalid, x509/fingerprint, x509/name ca_chain_path=PEM_encoded_CA_certificate_file_absolute_pat permitted_peers=permitted_peers`
Get the configured TLS parameters.	`nutanix@FSVM$ afs rsyslog.get_tls`
Disable all configured TLS parameters, and disable TLS on forwarded packets.	`nutanix@FSVM$ afs rsyslog.disable_tls`

TROUBLESHOOTING

Invalid Mounts After Authentication Change

Clients cannot mount share.

Changes to the authentication method.

Procedure

Remount the authentication type as a value for parameter sec.

```
host$ -o sec=authentication-type
```

For example, use `-o sec=krb5` for Kerberos 5.

Client Access Denial (NFS Protocol)

Linux client experiences a "permission denied" error while accessing NFS shares.

A user management (authentication) change on an existing file server.

Procedure

This problem might be fixed by restarting the RPC-GSSAPI service on the clients. For example, enter the following command on a Linux CentOS 6 client:

```
nutanix@fsvm$ sudo service rpcgssd restart
```

The command syntax to restart the RPC-GSSAPI service varies among different Linux versions.

Clients Cannot Mount Shares

Clients in the same subnet as the Controller VM or in the storage network of the file server cannot mount the shares of the file server.

The file server was configured with two separate networks for the client-side and storage-side networks.

Procedure

To allow clients in the same subnet as the Controller VM or storage network to mount shares, configure the file server with the same network for both the client-side and storage-side networks.

Client Side Network Mapping

The file server client network does not map to a site on the AD and does not specify the client side.

The file server's client network does not map to any site on the AD. Files cannot find the local domain controllers in a multi-site AD environment and uses a geographically remote DC, which, can result in delayed domain operations.

Procedure

1. In a multi-site environment, map the Files client network to a local site in the AD.
2. In a single-site environment or with only a single geographic location, ignore the warning.

Connecting to Authentication Services

The file server cannot connect with the AD server or it cannot contact the LDAP server for the given domain.

The file server cannot reach the given domain name with the specified DNS server list. Possible reasons include spelling mistakes in the domain name, incorrect DNS name servers, or connectivity issues with the domain controller servers.

Procedure

1. Check the DNS server addresses, domain name, and status of the domain controllers.
2. Verify the DNS entries for the given domain name.

Constraint Violation

Domain controller reused an operation due to a possible constraint violation.

- Incorrect SPN configuration.
- An SPN is not unique in the forest and the conflict results in failure.

Procedure

1. Ensure that Files related SPN entries are not present in the forest.
2. Ensure that the domain controllers do not have any replication issues.

DNS Missing SRV Records

SRV records not found on the specified DNS servers.

The specified DNS servers do not resolve the SRV records with the appropriate domain controller names.

Procedure

Add the domain controller SRV records for the required protocols and services.

Domain Controller Issues

Cannot find a writable and reachable domain controller.

Files cannot discover an active non-RODC (writable) LDAP domain controller (at the site or domain level).

Procedure

Ensure that one writeable domain controller is working in the given domain.

Finding IP Addresses

Fetch the IP addresses for all FSVMs.

Various causes.

Procedure

Enter the following command from any FSVM: `nutanix@fsvm$ afs misc.fsvmips`
Output lists the IP addresses for the FSVMs in the node.

Identifying the Share Owner

Identify the owner of a share or export.

The Files UI does not list the name of share and export owners.

Procedure

Perform one of the following commands to identify the share owner.

» List the share owners of a standard share or export.

```
nutanix@fsvm$ afs share.owner_fsvm share-name
```

» List the share owners of a distributed share or export by specifying the *share-name* and the *share-path*(including the name of the top-level directory TLD).

```
nutanix@fsvm$ afs share.owner_fsvm share-name path=share-path
```

Invalid Credential

Invalid user name or password.

Files cannot authenticate on the AD using the given user name and password combinations.

Procedure

1. Ensure that the user name and password are correct.
2. Verify that the user is not expired, locked, or disabled.

NLM Locks

Unable to get Network Lock Manager (NLM) locks from Mac client.

NLM recovery does not work over the User Datagram Protocol (UDP). Use the transmission Control Protocol (TCP) instead.

Procedure

1. Add the following lines to the `/etc/nfs.conf` file:

```
nfs.lockd.send_using_tcp = 1
nfs.statd.send_using_tcp = 1
```

2. Restart services.

```
user@host$ launchctl stop com.apple.lockd ; launchctl start com.apple.lockd
```

```
user@host$ launchctl stop com.apple.statd ; launchctl start com.apple.statd
```

Network Cannot Expand

You changed the network ID and the file server cannot expand.

Update the UUIDs.

Procedure

1. Log into the CVM.
2. To update the file server network, enter the following:

```
nutanix@cvm$ afs infra.update_file_server_network
```

3. To update a specific file server within the network, enter the following:

```
nutanix@cvm$ afs infra.update_file_server_network fs_name
```

NTLM Authentication Issues

Authentication might be unsuccessful for NTLM when contacting read-only domain controllers (RODC).

The list of allowed password replication must include the machine account name or file server name. To resolve, follow troubleshooting steps on a domain controller.

Procedure

1. Add the host name.

```
C:\>repadmin /prp add RODC_host_name allow machine_account_DN
```

2. View the list of added names.

```
C:\>repadmin /prp view RODC_host_name allow
```

For example, viewing the list of added names would appear similar to the following:

```
C:\>repadmin /prp view MINDC03 allow
```

Output looks similar to the following:

```
Allow list (msDS-RevealOnDemandGroup):
RODC "CN=MINDC03,OU=Domain Controllers,DC=automation,DC=nutanix,DC=com":
CN=MNRVATST124803,CN=Computers,DC=automation,DC=nutanix,DC=com
CN=Allowed RODC Password Replication Group,CN=Users,DC=automation,DC=nutanix,DC=
com
```

Share Copying

Copy operation interrupted while using Microsoft robocopy to copy large files to Files shares.

Various causes (for example network bandwidth issues).

Procedure

Use robocopy with the `/z` option. This option resumes any interrupted copy operation.

Stale Statistics

Windows client experiences stale statistics.

Default cache entry time is approximately 5 minutes.

Procedure

To change the default cache entry time of 5 minutes, log into the FSVM and run the following :

```
nutanix@fsvm$ afs smb.set_conf “stats cache ttl” “value” section=global
```

Time Difference

A time difference exists between Files and the domain controller.

Files uses Kerberos protocol for authentication on the AD. Kerberos is a time sensitive protocol and cannot sync the correct time when the client and servers are out of sync for several minutes.

Procedure

Use the same NTP server for the domain controller and Files.

Unsuccessful Authentication

Authentication might be unsuccessful for the NT LAN manager (NTL when contacting read-only domain controllers (RODC).

The list of allowed password replication must include the machine account name or file server name. To resolve, follow troubleshooting steps on a domain controller.

Procedure

1. Add the host name.

   ```
   C:\>repadmin /prp add RODC_host_name allow machine_account_DN
   ```

2. View the list of added names.

   ```
   C:\>repadmin /prp view RODC_host_name allow
   ```

   For example, viewing the list of added names would appear similar to the following:

   ```
   C:\>repadmin /prp view MINDC03 allow
   ```

   Output looks similar to the following:

   ```
   Allow list (msDS-RevealOnDemandGroup):
   RODC "CN=MINDC03,OU=Domain Controllers,DC=automation,DC=nutanix,DC=com":
   CN=MNRVATST124803,CN=Computers,DC=automation,DC=nutanix,DC=com
   CN=Allowed RODC Password Replication Group,CN=Users,DC=automation,DC=nutanix,DC=
   com
   ```

COPYRIGHT

Copyright 2022 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

Platform ANY

NX Series Hardware Administration Guide

March 3, 2022

Contents

System Specifications and Hardware Information..... v

Firewall Port Requirements for IPMI.....vi

Product Mixing Restrictions.....vii

Visio Stencils..... xii

1. Manual Firmware Updates..... 13
- Shutting Down a Single-node Cluster..... 13
  - Node Shutdown Prechecks.....13
  - Preparing a Single Node for Shutdown.....14
  - Shutting Down a Single Node (vSphere Client)..... 14
  - Shutting Down a Single Node (AHV).....15
- Shutting Down a Node in a Multinode Cluster..... 15
  - Node Shutdown Prechecks.....16
  - Preparing Nodes for Shutdown..... 16
  - Shutting Down a Node in a Cluster (vSphere Client)..... 17
  - Shutting Down a Node in a Cluster (vSphere Command Line).....17
  - Shutting Down a Node in a Cluster (AHV)..... 18
  - Shutting Down a Node in a Cluster (Hyper-V).....19
- Manually Updating M.2 RAID Hypervisor Boot Drive Firmware..... 21
- Manually Updating NVMe Drive Firmware..... 27
- Manually Updating Data Drive Firmware..... 33
- Manually Updating SATA DOM Firmware..... 39
- Manually Updating the BMC and BIOS..... 43
- Manually Updating HBA Controller Firmware..... 43
- Manually Updating NIC Firmware..... 46
- Starting a Single-node Cluster..... 47
  - Starting a Single Node (vSphere Client)..... 47
  - Starting a Single Node (AHV).....48
  - Node Startup Post-check..... 49
- Starting a Node in a Multinode Cluster.....50
  - Starting a Node in a Cluster (vSphere Client).....50
  - Starting a Node in a Cluster (AHV)..... 51
  - Starting a Node in a Cluster (Hyper-V).....53
  - Node Startup Post-check..... 55

2. Changing an IPMI IP Address..... 56
- Configuring the Remote Console IP Address (IPMI Web Interface)..... 57
- Configuring the Remote Console IP Address (Command Line)..... 57
- Configuring the Remote Console IP Address (BIOS).....59

3. Changing the IPMI Password 61
IPMI Password Requirements for Nutanix G8 Platforms 61
Changing the IPMI Password for ESXi 62
Changing the IPMI Password for Hyper-V 62
Changing the IPMI Password for AHV 63
Changing the IPMI Password without Operating System Access 63

4. Failed Component Diagnosis 64
SATA SSD Boot Drive 64
Rescue Shell 64
SATA HDD or SATA SSD Data Drive 68
Node 69
Chassis or Node Fan 69
Memory 70
Power Supply 71
Chassis 72

5. Adding a Drive 73

6. Remote Direct Memory Access 77

7. CMOS Battery Replacement 79

8. Memory Configurations 80
Supported Memory Configurations (G4 and G5 Platforms) 80
Supported Memory Configurations (G6 and G7 Platforms) 84
Supported Memory Configurations 86

9. Host Secure Boot (UEFI) 91
Host Secure Boot Overview 91
Enabling Host Secure Boot on Nutanix NX Platforms (Redfish API Method) 91
Enabling Host Secure Boot on Nutanix NX Platforms (BIOS Method) 93
Secure Boot Error Handling and Failure Conditions 96
Verifying Enabled Host Secure Boot Settings With the Redfish API or BIOS 97
Verifying Enabled Host Secure Boot Settings in the Redfish API 97
Imaging with Foundation on a Host Secure Boot Enabled Node 99
Verifying Enabled Host Secure Boot Settings With the Hypervisor or CVM 100
Disabling Host Secure Boot on Nutanix NX Platforms 100
Verifying the Disabled Host Secure Boot Settings in the Redfish API (Optional) 102
Confirming the Disabled Secure Boot Status from the Hypervisor and CVM 105

Copyright 106
License 106

Conventions..... 106
Default Cluster Credentials..... 106
Version.....107

SYSTEM SPECIFICATIONS AND HARDWARE INFORMATION

Find system specifications.

For system specifications, wiring diagrams, and other platform-specific hardware information, see the system specifications. Go to the Nutanix Support portal and select **Documentation** > **Hardware Compatibility Lists** > **NX Platforms**. Select your NX generation to see the system specifications.

FIREWALL PORT REQUIREMENTS FOR IPMI

Port numbers for the IPMI interface. Ensure that these ports are open on the firewall.

Interface	Port number
HTTP	80 (TCP)
HTTPS	443 (TCP)
IPMI	623 (UDP)
Remote console	5900 (TCP)
Virtual media	623 (TCP)
SMASH	22 (TCP)
WS-MAN	8889 (TCP)
SSH to virtual media	5120 (TCP)
AHV console	2937 (TCP)
Remote console (AHV over IP)	5900 (TCP)
Video (remote console)	5901 (TCP)
CD (remote console)	5120 (TCP)
Floppy (remote console)	5123 (TCP)

PRODUCT MIXING RESTRICTIONS

Hardware and software restrictions for Nutanix platforms and clusters.

> CAUTION: Do not configure a cluster that violates any of the following rules.

Compatibility

The Nutanix Support portal includes a compatibility matrix, available from the Compatibility Matrix link. Filter and display compatibility by Nutanix NX model, AOS release, hypervisor, or feature (platform/cluster intermixing).

Nutanix recommends that you consult the matrix before installing or upgrading software on your cluster.

Hardware Mixing Restrictions

- Nutanix clusters can support four generations of Nutanix platforms. For example, you can mix G8 nodes in a cluster with G7, G6, and G5 platforms, but not with G4 or earlier platforms.
- You can mix nodes that use different CPU families in the same cluster, but not in the same block. For example, a cluster can contain nodes from any NX generation, but a G7 block must contain only G7 nodes, a G8 block must contain only G8 nodes, and so on.
- Nutanix does not support mixing Nutanix NX nodes in the same cluster with nodes from any other hardware vendor. However, you can manage separate clusters using the Prism web console regardless of the hardware type.
- Nutanix does not support mixing nodes with RDMA enabled and nodes without RDMA enabled in the same cluster.

  For hardware vendors other than Nutanix (NX generation), you can mix nodes from the same hardware vendor product series in the same cluster. You cannot mix nodes from different hardware vendors in the same cluster. However, you can manage separate clusters using the Prism or Prism Central web console regardless of vendor.

  Consider these mixing examples.

- You can mix Cisco UCS (B and C Series) nodes in the same cluster.
- You cannot mix Cisco UCS (B and C Series) nodes and Nutanix NX Series nodes in the same cluster.
- You cannot mix Dell PowerEdge and Dell XC Series nodes in the same cluster.
- Mixing of Intel nodes and AMD nodes in the same cluster is not supported.
- You can mix HPE DX and non-DX HPE HFCL-certified platforms.

Operating System Restrictions

- All Controller VMs in a cluster must use the same version of AOS.

CPU Restrictions

- All CPUs in a block must be identical. When adding a node to a multinode block, make sure the new node and the existing nodes use identical CPUs.

NIC Restrictions

- All installed NICs in a node must be identical.
- Nutanix does not support third-party SFP+ to Base-T transceivers.
- Nutanix does not support long range SFP+ transceivers.

Fan Restrictions

- Nutanix does not support fan redundancy in any platform. You must replace a failed fan immediately.

Storage Restrictions

- Do not move HDD or SSD drives from one node to another.
- If you configure a cluster, and later disassemble the chassis (for example, to move the hardware to a different site), then when you reassemble the chassis, you must replace all drives in their original nodes. If you move a drive from one node to another, cluster services may not come back up correctly.
- You can mix all-SSD nodes and hybrid SSD/HDD nodes in the same cluster only under the following conditions:
  - Mixed all-flash/hybrid clusters must have a minimum of two all-SSD nodes.
  - The minimum number of each node type (all-SSD/hybrid) must be equal to the cluster redundancy factor. For example, clusters with a redundancy factor of 2 must have a minimum of two hybrid and two all-SSD nodes.
  - You can run guest VMs on the all-SSD nodes in clusters that also have hybrid nodes, provided that the cluster includes two or more all-SSD nodes.
- You can mix all-SSD nodes in the same cluster with all-NVMe nodes or mixed SSD/NVMe nodes.
- You can mix all-NVMe nodes in the same cluster with all-SSD nodes or mixed SSD/NVMe nodes.
- Nutanix does not support mixing hybrid SSD/HDD nodes in the same cluster with all-NVMe nodes or mixed SSD/NVMe nodes.

Encryption Restrictions

- Encrypted drives (SED) can be mixed with unencrypted (non-SED) drives in the same node, if encryption has never been enabled.

  Encrypted nodes can be mixed with unencrypted nodes in the same cluster if encryption was never enabled and remains disabled.
- NVMe drives do not support self-encryption (SED).

DIMM Restrictions

DIMM types

For all platforms: within a node, all DIMMs must be of the same type. For example, you cannot mix RDIMMs and LRDIMMs in the same node.

DIMM capacity

For all platforms: within a node, all DIMMs must have the same memory capacity. For example, you cannot mix 16 GB and 32 GB DIMMs in the same node.

DIMM manufacturers

For platforms earlier than G5, you cannot mix DIMMs from different manufacturers in the same node.

For G5 and later platforms, Nutanix supports mixing DIMMs from different manufacturers within the same node, but not within the same channel.

- DIMM slots are arranged on the motherboard in groups called channels.
  - On G5 platforms, channels contain either two DIMM slots (one blue and one black) or three DIMM slots (one blue and two black.)
  - On G6 and G7 platforms, all channels contain two DIMM slots (one blue and one black.)
  - On G8 platforms, channels contain either one DIMM slot (blue) or two DIMM slots (one blue and one black.)
- Within a channel, all DIMMs must be from the same manufacturer.
- When replacing a failed DIMM, ensure that you are replacing the original DIMM like-for-like.
- When adding new DIMMs to a node, if the new DIMMs and the original DIMMs are from different manufacturers, arrange the DIMMs so that the original DIMMs and the new DIMMs are not mixed in the same channel.

  Note: You do not need to balance numbers of DIMMs from different manufacturers within a node, so long as you never mix them in the same channel.

DIMM speed

For platforms earlier than G5, you cannot mix DIMMs that run at different speeds in the same node.

For G5 and later platforms, Nutanix supports higher-speed replacement DIMMs. You can mix DIMMs that use different speeds in the same node, and in the same memory channel, under the following conditions.

- You can only use higher-speed replacement DIMMs from one NX generation later than your platform.

- You can mix DIMMs that use different speeds in the same node but not in the same channel. Within a channel, Nutanix supports only the following mixes of DIMM speeds:
  - G5 platforms:
    - Samsung 2400 MHz B-die and Samsung 2666 MHz B-die
  - G6 platforms:
    - Samsung 2666 MHz B-die and Samsung 2666 MHz C-die
    - Samsung 2666 MHz C-die and Samsung 2933 MHz C-die
    - Samsung 2666 MHz B-die and Samsung 2933 MHz C-die
  - G7 platforms:
    - Samsung 2933 MHz 32 GB and Samsung 3200 MHz 32 GB
    - Samsung 2933 MHz 64 GB and Samsung 3200 MHz 64 GB
    - Micron 2933 MHz 32 GB and Micron 3200 MHz 32 GB
    - Micron 2933 MHz 64 GB and Micron 3200 MHz 64 GB
  - G8 platforms ship with 3200 MHz DIMMs. 3200 MHz is the highest DIMM speed that Nutanix currently supports, so you cannot currently mix DIMM speeds in any G8 platform.

  Note: When you mix DIMMs of different speeds in the same node, your system operates at the lowest common DIMM speed or CPU supported frequency.

Hypervisor Restrictions

- All nodes in a cluster must use the same hypervisor type and version. This restriction does not apply if the cluster contains the following nodes:
  - Storage-only nodes always run AHV, but you can add them to clusters that run on other hypervisors.

    Note: With Nutanix Foundation version 4.0 and later, any node can act as a storage-only node. With Foundation versions earlier than 4.0, only dedicated storage platforms such as the NX-6035C can act as storage-only nodes.

    Note: Citrix Hypervisor nodes do not support storage-only nodes.
  - A cluster consisting of only NX-1065 series nodes or only Lenovo HX1310 nodes. These nodes can form a mixed-hypervisor cluster running ESXi and AHV, with the AHV node used for storage only. For information about creating a multi-hypervisor cluster, see the *Field Installation Guide: Imaging Bare Metal Nodes*.
  - If you expand a cluster by adding a node with older generation hardware to a cluster that was initially created with later generation hardware, power cycle (do not restart)

any guest VMs before migrating them to the added older generation node or before upgrading the cluster.

Guest VMs are migrated during hypervisor and firmware upgrades (but not AOS upgrades).

For example, if you are adding a node with G6 Skylake CPUs to a cluster that also has newer G7 nodes with Cascade Lake CPUs, you must power cycle guest VMs hosted on the G7 nodes before you can migrate the VMs to the node with G6 CPUs. Power cycling the guest VMs enables them to discover G6 processor changes.

Power cycle guest VMs from the Prism web console VM dashboard. Do not perform a Guest Reboot; this case requires a VM power cycle.

- G8 platforms do not currently support Hyper-V.

vSphere Restrictions

- Nutanix supports mixing nodes with different processor architectures in the same cluster. However, vSphere only supports enhanced/live vMotion of VMs from one type of node to another when you have enhanced vMotion compatibility (EVC) enabled. For more information about EVC, see the vSphere 5 documentation and the following VMware knowledge base articles:

  - *Enhanced vMotion Compatibility (EVC) Processor Support* [1003212]
  - *EVC and CPU Compatibility FAQ* [1005764]

  Enabling EVC on an existing vCenter cluster requires shutting down all VMs. Because EVC is a vSphere cluster setting, it applies to all hosts at the same time. For these reasons, adding a node with a different processor type to a vCenter cluster requires shutting down the Nutanix cluster.

- *ENG-243562* AOS Controller VMs and Prism Central VMs require a minimum CPU micro-architecture version of Intel Sandy Bridge. For AOS clusters with ESXi hosts, or when deploying Prism Central VMs on any ESXi cluster: if you have set the vSphere cluster Enhanced vMotion Compatibility (EVC) level, the minimum level must be **L4 - Sandy Bridge**.

- Clusters are block-aware only under certain conditions. For complete explanation of the requirements, see the *Block Fault Tolerance* topic in the *Prism Web Console Guide*.

VISIO STENCILS

Visio stencils for Nutanix products are available on VisioCafe.

1

MANUAL FIRMWARE UPDATES

Nutanix recommends that you use the Prism Life Cycle Manager to perform firmware updates. However, firmware must be updated manually if:

- The component is located in a single-node cluster, OR
- The component is located in a multi-node cluster, but the hypervisor the cluster is running does not support LCM firmware updates.

To find the latest binary firmware files, consult KB 6937.

Shutting Down a Single-node Cluster

Before updating firmware in a single-node cluster, shut down the node, following the method for your hypervisor.

> Note: Single-node clusters do not support Hyper-V.

These topics cover single-node cluster shutdown:

- Node Shutdown Prechecks: Node Shutdown Prechecks on page 13
- Preparation: Preparing a Single Node for Shutdown on page 14
- vSphere Web Client: Shutting Down a Single Node (vSphere Client) on page 14
- AHV: Shutting Down a Single Node (AHV) on page 15

Node Shutdown Prechecks

Check to make sure that there are no issues that might prevent the node from being shut down safely. (Even if the node does not need to be shutdown these checks are helpful.)

About this task

Procedure

1. In Prism, go to the **Home** page and make sure **Data Resiliency Status** displays a green **OK**.
2. In Prism, go to the **Health** page and select **Actions** > **Run NCC Checks**.
3. In the dialog box that appears, select **All Checks** and click **Run**.

   Alternatively, issue the following command from the CVM:

   ```
   nutanix@cvm$ ncc health_checks run_all
   ```

4. If any checks fail, see the related KB article provided in the output and the *Nutanix Cluster Check Guide: NCC Reference* for information on resolving the issue.

5. If you have any unresolvable failed checks, contact Nutanix Support before shutting down the node.

6. Gather component details by running the NCC `show_hardware_info` command.

```
nutanix@cvm$ ncc hardware_info show_hardware_info
```

7. Save the output of the `show_hardware_info` command so that you can compare details when verifying the component replacement later.

Preparing a Single Node for Shutdown

Prepare the node for shutting down.

About this task

Procedure

1. Log on to the CVM and make a note of the BIOS, BMC, and SATA DOM versions.

2. Verify the versions from the configuration file `/etc/nutanix/firmware_config.json`.

```
cat /etc/nutanix/firmware_config.json
      {"satadom": {"model": "SATADOM-SL 3IE3 V2", "firmware_version": "S560301N"}, "bmc":
 {"model": "X10_ATEN", "firmware_version": "03.56"}, "motherboard_model": "X10SRW-F", "bios":
 {"model": "0833", "firmware_version": "20170425"}}
```

3. Find the IPMI IP address of the node (necessary in order to access the IPMI web UI.)

```
nutanix@cvm$ ipmiips
```

4. Shut down all running guest VMs on the cluster.

Shutting Down a Single Node (vSphere Client)

Before you begin

Shut down any guest VMs that are running on the node.

About this task

Procedure

1. Log on to the vSphere Client.

2. Shut down any VMs other than the Controller VM.

3. Right-click the host and select **Maintenance Mode** > **Enter Maintenance Mode**.

4. In the Enter Maintenance Mode dialog box, click **OK**.
   The host gets ready to go into maintenance mode, which prevents VMs from running on this host.

5. Log on to the Controller VM with SSH and shut down the Controller VM.

```
nutanix@cvm$ cvm_shutdown -P now
```

> Note: Do not reset or shutdown the Controller VM in any way other than the `cvm_shutdown` command to ensure that the cluster is aware that the Controller VM is unavailable

6. After the Controller VM shuts down, wait for the host to go into maintenance mode.
7. Right-click the host and select **Shut Down**.

   Wait until vCenter Server displays that the host is not responding, which may take several minutes. If you are logged on to the ESXi host rather than to vCenter Server, the vSphere web client disconnects when the host shuts down.

Shutting Down a Single Node (AHV)

Before you begin

Shut down guest VMs that are running on the node.

About this task

Procedure

1. If the Controller VM is running, shut down the Controller VM.
   a. Log on to the Controller VM with SSH.
   b. Run the command `acli host.list`.
   c. Note the value of **Hypervisor address** for the node.
   d. Put the node into maintenance mode.

      ```
      nutanix@cvm$ acli host.enter_maintenance_mode Hypervisor-address [wait="{ true |
       false }" ]
      ```

      Replace *`Hypervisor-address`* with the value of **Hypervisor address** for the node. Value of **Hypervisor address** is either the IP address of the AHV host or the host name.

   e. Shut down the Controller VM.

      ```
      nutanix@cvm$ cvm_shutdown -P now
      ```

2. Log on to the AHV host with SSH.
3. Shut down the host.

   ```
   root@ahv# shutdown -h now
   ```

Shutting Down a Node in a Multinode Cluster

Before updating firmware in a cluster, shut down the node, following the method for your hypervisor.

These topics cover node shutdown in a multinode cluster:

- Node Shutdown Prechecks: Node Shutdown Prechecks on page 13
- Preparation: Preparing Nodes for Shutdown on page 16
- vSphere Web Client: Shutting Down a Node in a Cluster (vSphere Client) on page 17
- vSphere Command Line: Shutting Down a Node in a Cluster (vSphere Command Line) on page 17
- AHV: Shutting Down a Node in a Cluster (AHV) on page 18

- Hyper-V: Shutting Down a Node in a Cluster (Hyper-V) on page 19

Node Shutdown Prechecks

Check to make sure that there are no issues that might prevent the node from being shut down safely. (Even if the node does not need to be shutdown these checks are helpful.)

About this task

Procedure

1. In Prism, go to the **Home** page and make sure **Data Resiliency Status** displays a green **OK**.
2. In Prism, go to the **Health** page and select **Actions** > **Run NCC Checks**.
3. In the dialog box that appears, select **All Checks** and click **Run**.

   Alternatively, issue the following command from the CVM:

   ```
   nutanix@cvm$ ncc health_checks run_all
   ```

4. If any checks fail, see the related KB article provided in the output and the *Nutanix Cluster Check Guide: NCC Reference* for information on resolving the issue.
5. If you have any unresolvable failed checks, contact Nutanix Support before shutting down the node.
6. Gather component details by running the NCC `show_hardware_info` command.

   ```
   nutanix@cvm$ ncc hardware_info show_hardware_info
   ```

7. Save the output of the `show_hardware_info` command so that you can compare details when verifying the component replacement later.

Preparing Nodes for Shutdown

Prepare each node in the cluster for shutting down.

About this task

Procedure

1. Log on to the CVM and make a note of the BIOS, BMC, and SATA DOM versions.
2. Verify the versions from the configuration file `/etc/nutanix/firmware_config.json`.

   ```
   cat /etc/nutanix/firmware_config.json
         {"satadom": {"model": "SATADOM-SL 3IE3 V2", "firmware_version": "S560301N"}, "bmc":
    {"model": "X10_ATEN", "firmware_version": "03.56"}, "motherboard_model": "X10SRW-F", "bios":
    {"model": "0833", "firmware_version": "20170425"}}
   ```

3. Find the IPMI IP address of each node (necessary in order to access the IPMI web UI.)

   ```
   nutanix@cvm$ ipmiips
   ```

4. Shut down all running guest VMs on the cluster.

Shutting Down a Node in a Cluster (vSphere Client)

About this task

> CAUTION: Verify the data resiliency status of your cluster. If the cluster only has replication factor 2 (RF2), you can only shut down one node for each cluster. If an RF2 cluster would have more than one node shut down, shut down the entire cluster.

Procedure

1. Log on to vCenter with the vSphere Client.

   If vCenter is not available, log on to the ESXi host IP address with vSphere.

2. If DRS is not enabled, manually migrate all the VMs except the Controller VM to another host in the cluster or shut down any VMs other than the Controller VM that you do not want to migrate to another host.

   If DRS is enabled on the cluster, you can skip this step.

3. Right-click the host and select **Maintenance Mode** > **Enter Maintenance Mode**.

4. In the Enter Maintenance Mode dialog box, click **OK**.

5. In the Confirm Maintenance Mode dialog box, uncheck Move powered off and suspended virtual machines to other hosts in the cluster and click Yes.
   The host gets ready to go into maintenance mode, which prevents VMs from running on this host. DRS automatically attempts to migrate all the VMs to another host in the cluster.

   > Note: If DRS is not enabled, manually migrate or shut down all the VMs excluding the Controller VM. The VMs that are not migrated automatically even when the DRS is enabled can be because of a configuration option in the VM that is not present on the target host.

6. Log on to the Controller VM with SSH and shut down the Controller VM.

   ```
   nutanix@cvm$ cvm_shutdown -P now
   ```

   > Note: Do not reset or shutdown the Controller VM in any way other than the `cvm_shutdown` command to ensure that the cluster is aware that the Controller VM is unavailable

7. After the Controller VM shuts down, wait for the host to go into maintenance mode.

8. Right-click the host and select **Shut Down**.

   Wait until vCenter Server displays that the host is not responding, which may take several minutes. If you are logged on to the ESXi host rather than to vCenter Server, the vSphere client disconnects when the host shuts down.

Shutting Down a Node in a Cluster (vSphere Command Line)

Before you begin

If DRS is not enabled, manually migrate all the VMs except the Controller VM to another host in the cluster or shut down any VMs other than the Controller VM that you do not want to migrate to another host. If DRS is enabled on the cluster, you can skip this pre-requisite.

About this task

> CAUTION: Verify the data resiliency status of your cluster. If the cluster only has replication factor 2 (RF2), you can only shut down one node for each cluster. If an RF2 cluster would have more than one node shut down, shut down the entire cluster.

You can put the ESXi host into maintenance mode and shut it down from the command line or by using the vSphere web client.

Procedure

1. Log on to the Controller VM with SSH and shut down the Controller VM.

   ```
   nutanix@cvm$ cvm_shutdown -P now
   ```

2. Log on to another Controller VM in the cluster with SSH.

3. Shut down the host.

   ```
   nutanix@cvm$ ~/serviceability/bin/esx-enter-maintenance-mode -s cvm_ip_addr
   ```

   If successful, this command returns no output. If it fails with a message like the following, VMs are probably still running on the host.

   ```
   CRITICAL esx-enter-maintenance-mode:42 Command vim-cmd hostsvc/maintenance_mode_enter failed
    with ret=-1
   ```

   Ensure that all VMs are shut down or moved to another host and try again before proceeding.

   ```
   nutanix@cvm$ ~/serviceability/bin/esx-shutdown -s cvm_ip_addr
   ```

   Replace *cvm_ip_addr* with the IP address of the Controller VM on the ESXi host.

   Alternatively, you can put the ESXi host into maintenance mode and shut it down using the vSphere Web Client.

   If the host shuts down, a message like the following is displayed.

   ```
   INFO esx-shutdown:67 Please verify if ESX was successfully shut down using
    ping hypervisor_ip_addr
   ```

4. Confirm that the ESXi host has shut down.

   ```
   nutanix@cvm$ ping hypervisor_ip_addr
   ```

   Replace *hypervisor_ip_addr* with the IP address of the ESXi host.

   If no ping packets are answered, the ESXi host shuts down.

Shutting Down a Node in a Cluster (AHV)

About this task

> CAUTION: Verify the data resiliency status of your cluster. If the cluster only has replication factor 2 (RF2), you can only shut down one node for each cluster. If an RF2 cluster would have more than one node shut down, shut down the entire cluster.

You must shut down the Controller VM to shut down a node. When you shut down the Controller VM, you must put the node in maintenance mode.

When a host is in maintenance mode, VMs that can be migrated are moved from that host to other hosts in the cluster. After exiting maintenance mode, those VMs are returned to the original host, eliminating the need to manually move them.

If a host is put in maintenance mode, the following VMs are not migrated:

- VMs with GPUs, CPU passthrough, PCI passthrough, and host affinity policies are not migrated to other hosts in the cluster. You can shut down such VMs by setting the **non_migratable_vm_action** parameter to **acpi_shutdown**. If you do not want to shut down these VMs for the duration of maintenance mode, you can set the **non_migratable_vm_action** parameter to **block**, or manually move these VMs to another host in the cluster.
- Agent VMs are always shut down if you put a node in maintenance mode and are powered on again after exiting maintenance mode.

Perform the following procedure to shut down a node.

Procedure

1. If the Controller VM is running, shut down the Controller VM.
   a. Log on to the Controller VM with SSH.
   b. List all the hosts in the cluster.
      ```
      acli host.list
      ```
      Note value of **Hypervisor address** for the node you want to shut down.
   c. Put the node into maintenance mode.
      ```
      nutanix@cvm$ acli host.enter_maintenance_mode Hypervisor address [wait="{ true |
       false }" ] [non_migratable_vm_action="{ acpi_shutdown | block }" ]
      ```
      Replace *`Hypervisor address`* with either the IP address or host name of the AHV host you want to shut down.

      Set **wait=true** to wait for the host evacuation attempt to finish.

      Set **non_migratable_vm_action=acpi_shutdown** if you want to shut down VMs such as VMs with GPUs, CPU passthrough, PCI passthrough, and host affinity policies during the maintenance mode.

      If you do not want to shut down these VMs during maintenance mode, you can set the **non_migratable_vm_action** parameter to **block**, or manually move these VMs to another host in the cluster.

      If you set the **non_migratable_vm_action** parameter to **block** and the operation to put the host into the maintenance mode fails, exit the maintenance mode and then either manually migrate the VMs to another host or shut down the VMs by setting the **non_migratable_vm_action** parameter to **acpi_shutdown**.
   d. Shut down the Controller VM.
      ```
      nutanix@cvm$ cvm_shutdown -P now
      ```
2. Log on to the AHV host with SSH.
3. Shut down the host.
   ```
   root@ahv# shutdown -h now
   ```

Shutting Down a Node in a Cluster (Hyper-V)

Shut down a node in a Hyper-V cluster.

Before you begin

Shut down guest VMs that are running on the node, or move them to other nodes in the cluster.

About this task

> CAUTION: Verify the data resiliency status of your cluster. If the cluster only has replication factor 2 (RF2), you can only shut down one node for each cluster. If an RF2 cluster would have more than one node shut down, shut down the entire cluster.

Perform the following procedure to shut down a node in a Hyper-V cluster.

Procedure

1. Log on to the Hyper-V host with Remote Desktop Connection.
2. Open the Failover Cluster Manager.
3. Click on **Nodes**.
4. Right-click on the node name.
5. Under **Pause** click **Drain Roles**.
   Under Status the node will appear as **Paused**.

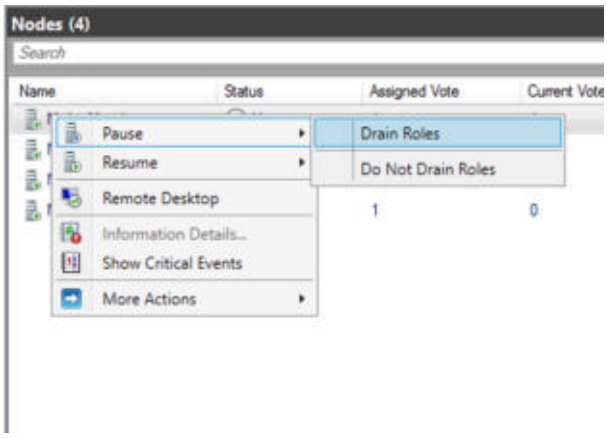


Figure 1: Pause Node Options in Failover Cluster Manager

6. (Optional) At the bottom of the center pane click on the **Roles** tab.
   Once all roles have moved off this node, it is safe to shut down the node.
7. Log on to the Controller VM with SSH and shut down the Controller VM.

```
nutanix@cvm$ cvm_shutdown -P now
```

> Note:
>
> Always use the `cvm_shutdown` command to reset, or shutdown the Controller VM. The `cvm_shutdown` command notifies the cluster that the Controller VM is unavailable.

8. Log on to the Hyper-V host with Remote Desktop Connection and start PowerShell.

9. Do one of the following to shut down the node.

   » `> shutdown /s /t 0`

   » > `Stop-Computer -ComputerName localhost`

   See the Microsoft documentation for up-to-date and additional details about how to shut down a Hyper-V node.

Manually Updating M.2 RAID Hypervisor Boot Drive Firmware

Manually update the firmware on HW RAID M.2 boot drives on -G7 or newer platforms in cases when LCM cannot be used.

About this task

Starting with the -G7 model, Nutanix platforms use a host boot device with a hardware RAID. Whenever possible, you should use LCM to update the firmware on M.2 RAID drives. LCM automates the upgrade. In some cases you might not be able to use LCM. Use this topic for all manual firmware upgrades for RAID M.2 boot drives.

During this procedure, you must restart your node into the Phoenix ISO and fetch a firmware update binary from Nutanix. KB 6937 lists the update binaries KB 6937.

> Note: If you need to update M.2 firmware on -G6 platforms, or any -G6 or newer platforms with a single M.2 drive, refer to the topic: Manually Updating Data Drive Firmware in this Hardware Administration Guide. Do not attempt to follow this procedure on older platforms.

> Note: Always use the latest Phoenix version.

> CAUTION: Only use this process on Nutanix (NX-) platforms.

Procedure

1. Go to the Nutanix portal at https://portal.nutanix.com and select **Downloads** > **Phoenix** to reach the Phoenix landing page. Download the Phoenix ISO to your system.

2. From the CVM prompt, identify an active 10G interface on the host. Make a note of the interface for use in a later step.

   » AHV:

   ```
   nutanix@cvm$ manage_ovs show_interfaces
   ```

   » ESXi:

   ```
   nutanix@cvm$ ssh root@host-IP-address esxcli network nic list
   ```

3. Find an unused IP address that is in the same subnet as the CVM. Make a note of it so you can assign this IP address to the Phoenix ISO in a later step.

4. Check to see if the CVM has a VLAN configured. If it does, make a note of the VLAN ID.

5. Check the state of the cluster.

   a. Make sure that no services are down on the cluster.

   ```
   nutanix@cvm$ cluster status | grep -v UP
   ```

   b. Make sure that all hosts are part of the metadata ring.

   ```
   nutanix@cvm$ nodetool -h 0 ring
   ```

   c. Check cluster data resiliency.

   ```
   nutanix@cvm$ ncli cluster get-domain-fault-tolerance-status type=node
   ```

6. Put the host into maintenance mode and shut down the CVM and the node.

   » If the drive is part of a single-node cluster, follow the procedures that start with Shutting Down a Single-node Cluster on page 13 .

   » If the drive is part of a multinode cluster, follow the procedures that start with Shutting Down a Node in a Multinode Cluster on page 15.

7. From the system where you downloaded the Phoenix ISO, enter the IPMI IP address of the node in a web browser to reach the IPMI web UI for the node.

8. In the IPMI web UI, launch the remote console by selecting **Remote Control** > **Console Redirection**.

Figure 2: Console redirection

9. Click the **Launch Console** button.

10. From the console menu, select **Virtual Media** > **Virtual Storage**.

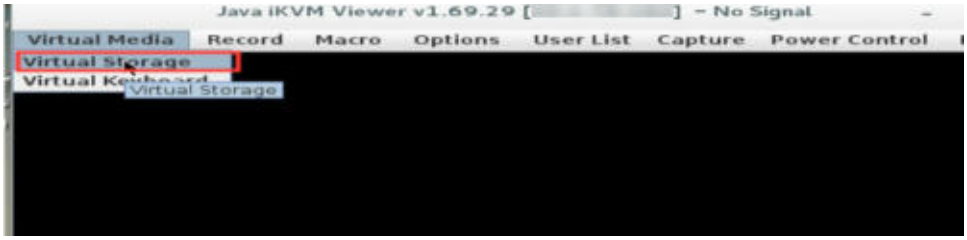


Figure 3: Virtual storage

11. In the **Virtual Storage** dialog box, specify the storage settings.

    a. In the **Logical Drive Type** field, select **ISO File**.

    b. In the **Image File Name and Full Path** field, select **Open Image**, browse to the location where you downloaded the Phoenix ISO file, and click **Open**.

    c. Click **Plug In** to mount the ISO on the node.

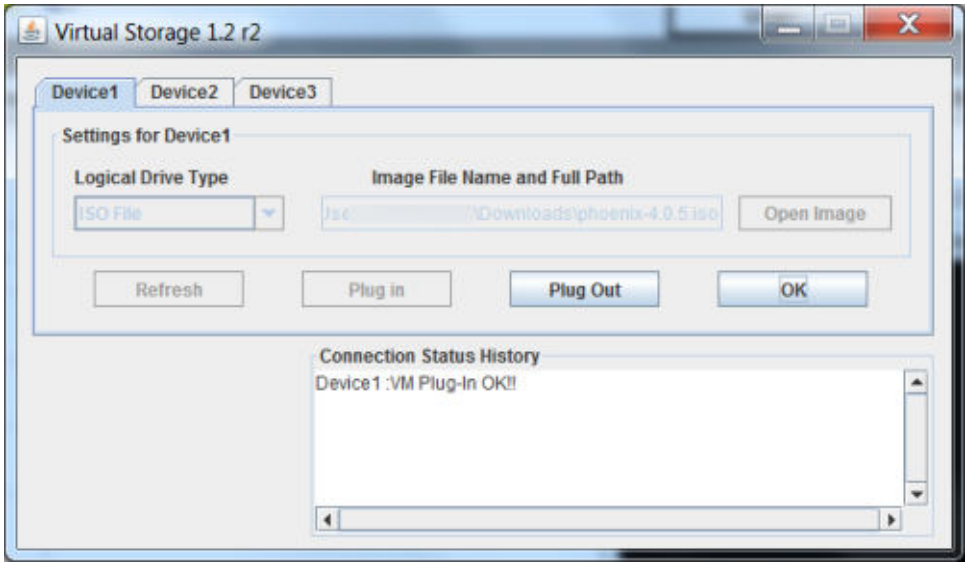


Figure 4: Specifying the image

12. Click **OK**.

13. From the console menu, restart the node by selecting **Power Control** > **Set Power On**.

Figure 5: Restart the node with Set Power On

The node restarts, showing the Phoenix prompt.

```
phoenix ~#
```

14. Set an IP address, netmask, and default gateway.

    » If the CVM has a VLAN configured:

```
phoenix ~# ifconfig interface up
phoenix ~# ip link add link interface name interface.VLAN_ID type vlan id VLAN_ID
phoenix ~# ip addr add PHX_IP_address/24 dev interface.VLAN_ID
phoenix ~# ip link set dev interface.VLAN_ID up
```

```
phoenix ~# ip link show
phoenix ~# ip route add default via CVM_default_gateway dev interface.VLAN_ID
```

- For *interface*, use the active interface you identified in step 2.
- For *PHX_IP_address*, use the IP address that you identified in step 3.

In the following example, the interface is `eth2` and the VLAN ID is `691`.

```
phoenix ~# ifconfig eth2 up
phoenix ~# ip link add link eth2 name eth2.691 type vlan id 691
phoenix ~# ip addr add 198.51.100.10/24 dev eth2.691
phoenix ~# ip link set dev eth2.691 up
phoenix ~# ip link show
phoenix ~# ip route add default via 198.51.100.16 dev eth2.691
```

» If the CVM does not have a VLAN configured:

```
phoenix ~# ip link set dev interface up
phoenix ~# ifconfig interface PHX_IP_address
phoenix ~# ifconfig interface netmask CVM_subnet_mask
phoenix ~# ip route add default via CVM_default_gateway dev interface
```

- For *interface*, use the active interface you identified in step 2.
- For *PHX_IP_address*, use the IP address that you identified in step 3.

In the following example, the interface is `eth2`.

```
phoenix ~# ip link set dev eth2 up
phoenix ~# ifconfig eth2 198.51.100.10
phoenix ~# ifconfig eth2 netmask 255.255.255.0
phoenix ~# ip route add default via 198.51.100.16 dev eth2
```

15. Test connectivity by pinging another CVM or the gateway.
16. Enable the rescue shell.
    a. Change your working directory to the directory that contains the `do_rescue_shell.sh` script.

    ```
    phoenix ~# cd /root
    ```

    b. Set executable permissions on the rescue shell script.

    ```
    phoenix/root~# chmod +x do_rescue_shell.sh
    ```

    c. Run the rescue shell script.

    ```
    phoenix/root~# sh do_rescue_shell.sh
    ```

17. Open a new SSH session to the CVM where the binaries are staged, and log on as root with the password `nutanix/4u`.
18. Run the `lsscsi` command.

```
[root@phoenix ~]# lsscsi
[0:0:0:0]    disk    ATA      SAMSUNG MZ7LH3T8 404Q  /dev/sda
```

```
[0:0:1:0]    disk    ATA      ST4000NM0035-1V4 TN05  /dev/sdb
[0:0:2:0]    disk    ATA      ST4000NM0035-1V4 TN05  /dev/sdc
[1:0:0:0]    cd/dvd  ATEN     Virtual CDROM    YS0J  /dev/sr0
[16:0:0:0]   disk    ATA      SMC VD           00-0  /dev/sdd
[18:0:0:0]   process Marvell  Console          1.01  -
```

The enumeration of the disks might vary. You can see a device labeled "`SMC VD`" (in this case it is `/dev/sdd`). This disk represents the hardware RAID for the boot partition, rather than showing the individual disks in the RAID. This is one difference which prevents a direct update of the firmware. This also prevents you from directly checking the firmware of the disks with `smartctl`.

19. Check the version details for the disks (and HBA) with the`mvcli` utility in phoenix.

    a. Check the version of disk in port 0:

```
[root@phoenix ~]# mvcli info -o pd -i 0
SG driver version 3.5.36.

Physical Disk Information
----------------------------
Adapter:             0
PD ID:               0
Type:                SATA PD
Linked at:           HBA port 0
Size:                234431064 K
Write cache:         not supported
SMART:               supported (on)
NCQ:                 supported (on)
48 bits LBA:         supported
supported speed:     1.5 3 6 Gb/s
Current speed:       6 Gb/s
model:               Micron_5100_MTFDDAV240TCB
Serial:              18381E845493
Firmware version:     D0MU071
Locate LED status:   Not Support
Running OS:          no
SSD Type:            SSD
block ids:           0
associated VDs:      0
PD valid size:       0 K
```

    b. Check the version of disk in port 1:

```
[root@phoenix ~]# mvcli info -o pd -i 1
SG driver version 3.5.36.

Physical Disk Information
----------------------------
Adapter:             0
PD ID:               1
Type:                SATA PD
Linked at:           HBA port 1
Size:                234431064 K
Write cache:         not supported
SMART:               supported (on)
NCQ:                 supported (on)
48 bits LBA:         supported
supported speed:     1.5 3 6 Gb/s
Current speed:       6 Gb/s
model:               Micron_5100_MTFDDAV240TCB
Serial:              18381E8454D3
Firmware version:     D0MU071
```

```
Locate LED status:   Not Support
Running OS:          no
SSD Type:            SSD
block ids:           4
associated VDs:      0
PD valid size:       0 K
```

c. Check the version of the raid controller:

```
[root@phoenix ~]# mvcli info -o hba
SG driver version 3.5.36.

Adapter ID:                          0
Product:                             1b4b-9230
Sub Product:                         15d9-1b22
Chip revision:                       A1
slot number:                         0
Max PCIe speed:                      5Gb/s
Current PCIe speed:                  5Gb/s
Max PCIe link:                       2
Current PCIe link:                   2
BIOS version:                        1.0.0.1031
Firmware version:                    2.3.21.1003
Boot loader version:                 2.1.0.1009
# of ports:                          3
Buzzer:                              Not supported
Supported port type:                 SATA
Supported RAID mode:                 RAID1 JBOD
Maximum disk in one VD:              2
PM:                                  Not supported
Expander:                            Not supported
Maximum supported disk:              2
Maximum supported VD:                1
Max total blocks:                    128
Features:                            rebuild,media patrol
Advanced features:                   event sense code,multi VD,spc 4,image
 health,timer,ata pass through,oem data
Advanced features 2:                 scsi pass through,flash
Max buffer size:                     3
Stripe size supported:               32K 64K
Image health:                        Healthy
Autoload image health:               Healthy
Boot loader image health:            Healthy
Firmware image health:               Healthy
Boot ROM image health:               Healthy
HBA info image health:               Healthy
```

20. Download the firmware update binary for your drive from the list in KB 6937.

21. Copy the firmware update binary to the Phoenix IP address that you assigned earlier: use a secure copy command such as `scp`.

22. Use the following steps to update the firmware.

    a. Copy the firmware to the phoenix host over `ssh` (to the `/root` directory):

    ```
    scp firmware_binary.bin root@phoenix_ip_address:/root/
    ```

    b. Change to the `/root` directory if you are not there already.

    ```
    cd /root
    ```

    c. If the binary was zipped, unzip the firmware that you uploaded in step a to get access to the .bin file.

    ```
    unzip firmware_bundle.zip
    ```

    d. To update the drive in port 0, type:

    ```
    mvcli pdflash -i 0 -f firmware_binary.bin
    ```

    e. To update the drive in port 1, type:

    ```
    mvcli pdflash -i 1 -f firmware_binary.bin
    ```

    f. Verify the new firmware version of each M.2 device:

    ```
    [root@phoenix ~]# mvcli info -o pd -i 1
    ```

    ```
    [root@phoenix ~]# mvcli info -o pd -i 0
    ```

23. From the console menu, select **Virtual Media** > **Virtual Storage** to unmount the ISO.

24. Disconnect from the IPMI web UI and restart the host.

    » If the drive is part of a single-node cluster, follow the procedures in Starting a Single-node Cluster on page 47.

    » If the drive is part of a multinode cluster, follow the procedures in Starting a Node in a Multinode Cluster on page 50.

25. Check the state of the cluster.

    a. Make sure that all hosts are part of the metadata ring.

    ```
    nutanix@cvm$ nodetool -h 0 ring
    ```

    b. Check cluster data resiliency.

    ```
    ncli> cluster get-domain-fault-tolerance-status type=node
    ```

Manually Updating NVMe Drive Firmware

You should update firmware on NVMe drives using LCM, rather than through the CVM.

About this task

To manually update an NVMe drive, you must restart the node and download firmware update binary provided by Nutanix. KB 6937 lists the update binaries KB 6937. Contact Nutanix support if you need further assistance.

Procedure

1. Go to the Nutanix portal at https://portal.nutanix.com and select **Downloads** > **Phoenix** to reach the Phoenix landing page. Download the Phoenix ISO to your system.

2. From the CVM prompt, identify an active 10G interface on the host. Make a note of the interface for use in a later step.

   » AHV:

   ```
   nutanix@cvm$ manage_ovs show_interfaces
   ```

   » ESXi:

   ```
   nutanix@cvm$ ssh root@host-IP-address esxcli network nic list
   ```

3. Find an unused IP address that is in the same subnet as the CVM. Make a note of it so you can assign this IP address to the Phoenix ISO in a later step.

4. Check to see if the CVM has a VLAN configured. If it does, make a note of the VLAN ID.

5. Check the state of the cluster.

   a. Make sure that no services are down on the cluster.

   ```
   nutanix@cvm$ cluster status | grep -v UP
   ```

   b. Make sure that all hosts are part of the metadata ring.

   ```
   nutanix@cvm$ nodetool -h 0 ring
   ```

   c. Check cluster data resiliency.

   ```
   ncli> cluster get-domain-fault-tolerance-status type=node
   ```

6. Put the host into maintenance mode and shut down the CVM and the node.

   » If the drive is part of a single-node cluster, follow the procedures in Shutting Down a Single-node Cluster on page 13.

   » If the drive is part of a multinode cluster, follow the procedures in Shutting Down a Node in a Multinode Cluster on page 15.

7. From the system where you downloaded the Phoenix ISO, enter the IPMI IP address of the node in a web browser to reach the IPMI web UI for the node.

8. In the IPMI web UI, launch the remote console by selecting **Remote Control** > **Console Redirection**.

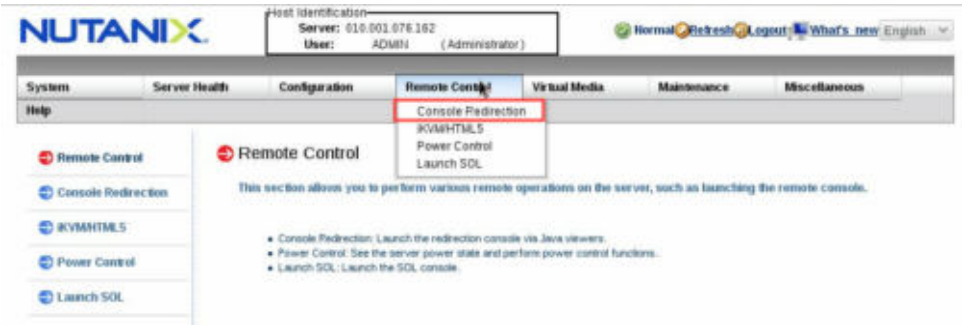


Figure 6: Console redirection

9. Click the **Launch Console** button.

10. From the console menu, select **Virtual Media** > **Virtual Storage**.

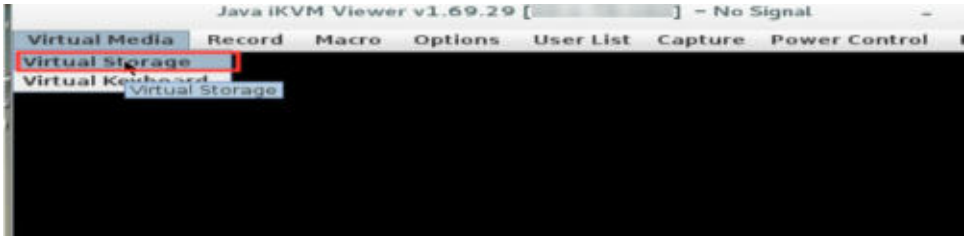


Figure 7: Virtual storage

11. In the **Virtual Storage** dialog box, specify the storage settings.

    a. In the **Logical Drive Type** field, select **ISO File**.

    b. In the **Image File Name and Full Path** field, select **Open Image**, browse to the location where you downloaded the Phoenix ISO file, and click **Open**.

    c. Click **Plug In** to mount the ISO on the node.

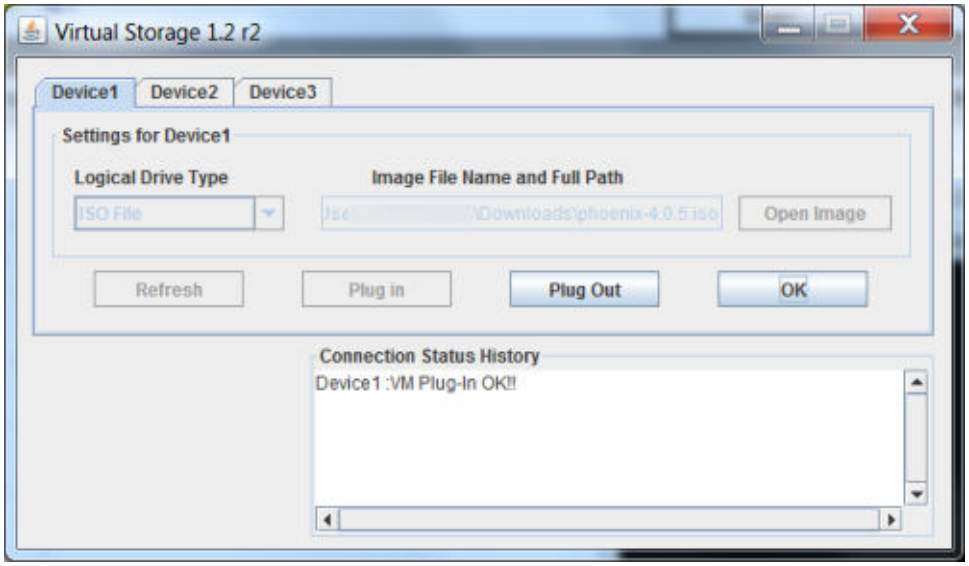


Figure 8: Specifying the image

12. Click **OK**.

13. From the console menu, restart the node by selecting **Power Control** > **Set Power Reset**.

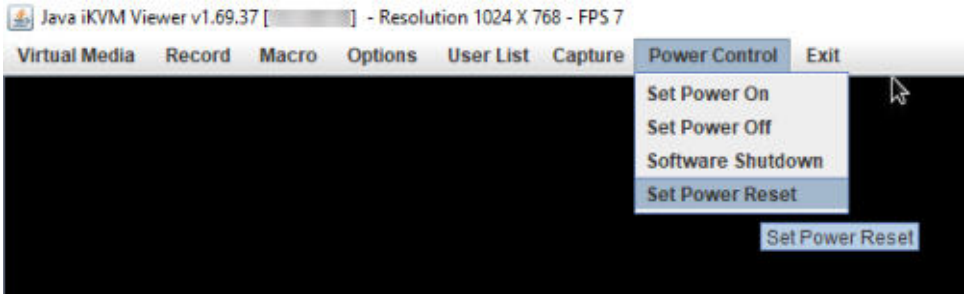


Figure 9: Restart the node

The node restarts, showing the Phoenix prompt.

```
phoenix ~#
```

14. Set an IP address, netmask, and default gateway.

    » If the CVM has a VLAN configured:

```
phoenix ~# ifconfig interface up
phoenix ~# ip link add link interface name interface.VLAN_ID type vlan id VLAN_ID
```

```
phoenix ~# ip addr add PHX_IP_address/24 dev interface.VLAN_ID
phoenix ~# ip link set dev interface.VLAN_ID up
phoenix ~# ip link show
phoenix ~# ip route add default via CVM_default_gateway dev interface.VLAN_ID
```

- For *interface*, use the active interface that you identified in step 2.
- For *PHX_IP_address*, use the IP address that you identified in step 3.

In the following example, the interface is `eth2` and the VLAN ID is `691`.

```
phoenix ~# ifconfig eth2 up
phoenix ~# ip link add link eth2 name eth2.691 type vlan id 691
phoenix ~# ip addr add 198.51.100.10/24 dev eth2.691
phoenix ~# ip link set dev eth2.691 up
phoenix ~# ip link show
phoenix ~# ip route add default via 198.51.100.16 dev eth2.691
```

» If the CVM does not have a VLAN configured:

```
phoenix ~# ip link set dev interface up
phoenix ~# ifconfig interface PHX_IP_address
phoenix ~# ifconfig interface netmask CVM_subnet_mask
phoenix ~# ip route add default via CVM_default_gateway dev interface
```

- For *interface*, use the active interface you identified in step 2.
- For *PHX_IP_address*, use the IP address that you identified in step 3.

In the following example, the interface is `eth2`.

```
phoenix ~# ip link set dev eth2 up
phoenix ~# ifconfig eth2 198.51.100.10
phoenix ~# ifconfig eth2 netmask 255.255.255.0
phoenix ~# ip route add default via 198.51.100.16 dev eth2
```

15. Test connectivity by pinging another CVM or the gateway.
16. Enable the rescue shell.
    a. Change your working directory to the directory that contains the `do_rescue_shell.sh` script.

    ```
    phoenix ~# cd /root
    ```

    b. Set executable permissions on the rescue shell script.

    ```
    phoenix/root~# chmod +x do_rescue_shell.sh
    ```

    c. Run the rescue shell script.

    ```
    phoenix/root~# sh do_rescue_shell.sh
    ```

17. Return to the Phoenix prompt.
18. List the NVMe drives installed on the system.

    ```
    [root@phoenix ~]# nvme list
    ```

```
Node           SN  Model        Namespace Usage    Format    FW Rev
---------------------------------------------------------------------------
/dev/nvme0n1 S3HDNX0KC00402 SAMSUNG MZWLL1T6HEHP-00003 1 1.60 TB / 1.60  TB 512 B + 0 B
 GPNA9B3Q
/dev/nvme1n1 S3HDNX0KC00400 SAMSUNG MZWLL1T6HEHP-00003 1 1.60 TB / 1.60  TB 512 B + 0 B
 GPNA9B3Q
```

19. Check the NVMe firmware version and firmware slots:

```
[root@phoenix ~]# nvme id-ctrl /dev/nvme0 |egrep 'mn |fr'
mn        : SAMSUNG MZWLL1T6HEHP-00003
fr        : GPNA9B3Q
frmw      : 0x17
```

20. Convert the hex value reported for frmw to determine the number of firmware "slots" on the drive according the following mapping.

    Bit 0 if set to '1' indicates that the first firmware slot (slot 1) is read only. If cleared to '0' then the first firmware slot (slot 1) is read/write. Implementations may choose to have a baseline read only firmware image.

    Bits 3:1 indicates the number of firmware slots that the controller supports. This field shall specify a value between one and seven, indicating that at least one firmware slot is supported and up to seven maximum. This corresponds to firmware slots 1 through 7.

    In this example, we see that the 0x17 value converts to a binary 00010111

    Bit [3:1] = "011": There are 3 firmware slots.

    Bit [0:0] = "1": Firmware Slot 1 is read only.

    > Note: The above output will vary depending on the firmware version of the drive. If slot 1 is read-only and the drive supports three firmware slots, you need to activate the firmware image from slot 2 which is read/write. If slot 1 is read/write, you should activate slot 1.

    This mapping information is based on the NVMe specifications at the NVME express website: http://www.nvmexpress.org/

21. Download the firmware update binary for your drive from the list in KB 6937 KB 6937.

22. Copy the firmware update binary to the Phoenix IP address you assigned in step 14, using a secure copy command such as `scp`.

23. Download the firmware.

```
[root@phoenix tmp]# nvme fw-download /dev/nvme0 --fw=/root/tmp/GPNABB3Q.bin
Firmware download success
```

24. Activate the firmware in the appropriate slot.

```
[root@phoenix tmp]# nvme fw-activate /dev/nvme0 --slot=2 --action=1
Success committing firmware action:1 slot:2
```

25. Use the `id-ctrl` command to verify that the new firmware is active.

```
[root@phoenix tmp]# nvme id-ctrl /dev/nvme0 |egrep 'mn |fr'
mn        : SAMSUNG MZWLL1T6HEHP-00003
fr        : GPNABB3Q
```

```
frmw      : 0x17
```

If you cannot see the new firmware version after running the **id-ctrl** command, the firmware version will appear later when you reset the host. To see the firmware version, you can also issue the following command:

```
[root@phoenix tmp]# nvme reset /dev/nvme0
```

26. Repeat steps 18 through step 25 for all the NVMe drives.
27. From the console menu, select Virtual Media > Virtual Storage to unmount the ISO.
28. Disconnect from the IPMI web UI and restart the host.
    - If the drive is part of a single-node cluster, follow the procedures in Starting a Single-node Cluster on page 47.
    - If the drive is part of a multinode cluster, follow the procedures in Starting a Node in a Multinode Cluster on page 50.
29. Check the state of the cluster.
30. Ensure that all hosts are part of the metadata ring.

    ```
    nutanix@cvm$ nodetool -h 0 ring
    ```

31. Check the cluster data resiliency.

    ```
    ncli> cluster get-domain-fault-tolerance-status type=node
    ```

Manually Updating Data Drive Firmware

Manually update the firmware for a SAS or SATA data drive, or update the hypervisor boot drive firmware on a non-RAID M.2 drive.

About this task

All G6 platforms have non-RAID M.2 drives. The same procedure applies to G7 platforms with only one M.2 device, such as the NX-1175S-G7, the NX-1120S-G7 or newer platforms with a single M.2 drive. To update RAID M.2 devices, see Manually Updating M.2 RAID Hypervisor Boot Drive Firmware on page 21.

To update a data drive manually, you must restart your node into the Phoenix ISO and fetch a firmware update binary from Nutanix. KB 6937 lists the update binaries.

Procedure

1. Go to the Nutanix portal at https://portal.nutanix.com and select **Downloads** > **Phoenix** to reach the Phoenix landing page. Download the Phoenix ISO to your system.
2. From the CVM prompt, identify an active 10G interface on the host. Make a note of the interface for use in a later step.
    - AHV:

      ```
      nutanix@cvm$ manage_ovs show_interfaces
      ```

    - ESXi:

      ```
      nutanix@cvm$ ssh root@host-IP-address esxcli network nic list
      ```

3. Find an unused IP address that is in the same subnet as the CVM. Make a note of it so you can assign this IP address to the Phoenix ISO in a later step.

4. Check to see if the CVM has a VLAN configured. If it does, make a note of the VLAN ID.

5. Check the state of the cluster.

   a. Make sure that no services are down on the cluster.

   ```
   nutanix@cvm$ cluster status | grep -v UP
   ```

   b. Make sure that all hosts are part of the metadata ring.

   ```
   nutanix@cvm$ nodetool -h 0 ring
   ```

   c. Check cluster data resiliency.

   ```
   ncli> cluster get-domain-fault-tolerance-status type=node
   ```

6. Put the host into maintenance mode and shut down the CVM and the node.

   » If the drive is part of a single-node cluster, follow the procedures in Shutting Down a Single-node Cluster on page 13.

   » If the drive is part of a multinode cluster, follow the procedures in Shutting Down a Node in a Multinode Cluster on page 15.

7. From the system where you downloaded the Phoenix ISO, enter the IPMI IP address of the node in a web browser to reach the IPMI WebUI for the node.

8. In the IPMI WebUI, launch the remote console by selecting **Remote Control** > **Console Redirection**.

Figure 10: Console redirection

9. Click the **Launch Console** button.

10. From the console menu, select **Virtual Media** > **Virtual Storage**.

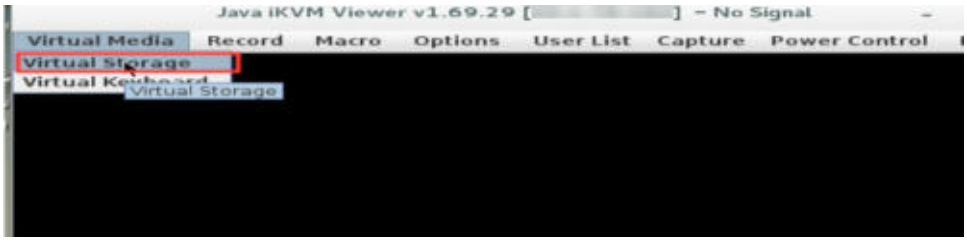


Figure 11: Virtual storage

11. In the **Virtual Storage** dialog box, specify the storage settings.

    a. In the **Logical Drive Type** field, select **ISO File**.

    b. In the **Image File Name and Full Path** field, select **Open Image**, browse to the location where you downloaded the Phoenix ISO file, and click **Open**.

    c. Click **Plug In** to mount the ISO on the node.

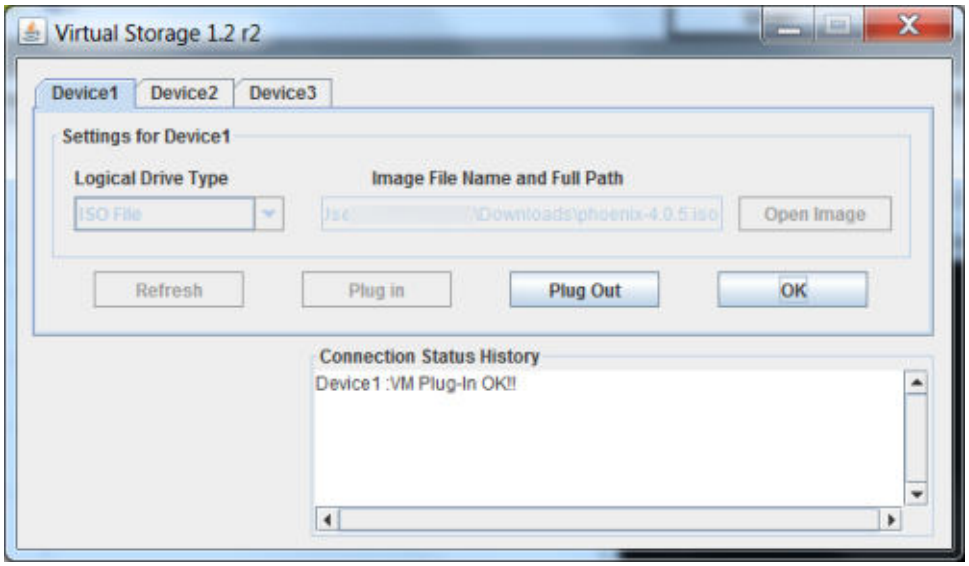


Figure 12: Specifying the image

12. Click **OK**.

13. From the console menu, restart the node by selecting **Power Control** > **Set Power Reset**.

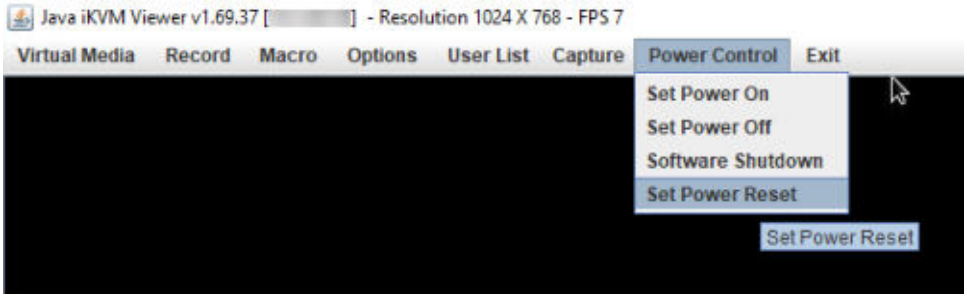


Figure 13: Restart the node

The node restarts, showing the Phoenix prompt.

```
phoenix ~#
```

14. Set an IP address, netmask, and default gateway.

» If the CVM has a VLAN configured:

```
phoenix ~# ifconfig interface up
phoenix ~# ip link add link interface name interface.VLAN_ID type vlan id VLAN_ID
phoenix ~# ip addr add PHX_IP_address/24 dev interface.VLAN_ID
phoenix ~# ip link set dev interface.VLAN_ID up
phoenix ~# ip link show
phoenix ~# ip route add default via CVM_default_gateway dev interface.VLAN_ID
```

- For *`interface`*, use the active interface you identified in step 2.
- For *`PHX_IP_address`*, use the IP address that you identified in step 3.

In the following example, the interface is `eth2` and the VLAN ID is `691`.

```
phoenix ~# ifconfig eth2 up
phoenix ~# ip link add link eth2 name eth2.691 type vlan id 691
phoenix ~# ip addr add 198.51.100.10/24 dev eth2.691
phoenix ~# ip link set dev eth2.691 up
phoenix ~# ip link show
phoenix ~# ip route add default via 198.51.100.16 dev eth2.691
```

» If the CVM does not have a VLAN configured:

```
phoenix ~# ip link set dev interface up
phoenix ~# ifconfig interface PHX_IP_address
phoenix ~# ifconfig interface netmask CVM_subnet_mask
phoenix ~# ip route add default via CVM_default_gateway dev interface
```

- For *`interface`*, use the active interface you identified in step 2.
- For *`PHX_IP_address`*, use the IP address that you identified in step 3.

In the following example, the interface is `eth2`.

```
phoenix ~# ip link set dev eth2 up
```

```
phoenix ~# ifconfig eth2 198.51.100.10
phoenix ~# ifconfig eth2 netmask 255.255.255.0
phoenix ~# ip route add default via 198.51.100.16 dev eth2
```

15. Test connectivity by pinging another CVM or the gateway.
16. Enable the rescue shell.
    a. Change your working directory to the directory that contains the `do_rescue_shell.sh` script.

    ```
    phoenix ~# cd /root
    ```

    b. Set executable permissions on the rescue shell script.

    ```
    phoenix/root~# chmod +x do_rescue_shell.sh
    ```

    c. Run the rescue shell script.

    ```
    phoenix/root~# sh do_rescue_shell.sh
    ```

17. Return to the Phoenix prompt.
18. Enter the `lsscsi` command to find the device name of the drive you want to update.

```
phoenix ~# lsscsi
[0:0:0:0]    disk    ATA    SAMSUNG MZ7KM1T9 104Q    /dev/sda
[0:0:1:0]    disk    ATA    SAMSUNG MZ7KM1T9 104Q    /dev/sdb
[0:0:2:0]    disk    ATA    ST6000NM0115-1YZ SN04    /dev/sdc
[0:0:3:0]    disk    ATA    ST6000NM0115-1YZ SN04    /dev/sdd
[1:0:0:0]    cd/dvd  ATEN   Virtual CDROM    YS0J    /dev/sr0
[10:0:0:0]   disk    ATA    SATADOM-SL 3IE3  301N    /dev/sde
```

19. Run the `smartctl` command (where `/dev/sdx` is the device name of the drive you want to update.) Look in the **Information** section of the output to find out whether your drive uses a SAS or SATA interface. At the same time, make a note of the current firmware version (so you can verify the version after the update.)

```
phoenix ~# smartctl -i /dev/sdx

smartctl 6.4 2015-06-04 r4109 [x86_64-linux-4.10.1] (local build)
Copyright (C) 2002-15, Bruce Allen, Christian Franke, www.smartmontools.org

=== START OF INFORMATION SECTION ===
Device Model:     SAMSUNG MZ7KM1T9HMJP-00005
Serial Number:    S3F6NY0J400685
LU WWN Device Id: 5 002538 c000dae18
Firmware Version: GXM5104Q
User Capacity:    1,920,383,410,176 bytes [1.92 TB]
Sector Size:      512 bytes logical/physical
Rotation Rate:    Solid State Device
Form Factor:      2.5 inches
Device is:        Not in smartctl database [for details use: -P showall]
ATA Version is:   ACS-2, ATA8-ACS T13/1699-D revision 4c
SATA Version is:  SATA 3.1, 6.0 Gb/s (current: 6.0 Gb/s)
Local Time is:    Tue Jul  3 00:36:40 2018 UTC
SMART support is: Available - device has SMART capability.
```

```
SMART support is: Enabled
```

20. Download the firmware update binary for your drive from the list in KB 6937.

21. Change to the `/root` directory if you are not there already.

```
cd /root
```

22. Copy the firmware update binary to the Phoenix IP address you assigned in step 16, using a secure copy command such as `scp`.

    For example, to copy the firmware to the `/root` directory, use the following command.

```
scp firmware_binary root@phoenix_ip_address:/root/
```

23. Update the firmware.

    » For a SATA drive or a -G6 platform M.2 drive: At the Phoenix prompt, enter the following command, where *`binary-file-name`* is the firmware update binary and `/dev/sdx` is the device name of the drive:

```
phoenix ~# hdparm --fwdownload binary-file-name --yes-i-know-what-i-am-doing --please-destroy-my-drive /dev/sdx
```

> Note: The command does not affect any data on your drive. It only updates the firmware. The `--please-destroy-my-drive` flag does not actually destroy the drive. It is safe to proceed with the firmware update.

    » For a SAS drive: At the Phoenix prompt, enter the following command, where *`binary-file-name`* is the firmware update binary and `/dev/sdx` is the device name of the drive:

```
phoenix ~# sg_write_buffer -v -I binary-file-name -m 7 -S 0 -b 0x10000 /dev/sdx
```

24. Run the `smartctl` command (where `/dev/sdx` is the device name of the updated drive) and look in the **Information** section of the output to verify the firmware update.

```
phoenix ~# smartctl -i /dev/sdx

smartctl 6.4 2015-06-04 r4109 [x86_64-linux-4.10.1] (local build)
Copyright (C) 2002-15, Bruce Allen, Christian Franke, www.smartmontools.org

=== START OF INFORMATION SECTION ===
Device Model:     SAMSUNG MZ7KM1T9HMJP-00005
Serial Number:    S3F6NY0HB00176
LU WWN Device Id: 5 002538 c000af0b6
Firmware Version: GXM5304Q
User Capacity:    1,920,383,410,176 bytes [1.92 TB]
Sector Size:      512 bytes logical/physical
Rotation Rate:    Solid State Device
Form Factor:      2.5 inches
Device is:        Not in smartctl database [for details use: -P showall]
ATA Version is:   ACS-2, ATA8-ACS T13/1699-D revision 4c
SATA Version is:  SATA 3.1, 6.0 Gb/s (current: 6.0 Gb/s)
Local Time is:    Tue Jul  3 01:40:43 2018 UTC
SMART support is: Available - device has SMART capability.
SMART support is: Enabled
```

25. From the console menu, select **Virtual Media** > **Virtual Storage** to unmount the ISO.

26. Disconnect from the IPMI WebUI and restart the host.

    » If the drive is part of a single-node cluster, follow the procedures in Starting a Single-node Cluster on page 47.

    » If the drive is part of a multinode cluster, follow the procedures in Starting a Node in a Multinode Cluster on page 50.

27. Check the state of the cluster.

    a. Make sure that all hosts are part of the metadata ring.

    ```
    nutanix@cvm$ nodetool -h 0 ring
    ```

    b. Check cluster data resiliency.

    ```
    ncli> cluster get-domain-fault-tolerance-status type=node
    ```

Manually Updating SATA DOM Firmware

Update firmware for a SATA DOM.

About this task

To update SATA DOM firmware, you need an ISO provided by Nutanix. Download the firmware update binary for your drive from the list in KB 6937. Contact Nutanix support if you need further assistance.

Procedure

1. Download the SATA DOM firmware update ISO.

2. Shut down the CVM and the node.

    » If the SATA DOM is part of a single-node cluster, follow the procedures in Shutting Down a Single-node Cluster on page 13.

    » If the SATA DOM is part of a multinode cluster, follow the procedures in Shutting Down a Node in a Multinode Cluster on page 15.

3. From the system where you downloaded the firmware ISO, enter the IPMI IP address of the node in a web browser to reach the IPMI web UI for the node.

4. In the IPMI web UI, launch the remote console by selecting **Remote Control** > **Console Redirection**.

Figure 14: Console redirection

5. Click the **Launch Console** button.
6. From the console menu, select **Virtual Media** > **Virtual Storage**.

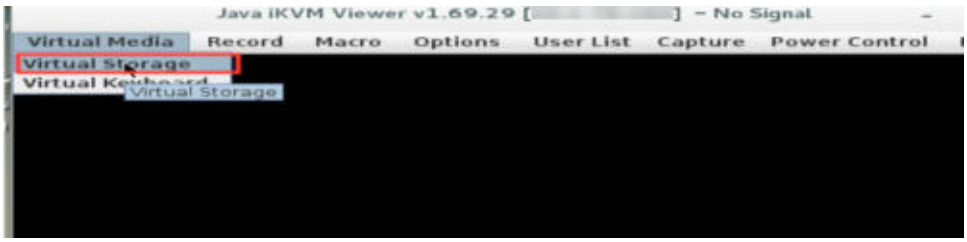


Figure 15: Virtual storage

7. In the **Virtual Storage** dialog box, specify the storage settings.
   a. In the **Logical Drive Type** field, select **ISO File**.
   b. In the **Image File Name and Full Path** field, select **Open Image**, browse to the location where you downloaded the ISO file, and click **Open**.
   c. Select **Plug In** to mount the ISO on the node.

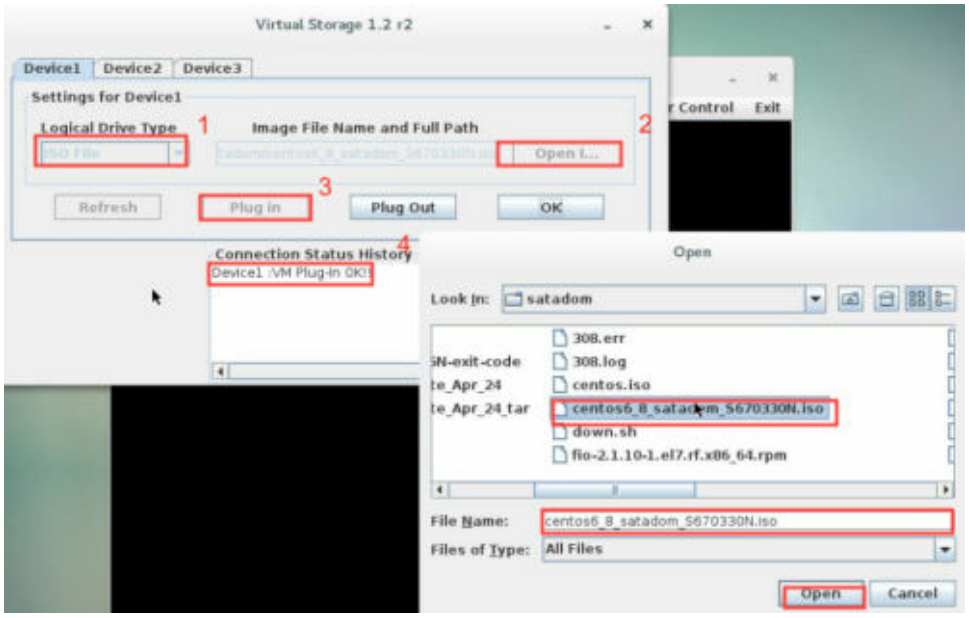


Figure 16: Specifying the image

8. Click **OK**.

9. Turn on the node in order to start from the ISO.

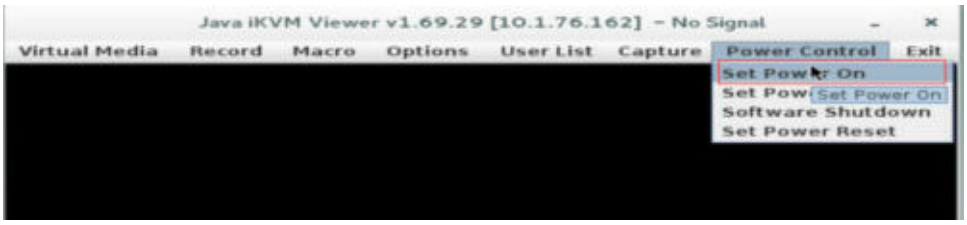


Figure 17: Power on the node

10. After the node starts up from the mounted ISO, log on as root (you do not need a password.)

11. Use the **lsscsi** command to verify that the SATA DOM is visible and present.

```
[root@centos6_8_satadom_S670330N ~]# lsscsi
```

The command returns output similar to the following:

```
[3:0:0:0]    disk    ATA    SATADOM-SL 3IE3    301N    /dev/sda
[10:0:0:0]   cd/dvd  ATEN   Virtual CDROM      YS0J    /dev/sr0
[11:0:0:0]   disk    ATA    SAMSUNG MZ7KM480   0N6Q    /dev/sdb
[11:0:1:0]   disk    ATA    ST2000NX0253       SN02    /dev/sdc
[11:0:2:0]   disk    ATA    ST2000NX0253       SN02    /dev/sdd
[root@centos6_8_satadom_S670330N ~]#
```

> Note: In this example, the SATA DOM-SL 3IE3 is visible, and the device name is **/dev/sda**.

12. Untar the file.

```
# cd /usr/local
# tar xvf filename.tar
# cd firmware-update-directory
```

For example: **tar** *Nutanix-S670330N_firmware_update_Apr_24* and **cd** *Nutanix-S670330N_firmware_update_Apr_24*

13. Change to the *firmware-image-file* directory created by the untarring command and confirm which **/dev/***xxx* is the SATA DOM device.

```
# lsscsi | grep SATADOM #
```

> Note: In the next step, run the command from the directory where you untarred the file.

14. Apply the firmware. Replace **/dev/***xxx* with the name of the SATA DOM device:

```
# ./mp_64 -d /dev/xxx -c 1 -u -k -r -v 0
```

> Note: The character **1** is the numeral one, and the final character is a zero.

> Note: The -r parameter restarts the SATA DOM device.

The command typically takes about a minute to complete. Output resembles the following:

```
**************************************************
* Innodisk MPTool V2.6.4              2018/04/20 *
**************************************************
```

```
Model Name : SATADOM-SL 3IE3 V2
Serial Num : BCA11602030230365
FW Version : S560301N

1. TH58TFT0DDLBA8H

Flash: TH58TFT0DDLBA8H

Write code ....... Pass!!
Reboot ........... ata4: exception Emask 0x50 SAct 0x0 SErr 0x40d0802 action 0xe frozen
ata4: irq_stat 0x00400040, connection status changed
ata4: SError: { RecovComm HostInt PHYRdyChg CommWake 10B8B DevExch }
ata4: hard resetting link
ata4: SATA link up 6.0 Gbps (SStatus 133 SControl 300)
ata4.00: configured for UDMA/133
ata4: EH complete
Pass!!
Upgrade Pass!!
#
```

> Note: The `-r` parameter causes kernel output messages that report **ata*x*** activity. The output indicates that the device has restarted successfully.

15. After you see a successful upgrade, turn off the node by selecting **Power Control** > **Set Power Off**; then turn the node on again by selecting **Power Control** > **Set Power On**.
16. Start again from the ISO (it should still be plugged in to the IPMI virtual media.)
17. Log on as root (you do not need a password.) Confirm that the firmware version is the same as shown in the name of the ISO file you downloaded from Nutanix in step 1, where `/dev/`*`xxx`* is the SATA DOM device:

```
# smartctl -a /dev/xxx | grep -i "Firmware Version"
```

18. Check the current wear level of the media, where `/dev`*`xxx`* is the SATA DOM device:

```
# smartctl -a /dev/xxx | grep -i “Media_Wearout_Indicator”
```

```
233 Media Wearout Indicator 0x0000    100    000    000    Old_age    Offline    -
100
~]#
```

> Note: Make a note of the wear level, and review it with Nutanix Support in order to determine whether the SATA DOM media needs replacing.

19. From the console menu, select **Virtual Media** > **Virtual Storage** to unmount the ISO.
20. Disconnect from the IPMI web UI and restart the host.

    » If the drive is part of a single-node cluster, follow the procedures in Starting a Single-node Cluster on page 47.

    » If the drive is part of a multinode cluster, follow the procedures in Starting a Node in a Multinode Cluster on page 50.

21. Check the state of the cluster.

    a. Make sure that all hosts are part of the metadata ring.

    ```
    nutanix@cvm$ nodetool -h 0 ring
    ```

    b. Check cluster data resiliency.

    ```
    ncli> cluster get-domain-fault-tolerance-status type=node
    ```

Manually Updating the BMC and BIOS

Manually update BMC and BIOS versions.

Before you begin

Before updating, follow the node preparation and shutdown procedures described above.

About this task

> Note: If you are upgrading both the BMC and the BIOS, upgrade the BMC first.

Procedure

1. Update the BMC by following the procedures described in the Nutanix BMC Manual Upgrade Guide.
2. Update the BIOS by following the procedures described in the Nutanix BIOS Manual Upgrade Guide.

Manually Updating HBA Controller Firmware

Update the firmware for an HBA card.

About this task

To update HBA firmware, you need an ISO provided by Nutanix that includes the binary and the `sas3flash` utility. Download the firmware update ISO for your HBA from the list in KB 6937. Contact Nutanix support if you need further assistance.

Procedure

1. Download the HBA controller firmware update ISO.
2. Shut down the CVM and the node.
   - » If the HBA card is in a single-node cluster, follow the procedures in Shutting Down a Single-node Cluster on page 13.
   - » If the HBA card is in a multinode cluster, follow the procedures in Shutting Down a Node in a Multinode Cluster on page 15.
3. From the system where you downloaded the update ISO, enter the IPMI IP address of the node in a web browser to reach the IPMI web UI for the node.

4. In the IPMI web UI, launch the remote console by selecting **Remote Control** > **Console Redirection**.

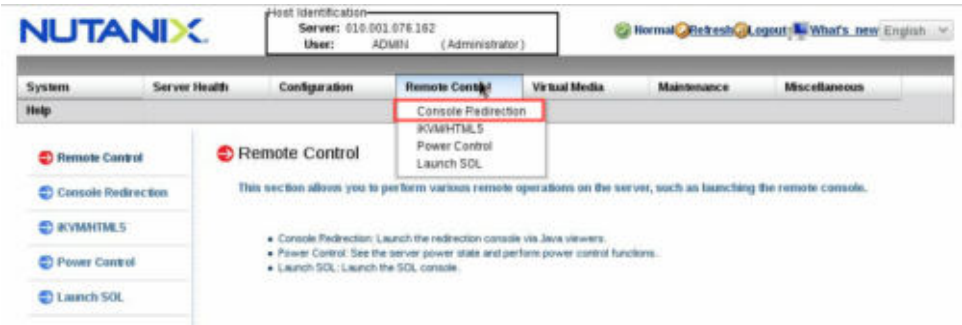


Figure 18: Console redirection

5. Click the **Launch Console** button.

6. From the console menu, select **Virtual Media** > **Virtual Storage**.

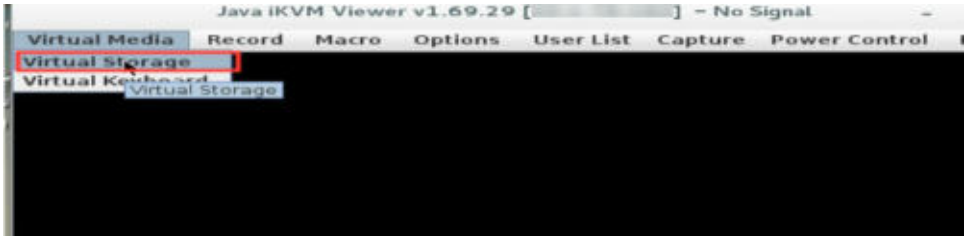


Figure 19: Virtual storage

7. In the **Virtual Storage** dialog box, specify the storage settings.

   a. In the **Logical Drive Type** field, select **ISO File**.

   b. In the **Image File Name and Full Path** field, select **Open Image**, browse to the location where you downloaded the ISO file, and click **Open**.

   c. Select **Plug In** to mount the ISO on the node.

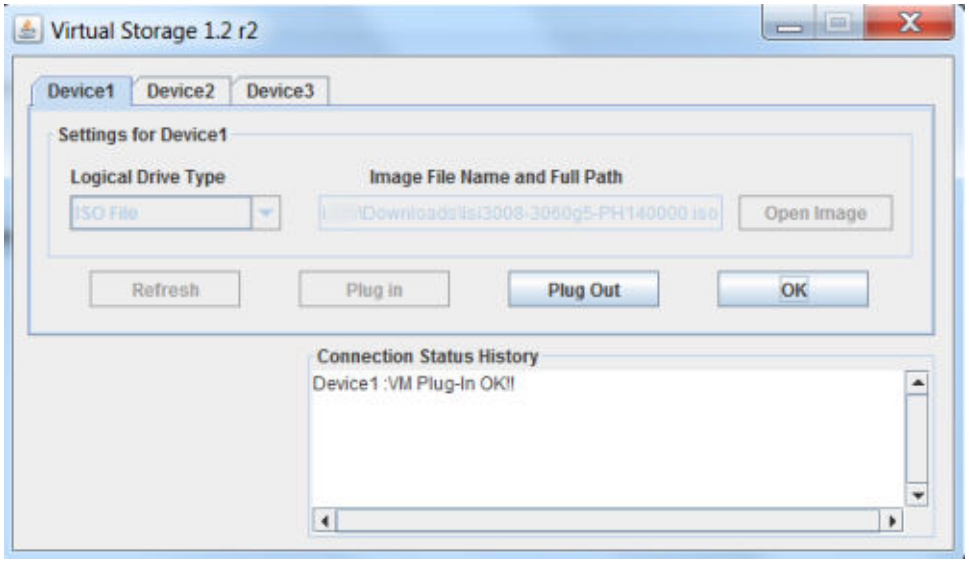


Figure 20: Specifying the image

8. Click **OK**.

9. Turn on the node in order to restart from the ISO.

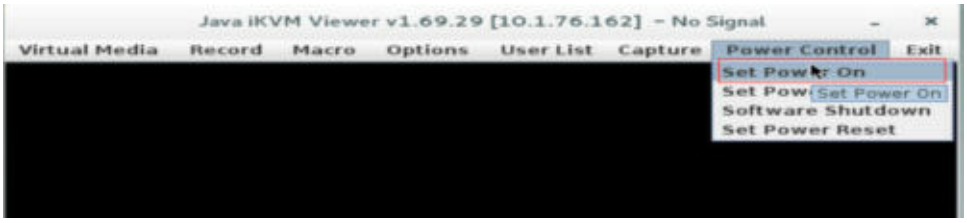


Figure 21: Power on the node

When the host restarts into the ISO, the `sas3flash` utility performs the update automatically.

```
Avago Technologies SAS3 Flash Utility
Version 09.00.00.00 (2015.02.03)
Copyright 2008-2015 Avago Technologies. All rights reserved.

Adapter Selected is a Avago SAS: SAS3008(C0)

Num   Ctlr            FW Ver        NVDATA        x86-BIOS         PCI Addr
----------------------------------------------------------------------------
0  SAS3008(C0)  14.00.00.00    0e.00.30.28    08.31.03.00     00:00:05:00

Finished Processing Commands Successfully.
```

```
Exiting SAS3Flash.
Reached End
>_
```

10. In the **Virtual Storage** dialog box, click **Plug Out** to unmount the ISO.
11. Disconnect from the IPMI web UI and restart the host.
    - » If the HBA card is in a single-node cluster, follow the procedures in Starting a Single-node Cluster on page 47.
    - » If the HBA card is in a multinode cluster, follow the procedures in Starting a Node in a Multinode Cluster on page 50.
12. Check the state of the cluster.
    a. Make sure that all hosts are part of the metadata ring.

    ```
    nutanix@cvm$ nodetool -h 0 ring
    ```

    b. Check cluster data resiliency.

    ```
    ncli> cluster get-domain-fault-tolerance-status type=node
    ```

13. From the CVM, verify the firmware version.

```
nutanix@cvm$ sudo /usr/local/nutanix/bootstrap/lib/lsi-sas/sas3flash -list

Avago Technologies SAS3 Flash Utility
Version 09.00.00.00 (2015.02.03)
Copyright 2008-2015 Avago Technologies. All rights reserved.

Adapter Selected is a Avago SAS: SAS3008(C0)

Controller Number              : 0
Controller                     : SAS3008(C0)
PCI Address                    : 00:00:05:00
SAS Address                    : 5003048-0-1977-6201
NVDATA Version (Default)       : 0e.00.30.28
NVDATA Version (Persistent)    : 0e.00.30.28
Firmware Product ID            : 0x2221 (IT)
Firmware Version               : 14.00.00.00
NVDATA Vendor                  : LSI
NVDATA Product ID              : LSI3008-IT
BIOS Version                   : 08.31.03.00
UEFI BSD Version               : 12.00.00.00
FCODE Version                  : N/A
Board Name                     : LSI3008-IT
Board Assembly                 : N/A
Board Tracer Number            : N/A

Finished Processing Commands Successfully.
Exiting SAS3Flash.
```

Manually Updating NIC Firmware

Manual update for network cards.

LCM does support some NIC firmware updates. Ensure that your platform has the latest version of LCM. Use LCM to verify if NIC firmware updates are available. Use LCM to update the NIC. If the NIC update is not listed in LCM and you need to update the firmware or drivers, contact Nutanix support.

Starting a Single-node Cluster

After updating firmware in a single-node cluster, restart the node using the method for your hypervisor.

Starting a Single Node (vSphere Client)

About this task

Procedure

1. Turn on the node with one of the following methods:
   - » Press the power button.
   - » Use the IPMI Web Interface.

   If the node is already powered on proceed to the next step.
2. Log on to the vSphere Client.
3. Right-click the ESXi host and select **Exit Maintenance Mode**.
4. Right-click the Controller VM and select **Power** > **Power on**.
   Wait approximately 5 minutes for all services to start on the Controller VM.
5. Right-click the ESXi host in the vSphere client and select **Rescan for Datastores**. Confirm that all Nutanix datastores are available.
6. Verify that all services are up on all Controller VMs.

   ```
   nutanix@cvm$ cluster status
   ```

   If the cluster is running properly, output similar to the following is displayed for each node in the cluster:

   ```
           CVM: 10.1.64.60 Up
                                   Zeus   UP       [5362, 5391, 5392, 10848, 10977, 10992]
                              Scavenger   UP       [6174, 6215, 6216, 6217]
                          SSLTerminator   UP       [7705, 7742, 7743, 7744]
                        SecureFileSync   UP       [7710, 7761, 7762, 7763]
                                 Medusa   UP       [8029, 8073, 8074, 8176, 8221]
                   DynamicRingChanger   UP       [8324, 8366, 8367, 8426]
                                 Pithos   UP       [8328, 8399, 8400, 8418]
                                   Hera   UP       [8347, 8408, 8409, 8410]
                               Stargate   UP       [8742, 8771, 8772, 9037, 9045]
                             InsightsDB   UP       [8774, 8805, 8806, 8939]
                  InsightsDataTransfer   UP       [8785, 8840, 8841, 8886, 8888, 8889, 8890]
                                  Ergon   UP       [8814, 8862, 8863, 8864]
                                Cerebro   UP       [8850, 8914, 8915, 9288]
                                Chronos   UP       [8870, 8975, 8976, 9031]
                                Curator   UP       [8885, 8931, 8932, 9243]
                                  Prism   UP       [3545, 3572, 3573, 3627, 4004, 4076]
                                    CIM   UP       [8990, 9042, 9043, 9084]
                           AlertManager   UP       [9017, 9081, 9082, 9324]
   ```

```
                    Arithmos   UP       [9055, 9217, 9218, 9353]
                     Catalog   UP       [9110, 9178, 9179, 9180]
                   Acropolis   UP       [9201, 9321, 9322, 9323]
                       Atlas   UP       [9221, 9316, 9317, 9318]
                       Uhura   UP       [9390, 9447, 9448, 9449]
                        Snmp   UP       [9418, 9513, 9514, 9516]
            SysStatCollector   UP       [9451, 9510, 9511, 9518]
                      Tunnel   UP       [9480, 9543, 9544]
               ClusterHealth   UP       [9521, 9619, 9620, 9947, 9976, 9977, 10301]
                       Janus   UP       [9532, 9624, 9625]
           NutanixGuestTools   UP       [9572, 9650, 9651, 9674]
                  MinervaCVM   UP       [10174, 10200, 10201, 10202, 10371]
               ClusterConfig   UP       [10205, 10233, 10234, 10236]
                 APLOSEngine   UP       [10231, 10261, 10262, 10263]
                       APLOS   UP       [10343, 10368, 10369, 10370, 10502, 10503]
                       Lazan   UP       [10377, 10402, 10403, 10404]
                       Orion   UP       [10409, 10449, 10450, 10474]
                      Delphi   UP       [10418, 10466, 10467, 10468]
```

Starting a Single Node (AHV)

About this task

Procedure

1. Turn on the node with one of the following methods:

   » Press the power button.

   » Use the IPMI Web Interface.

   If the node is already powered on proceed to the next step.

2. Log on to the AHV host with SSH.

3. Find the name of the Controller VM.

   ```
   root@ahv# virsh list --all | grep CVM
   ```

   Make a note of the Controller VM name in the second column.

4. Determine if the Controller VM is running.

   - If the Controller VM is off, a line similar to the following should be returned:

     ```
     -     NTNX-12AM2K470031-D-CVM        shut off
     ```

     Make a note of the Controller VM name in the second column.

   - If the Controller VM is on, a line similar to the following should be returned:

     ```
     -     NTNX-12AM2K470031-D-CVM        running
     ```

5. If the Controller VM is shut off, start it.

   ```
   root@ahv# virsh start cvm_name
   ```

   Replace `cvm_name` with the name of the Controller VM that you found from the preceding command.

   Wait approximately 5 minutes for all services to start on the Controller VM.

6. Verify that all services are up on all Controller VMs.

```
nutanix@cvm$ cluster status
```

If the cluster is running properly, output similar to the following is displayed for each node in the cluster:

```
        CVM: 10.1.64.60 Up
                                Zeus   UP       [5362, 5391, 5392, 10848, 10977, 10992]
                           Scavenger   UP       [6174, 6215, 6216, 6217]
                       SSLTerminator   UP       [7705, 7742, 7743, 7744]
                      SecureFileSync   UP       [7710, 7761, 7762, 7763]
                              Medusa   UP       [8029, 8073, 8074, 8176, 8221]
                  DynamicRingChanger   UP       [8324, 8366, 8367, 8426]
                              Pithos   UP       [8328, 8399, 8400, 8418]
                                Hera   UP       [8347, 8408, 8409, 8410]
                            Stargate   UP       [8742, 8771, 8772, 9037, 9045]
                          InsightsDB   UP       [8774, 8805, 8806, 8939]
                InsightsDataTransfer   UP       [8785, 8840, 8841, 8886, 8888, 8889, 8890]
                               Ergon   UP       [8814, 8862, 8863, 8864]
                             Cerebro   UP       [8850, 8914, 8915, 9288]
                             Chronos   UP       [8870, 8975, 8976, 9031]
                             Curator   UP       [8885, 8931, 8932, 9243]
                               Prism   UP       [3545, 3572, 3573, 3627, 4004, 4076]
                                 CIM   UP       [8990, 9042, 9043, 9084]
                        AlertManager   UP       [9017, 9081, 9082, 9324]
                            Arithmos   UP       [9055, 9217, 9218, 9353]
                             Catalog   UP       [9110, 9178, 9179, 9180]
                           Acropolis   UP       [9201, 9321, 9322, 9323]
                               Atlas   UP       [9221, 9316, 9317, 9318]
                               Uhura   UP       [9390, 9447, 9448, 9449]
                                Snmp   UP       [9418, 9513, 9514, 9516]
                   SysStatCollector   UP       [9451, 9510, 9511, 9518]
                              Tunnel   UP       [9480, 9543, 9544]
                       ClusterHealth   UP       [9521, 9619, 9620, 9947, 9976, 9977, 10301]
                              Janus   UP       [9532, 9624, 9625]
                  NutanixGuestTools   UP       [9572, 9650, 9651, 9674]
                          MinervaCVM   UP       [10174, 10200, 10201, 10202, 10371]
                       ClusterConfig   UP       [10205, 10233, 10234, 10236]
                         APLOSEngine   UP       [10231, 10261, 10262, 10263]
                               APLOS   UP       [10343, 10368, 10369, 10370, 10502, 10503]
                               Lazan   UP       [10377, 10402, 10403, 10404]
                               Orion   UP       [10409, 10449, 10450, 10474]
                              Delphi   UP       [10418, 10466, 10467, 10468]
```

7. If the node is in maintenance mode, log on to the Controller VM and take the node out of maintenance mode.

```
nutanix@cvm$ acli
```

```
<acropolis> host.exit_maintenance_mode AHV-hypervisor-IP-address
```

Replace *AHV-hypervisor-IP-address* with the AHV IP address.

```
<acropolis> exit
```

Node Startup Post-check

Check the health of the cluster after starting a node.

About this task

Procedure

1. In Prism, go to the **Health** page and select **Actions** > **Run NCC Checks**.
2. In the dialog box that appears, select **All Checks** and click **Run**.

   Alternatively, issue the following command from the CVM:

   ```
   nutanix@cvm$ ncc health_checks run_all
   ```

3. If any checks fail, see the related KB article provided in the output and the *Nutanix Cluster Check Guide: NCC Reference* for information on resolving the issue.
4. If you have any unresolvable failed checks, contact Nutanix Support.

Starting a Node in a Multinode Cluster

After updating firmware in a cluster, restart the node, using the method for your hypervisor.

Starting a Node in a Cluster (vSphere Client)

About this task

Procedure

1. Turn on the node with one of the following methods:

   » Press the power button.

   » Use the IPMI Web Interface.

   If the node is already powered on, proceed to the next step.

2. Log on to vCenter (or to the node if vCenter is not running) with the vSphere client.
3. Right-click the ESXi host and select **Exit Maintenance Mode**.
4. Right-click the Controller VM and select **Power** > **Power on**.

   Wait approximately 5 minutes for all services to start on the Controller VM.

5. Log on to any Controller VM in the cluster with SSH.
6. Verify that all services are up on all Controller VMs.

   ```
   nutanix@cvm$ cluster status
   ```

   If the cluster is running properly, output similar to the following is displayed for each node in the cluster:

   ```
           CVM: 10.1.64.60 Up
                                   Zeus   UP       [5362, 5391, 5392, 10848, 10977, 10992]
                              Scavenger   UP       [6174, 6215, 6216, 6217]
                          SSLTerminator   UP       [7705, 7742, 7743, 7744]
                        SecureFileSync   UP       [7710, 7761, 7762, 7763]
                                 Medusa   UP       [8029, 8073, 8074, 8176, 8221]
                     DynamicRingChanger   UP       [8324, 8366, 8367, 8426]
                                 Pithos   UP       [8328, 8399, 8400, 8418]
                                   Hera   UP       [8347, 8408, 8409, 8410]
                               Stargate   UP       [8742, 8771, 8772, 9037, 9045]
                             InsightsDB   UP       [8774, 8805, 8806, 8939]
                   InsightsDataTransfer   UP       [8785, 8840, 8841, 8886, 8888, 8889, 8890]
                                  Ergon   UP       [8814, 8862, 8863, 8864]
   ```

```
                    Cerebro   UP       [8850, 8914, 8915, 9288]
                    Chronos   UP       [8870, 8975, 8976, 9031]
                    Curator   UP       [8885, 8931, 8932, 9243]
                      Prism   UP       [3545, 3572, 3573, 3627, 4004, 4076]
                        CIM   UP       [8990, 9042, 9043, 9084]
               AlertManager   UP       [9017, 9081, 9082, 9324]
                   Arithmos   UP       [9055, 9217, 9218, 9353]
                    Catalog   UP       [9110, 9178, 9179, 9180]
                  Acropolis   UP       [9201, 9321, 9322, 9323]
                      Atlas   UP       [9221, 9316, 9317, 9318]
                      Uhura   UP       [9390, 9447, 9448, 9449]
                       Snmp   UP       [9418, 9513, 9514, 9516]
           SysStatCollector   UP       [9451, 9510, 9511, 9518]
                     Tunnel   UP       [9480, 9543, 9544]
              ClusterHealth   UP       [9521, 9619, 9620, 9947, 9976, 9977, 10301]
                      Janus   UP       [9532, 9624, 9625]
          NutanixGuestTools   UP       [9572, 9650, 9651, 9674]
                 MinervaCVM   UP       [10174, 10200, 10201, 10202, 10371]
              ClusterConfig   UP       [10205, 10233, 10234, 10236]
                APLOSEngine   UP       [10231, 10261, 10262, 10263]
                      APLOS   UP       [10343, 10368, 10369, 10370, 10502, 10503]
                      Lazan   UP       [10377, 10402, 10403, 10404]
                      Orion   UP       [10409, 10449, 10450, 10474]
                     Delphi   UP       [10418, 10466, 10467, 10468]
```

7. If necessary, add the host back to the metadata store.

   If the node is unavailable for more than 30 minutes, it might have been removed from metadata store. To add a host into the metadata store again, complete one of these steps.

   a. Select the target host in the diagram (Diagram view) or click the **Host tab** and select that host in the table (Table view).

   b. Click the **Enable Metadata Store** link on the right of the Summary line.

Starting a Node in a Cluster (AHV)

About this task

Procedure

1. Turn on the node with one of the following methods:

   » Press the power button.

   » Use the IPMI Web Interface.

   If the node is already powered on, proceed to the next step.

2. Log on to the AHV host with SSH.

3. Find the name of the Controller VM.

   ```
   root@ahv# virsh list --all | grep CVM
   ```

   Make a note of the Controller VM name in the second column.

4. Determine if the Controller VM is running.

   - If the Controller VM is off, a line similar to the following should be returned:

     ```
     -     NTNX-12AM2K470031-D-CVM        shut off
     ```

     Make a note of the Controller VM name in the second column.

   - If the Controller VM is on, a line similar to the following should be returned:

     ```
     -     NTNX-12AM2K470031-D-CVM        running
     ```

5. If the Controller VM is shut off, start it.

   ```
   root@ahv# virsh start cvm_name
   ```

   Replace `cvm_name` with the name of the Controller VM that you found from the preceding command.

   Wait approximately 5 minutes for all services to start on the Controller VM.

6. Log on to any Controller VM in the cluster with SSH.

7. Verify that all services are up on all Controller VMs.

   ```
   nutanix@cvm$ cluster status
   ```

   If the cluster is running properly, output similar to the following is displayed for each node in the cluster:

   ```
           CVM: 10.1.64.60 Up
                                 Zeus   UP       [5362, 5391, 5392, 10848, 10977, 10992]
                            Scavenger   UP       [6174, 6215, 6216, 6217]
                       SSLTerminator   UP       [7705, 7742, 7743, 7744]
                      SecureFileSync   UP       [7710, 7761, 7762, 7763]
                              Medusa   UP       [8029, 8073, 8074, 8176, 8221]
                 DynamicRingChanger   UP       [8324, 8366, 8367, 8426]
                              Pithos   UP       [8328, 8399, 8400, 8418]
                                Hera   UP       [8347, 8408, 8409, 8410]
                            Stargate   UP       [8742, 8771, 8772, 9037, 9045]
                          InsightsDB   UP       [8774, 8805, 8806, 8939]
                InsightsDataTransfer   UP       [8785, 8840, 8841, 8886, 8888, 8889, 8890]
                               Ergon   UP       [8814, 8862, 8863, 8864]
                             Cerebro   UP       [8850, 8914, 8915, 9288]
                             Chronos   UP       [8870, 8975, 8976, 9031]
                             Curator   UP       [8885, 8931, 8932, 9243]
                               Prism   UP       [3545, 3572, 3573, 3627, 4004, 4076]
                                 CIM   UP       [8990, 9042, 9043, 9084]
                        AlertManager   UP       [9017, 9081, 9082, 9324]
                            Arithmos   UP       [9055, 9217, 9218, 9353]
                             Catalog   UP       [9110, 9178, 9179, 9180]
                            Acropolis   UP       [9201, 9321, 9322, 9323]
                               Atlas   UP       [9221, 9316, 9317, 9318]
                               Uhura   UP       [9390, 9447, 9448, 9449]
                                Snmp   UP       [9418, 9513, 9514, 9516]
                   SysStatCollector   UP       [9451, 9510, 9511, 9518]
                              Tunnel   UP       [9480, 9543, 9544]
                       ClusterHealth   UP       [9521, 9619, 9620, 9947, 9976, 9977, 10301]
                              Janus   UP       [9532, 9624, 9625]
                  NutanixGuestTools   UP       [9572, 9650, 9651, 9674]
                          MinervaCVM   UP       [10174, 10200, 10201, 10202, 10371]
                       ClusterConfig   UP       [10205, 10233, 10234, 10236]
                         APLOSEngine   UP       [10231, 10261, 10262, 10263]
                              APLOS   UP       [10343, 10368, 10369, 10370, 10502, 10503]
                              Lazan   UP       [10377, 10402, 10403, 10404]
   ```

```
                                Orion   UP       [10409, 10449, 10450, 10474]
                               Delphi   UP       [10418, 10466, 10467, 10468]
```

8. If necessary, add the host back to the metadata store.

   If the node is unavailable for more than 30 minutes, it might have been removed from metadata store. To add a host into the metadata store again, complete one of these steps.

   a. Select the target host in the diagram (Diagram view) or click the **Host tab** and select that host in the table (Table view).

   b. Click the **Enable Metadata Store** link on the right of the Summary line.

9. If the node is in maintenance mode, log on to the Controller VM and take the node out of maintenance mode.

```
nutanix@cvm$ acli
```

```
<acropolis> host.exit_maintenance_mode AHV-hypervisor-IP-address
```

   Replace `AHV-hypervisor-IP-address` with the AHV IP address.

```
<acropolis> exit
```

Starting a Node in a Cluster (Hyper-V)

About this task

Procedure

1. If the node is off, turn it on by pressing the power button on the front. Otherwise, proceed to the next step.
2. Log on to the Hyper-V host with Remote Desktop Connection and start PowerShell.
3. Open the Failover Cluster Manager.
4. Click on **Nodes.**.
5. Right-click on the node name and select **Resume**.
6. Click the **Fail Roles Back** menu item.

   This resumes the node. It also restores the roles that were running on the node previously.

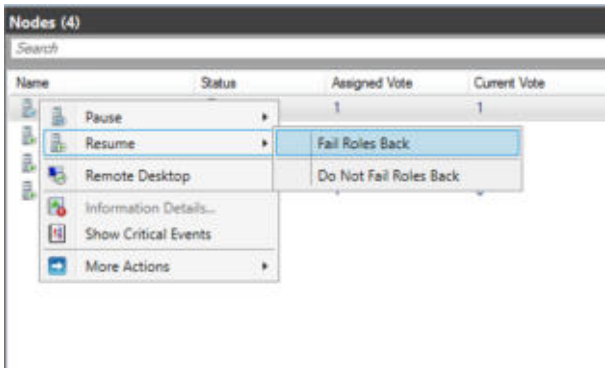


Figure 22: Resume Node Options in Failover Cluster Manager

7. From the Powershell, determine if the Controller VM is running.

```
> Get-VM | Where {$_.Name -match 'NTNX.*CVM'}
```

   - If the Controller VM is off, a line similar to the following should be returned:

     ```
     NTNX-13SM35230026-C-CVM Stopped -           -             - Opera...
     ```

     Make a note of the Controller VM name in the second column.

   - If the Controller VM is on, a line similar to the following should be returned:

     ```
     NTNX-13SM35230026-C-CVM Running 2           16384             05:10:51 Opera...
     ```

8. Start the Controller VM from the GUI or use the powershell.

```
> Start-VM -Name NTNX-*CVM
```

   Wait approximately 5 minutes for all services to start on the Controller VM.

9. Log on to any Controller VM in the cluster with SSH.

10. Verify that all services are up on all Controller VMs.

```
nutanix@cvm$ cluster status
```

   If the cluster is running properly, output similar to the following is displayed for each node in the cluster:

```
        CVM: 10.1.64.60 Up
                                Zeus   UP       [5362, 5391, 5392, 10848, 10977, 10992]
                           Scavenger   UP       [6174, 6215, 6216, 6217]
                       SSLTerminator   UP       [7705, 7742, 7743, 7744]
                      SecureFileSync   UP       [7710, 7761, 7762, 7763]
                              Medusa   UP       [8029, 8073, 8074, 8176, 8221]
                  DynamicRingChanger   UP       [8324, 8366, 8367, 8426]
                              Pithos   UP       [8328, 8399, 8400, 8418]
                                Hera   UP       [8347, 8408, 8409, 8410]
                            Stargate   UP       [8742, 8771, 8772, 9037, 9045]
                          InsightsDB   UP       [8774, 8805, 8806, 8939]
                 InsightsDataTransfer   UP       [8785, 8840, 8841, 8886, 8888, 8889, 8890]
                               Ergon   UP       [8814, 8862, 8863, 8864]
                            Cerebro   UP       [8850, 8914, 8915, 9288]
                            Chronos   UP       [8870, 8975, 8976, 9031]
                            Curator   UP       [8885, 8931, 8932, 9243]
                              Prism   UP       [3545, 3572, 3573, 3627, 4004, 4076]
                                CIM   UP       [8990, 9042, 9043, 9084]
                       AlertManager   UP       [9017, 9081, 9082, 9324]
                           Arithmos   UP       [9055, 9217, 9218, 9353]
                            Catalog   UP       [9110, 9178, 9179, 9180]
                          Acropolis   UP       [9201, 9321, 9322, 9323]
                              Atlas   UP       [9221, 9316, 9317, 9318]
                              Uhura   UP       [9390, 9447, 9448, 9449]
                               Snmp   UP       [9418, 9513, 9514, 9516]
                    SysStatCollector   UP       [9451, 9510, 9511, 9518]
                             Tunnel   UP       [9480, 9543, 9544]
                       ClusterHealth   UP       [9521, 9619, 9620, 9947, 9976, 9977, 10301]
                              Janus   UP       [9532, 9624, 9625]
                  NutanixGuestTools   UP       [9572, 9650, 9651, 9674]
                         MinervaCVM   UP       [10174, 10200, 10201, 10202, 10371]
                       ClusterConfig   UP       [10205, 10233, 10234, 10236]
                         APLOSEngine   UP       [10231, 10261, 10262, 10263]
                              APLOS   UP       [10343, 10368, 10369, 10370, 10502, 10503]
                              Lazan   UP       [10377, 10402, 10403, 10404]
                              Orion   UP       [10409, 10449, 10450, 10474]
```

```
                                Delphi   UP       [10418, 10466, 10467, 10468]
```

11. If necessary, add the host back to the metadata store.

    If the node is unavailable for more than 30 minutes, it might have been removed from metadata store. To add a host into the metadata store again, complete one of these steps.

    a. Select the target host in the diagram (Diagram view) or click the **Host tab** and select that host in the table (Table view).

    b. Click the **Enable Metadata Store** link on the right of the Summary line.

Node Startup Post-check

Check the health of the cluster after starting a node.

About this task

Procedure

1. In Prism, go to the **Health** page and select **Actions** > **Run NCC Checks**.
2. In the dialog box that appears, select **All Checks** and click **Run**.

   Alternatively, issue the following command from the CVM:

   ```
   nutanix@cvm$ ncc health_checks run_all
   ```

3. If any checks fail, see the related KB article provided in the output and the *Nutanix Cluster Check Guide: NCC Reference* for information on resolving the issue.
4. If you have any unresolvable failed checks, contact Nutanix Support.

2

CHANGING AN IPMI IP ADDRESS

Reset the IPMI address on a node. These steps are for NX- (Nutanix) nodes. For nodes from other hardware vendors, see the manufacturer instructions.

Before you begin

Ensure that the cluster is healthy. Resolve any critical alerts and run Nutanix Cluster Check (NCC) if necessary. If you are unfamiliar with restarting cluster services, contact Nutanix Support.

About this task

For initial setup, perform these steps once for every IPMI interface in the cluster. Complete the entire procedure on an interface before proceeding to the next interface. If you are reconfiguring the IPMI IP address because you have replaced a node, reconfigure the IPMI address only for that node.

> Note: Restart Genesis after changing the IPMI configuration. Otherwise, the cluster does not have access to the IPMI interface.

Procedure

1. Configure the IPMI IP addresses by using either the IPMI web interface or the hypervisor host command-line interface.

   » Configuring the Remote Console IP Address (IPMI Web Interface) on page 57

   » Configuring the Remote Console IP Address (Command Line) on page 57

   Alternatively, you can configure the IP address in the BIOS by following Configuring the Remote Console IP Address (BIOS) on page 59.

   > CAUTION: If you are unfamiliar with restarting Genesis, contact Nutanix support.

2. Restart Genesis with one of the following commands:

   » If you replace a node with a new node, you need to update the cluster with the IPMI address of the new node. You only need to restart Genesis on the Controller VM for the new node, using the `genesis restart` command:

   ```
   nutanix@cvm$ genesis restart
   ```

   » If you will restart every Controller VM in the cluster, log on to any Controller VM in the cluster and restart Genesis.

   ```
   nutanix@cvm$ allssh genesis restart
   ```

   If the restart is successful, output similar to the following is displayed:

   ```
   Stopping Genesis pids [1933, 30217, 30218, 30219, 30241]
   ```

```
Genesis started on pids [30378, 30379, 30380, 30381, 30403]
```

Configuring the Remote Console IP Address (IPMI Web Interface)

About this task

Procedure

1. Sign in to the IPMI web console.
2. Go to **Configuration** > **Network**.
3. Select DHCP or enter the new static IP address in the IPv4 Setting section.

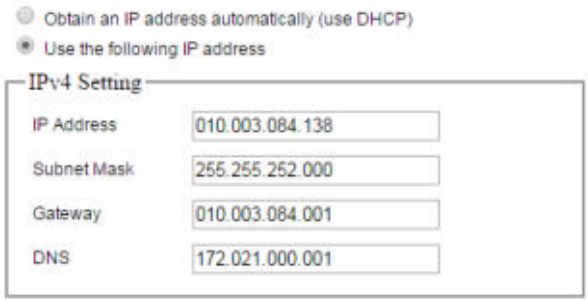


Figure 23: IPMI Network Configuration

4. Click **Save**.
   The new IPv4 configuration takes effect. This change terminates your connection to the web interface. To start a new connection, go to the new IP address of the IPMI interface.
5. Log on to every Controller VM in the cluster and restart Genesis.

   > Note: If you are unfamiliar with restarting Genesis, contact Nutanix support.

   > Note: If you are reconfiguring IPMI address on a node because you have replaced the motherboard, restart Genesis on the Controller VM only for that node.

   ```
   nutanix@cvm$ genesis restart
   ```

   If the restart is successful, output similar to the following is displayed:

   ```
   Stopping Genesis pids [1933, 30217, 30218, 30219, 30241]
   Genesis started on pids [30378, 30379, 30380, 30381, 30403]
   ```

Configuring the Remote Console IP Address (Command Line)

About this task

You can configure the management interface from the hypervisor host on the same node.

Perform these steps once from each hypervisor host in the cluster where you want to change the management network configuration.

Procedure

1. Log on to the hypervisor host with SSH (vSphere or AHV) or remote desktop connection (Hyper-V).
2. Set the networking parameters.

   » vSphere

   ```
   root@esx# /ipmitool lan set 1 ipsrc static
   root@esx# /ipmitool lan set 1 ipaddr mgmt_interface_ip_addr
   root@esx# /ipmitool lan set 1 netmask mgmt_interface_subnet_addr
   root@esx# /ipmitool lan set 1 defgw ipaddr mgmt_interface_gateway
   ```

   » Hyper-V

   ```
   > ipmiutil lan -e -I mgmt_interface_ip_addr -G mgmt_interface_gateway
    -S mgmt_interface_subnet_addr
   ```

   » AHV

   ```
   root@ahv# ipmitool lan set 1 ipsrc static
   root@ahv# ipmitool lan set 1 ipaddr mgmt_interface_ip_addr
   root@ahv# ipmitool lan set 1 netmask mgmt_interface_subnet_addr
   root@ahv# ipmitool lan set 1 defgw ipaddr mgmt_interface_gateway
   ```

   - Replace `mgmt_interface_ip_addr` with the new IP address for the remote console.
   - Replace `mgmt_interface_gateway` with the gateway IP address.
   - Replace `mgmt_interface_subnet_addr` with the subnet mask for the new IP address.

3. Show current settings.

   » vSphere

   ```
   root@esx# /ipmitool -v lan print 1
   ```

   » Hyper-V

   ```
   > ipmiutil lan -r
   ```

   » AHV

   ```
   root@ahv# ipmitool -v lan print 1
   ```

   Confirm that the parameters are set to the correct values.

4. Log on to every Controller VM in the cluster and restart Genesis.

   > Note: If you are unfamiliar with restarting Genesis, contact Nutanix support.

   > Note: If you are reconfiguring IPMI address on a node because you have replaced the motherboard, restart Genesis on the Controller VM only for that node.

   ```
   nutanix@cvm$ genesis restart
   ```

   If the restart is successful, output similar to the following is displayed:

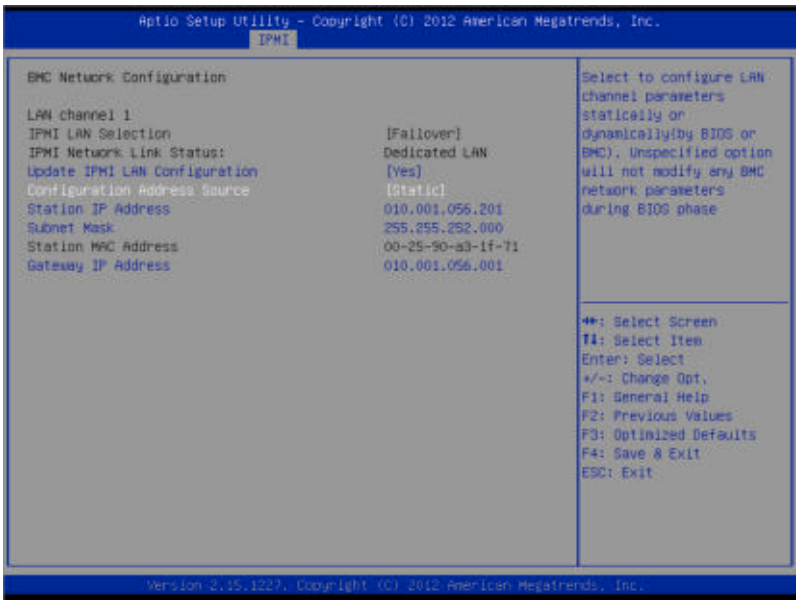
   ```
   Stopping Genesis pids [1933, 30217, 30218, 30219, 30241]
   Genesis started on pids [30378, 30379, 30380, 30381, 30403]
   ```

Configuring the Remote Console IP Address (BIOS)

About this task

Procedure

1. Connect a keyboard and monitor to a node in the Nutanix block.
2. Restart the node and press **Delete** to enter the BIOS setup utility.
   There is a limited amount of time to enter BIOS before the host completes the restart process.
3. Press the right arrow key to select the **IPMI** tab.
4. Press the down arrow key until **BMC network configuration** is highlighted and then press **Enter**.
5. Press down the arrow key until **Update IPMI LAN Configuration** is highlighted and press **Enter** to select **Yes**.
6. Select **Configuration Address source** and press **Enter**.
7. Select **Static** and press **Enter**.
8. Assign the **Station IP address**, **Subnet mask**, and **Router IP address**.



9. Review the BIOS settings and press **F4** to save the configuration changes and exit the BIOS setup utility.
   The node restarts.

10. Log on to every Controller VM in the cluster and restart Genesis.

> Note: If you are unfamiliar with restarting Genesis, contact Nutanix support.

> Note: If you are reconfiguring IPMI address on a node because you have replaced the motherboard, restart Genesis on the Controller VM only for that node.

```
nutanix@cvm$ genesis restart
```

If the restart is successful, output similar to the following is displayed:

```
Stopping Genesis pids [1933, 30217, 30218, 30219, 30241]
Genesis started on pids [30378, 30379, 30380, 30381, 30403]
```

3

CHANGING THE IPMI PASSWORD

About this task

> Note: This procedure helps prevent the BMC password from being retrievable on port 49152.

IPMI Password requirements: - Required password length: 8 to 20 characters - Password can not be the reverse of the user name - Password must include characters from at least 3 of the listed character classes - Allowed character classes are: a - z, A - Z, 0 - 9, and special characters.

> Note: The special characters for -G8 platforms are listed in the next topic *IPMI Password Requirements for Nutanix G8 Platforms*.

> Note: BMC 7.08 changes the default IPMI password so that every node ships from the factory with a unique password. The new default IPMI credentials are `username = ADMIN` and `password = node-serial-number`. To find the serial number, do one of the following. From the host, issue the command `ipmitool fru print`. In the output, search for Board serial. The Board serial value is the IPMI password. Look for the sticker on the node that reads IPMI PW/SN. That string is the default IPMI password.

Procedure

Change the administrative user password of all IPMI hosts.

Perform these steps on every IPMI host in the cluster.

a. Sign in to the IPMI web interface as the administrative user.

b. Click **Configuration**.

c. Click **Users**.

d. Select the administrative user and then click **Modify User**.

e. Type the new password in both text fields and then click **Modify**.

f. Click **OK** to close the confirmation window.

IPMI Password Requirements for Nutanix G8 Platforms

This topic lists the IPMI password requirements for NX- series G8 platforms.

Special Characters for NX Series G8 Platforms

These 29 special characters are supported in IPMI passwords for G8 platforms:

```
` ~ ! @ $ % ^ & * ( ) = [ ] \ / _ + ; , . { } < > ? | # -
```

Three special characters below are not supported on G8 platforms:

```
' : "
```

Changing the IPMI Password for ESXi

If you do not know the password of IPMI, but have root access to ESXi host and want to change the password of the IPMI, perform the following procedure. You can use the IPMI tool from SSH as a root to do an in-band reset.

Procedure

1. Log into the ESXi host with SSH.
2. Determine the user ID of the administrator for which you want to change the password.

   ```
   root@esx# /ipmitool user list
   ```

   A sample output is as follows.

   ```
   ID Name Callin Link Auth IPMI Msg Channel Priv Limit
   2 ADMIN true false false Unknown (0x00)
   ```

   In the sample output, the ID of the administrator for which you want to change the password is 2.
3. Set the new password.

   ```
   root@esx# /ipmitool user set password user_id new_password
   ```

   > Note: If you want to use more than 16 characters, you must append the parameter "20" and the maximum is 20 characters.
   >
   > ```
   > root@esx# /ipmitool user set password user_id new_password 20
   > ```

Changing the IPMI Password for Hyper-V

If you have administrator access to a Hyper-V host, you can change the IPMI password using an in-band reset, even if you do not know the current password.

Procedure

1. Log on to the Hyper-V host with Remote Desktop Connection and start PowerShell.
2. Determine the user ID of the administrator for which you want to change the password.

   ```
   PS C:\Program Files\Nutanix\ipmicfg> .\IPMICFG-Win.exe -user list
   ```

   A sample output is as follows.

   ```
   Maximum number of Users          : 10
   Count of currently enabled Users : 1
   User ID | User Name     |Privilege Level | Enable
   --------  ----------    |--------------- | -------
         2 | ADMIN         | Administrator  | Yes
   ```

   In the sample output, the user ID of the administrator for which you want to change the password is 2.

3. Set the new password.

```
PS C:\Program Files\Nutanix\ipmicfg> .\IPMICFG-Win.exe -user setpwd user_id new_password
```

> Note: If you want to use more than 16 characters, you must append the parameter "20" and the maximum is 20 characters.
>
> ```
> PS C:\Program Files\Nutanix\ipmicfg> .\IPMICFG-Win.exe -user
> setpwd user_id new_password 20
> ```

Changing the IPMI Password for AHV

If you do not know the password of IPMI, but have root access to AHV host and want to change the password of the IPMI, perform the following procedure. You can use the IPMI tool from SSH as a root to do an in-band reset.

Procedure

1. Log into the AHV host with SSH.
2. Determine the user ID of the administrator for which you want to change the password.

   ```
   root@ahv# ipmitool user list
   ```

   A sample output is as follows.

   ```
   ID Name Callin Link Auth IPMI Msg Channel Priv Limit
   2 ADMIN true false false Unknown (0x00)
   ```

   In the sample output, the ID of the administrator for which you want to change the password is 2.
3. Set the new password.

   ```
   root@ahv# ipmitool user set password user_id new_password
   ```

   > Note: If you want to use more than 16 characters, you must append the parameter "20" and the maximum is 20 characters.
   >
   > ```
   > root@ahv# ipmitool user set password user_id new_password 20
   > ```

Changing the IPMI Password without Operating System Access

Change the IPMI password.

About this task

If you do not know the IPMI password, and you do not have operating system access, you can contact Nutanix support to reset the IPMI password using a bootable DOS environment on a USB drive.

Procedure

1. Obtain a bootable USB drive.
2. Contact Nutanix support to reset the IPMI password.

4

FAILED COMPONENT DIAGNOSIS

Recognize key component failures of the Nutanix platforms.

SATA SSD Boot Drive

Applicable Platforms

All

Failure Indications

- The console of the Controller VM shows:
  - Repeated Ext4 filesystem errors.
  - The words `hard reset failed` or `kernel panic`
  - I/O errors on the `/dev/sda` device
- An alert on the Nutanix UI that a SATA HDD is marked offline.
- There are errors in the Cassandra log file indicating an issue reading the metadata.

Next Steps:

- If the Controller VM is showing errors, power the Controller VM down, reseat the SSD, and start the Controller VM. If that does not resolve the issue, replace the drive.
- If an alert indicates that the disk is marked offline, run the command

  ```
  ncli> disk ls id=xx
  ```

  (where *xx* is the ID shown in the alert) to verify the location before replacing the drive.

  - Storage tier: SSD-SATA
  - Location: 1
  - Online: false
- If there are errors indicating metadata problems, replace the drive.

Rescue Shell

Log on to a rescue shell to diagnose issues with the boot device.

To log on to a rescue shell, you must first create the `svmrescue.iso` image on another node.

Creating the Controller VM Recovery Image (Hyper-V)

About this task

Procedure

1. Log on to another Controller VM in the cluster with SSH.
2. Create the ISO image.

   ```
   nutanix@cvm$ cd ~/data/installer/*version*
   nutanix@cvm$ ./make_iso.sh svmrescue Rescue 10
   ```

   Replace *version* with the AOS version of the cluster.

   The `make_iso` command generates the `svmrescue.iso` file in the `/home/nutanix/data/installer` directory.

Launching the Recovery Shell

About this task

> CAUTION: This procedure is for the boot drive replacement in slot 1 of the node. Do not install the Controller VM on a metadata drive in slot 2 of the node.

Procedure

1. Log on to the ESXi host as root with the vSphere client.
2. If the Controller VM is running, right-click the Controller VM and select **Power** > **Power Off**.

3. Set up the BIOS.

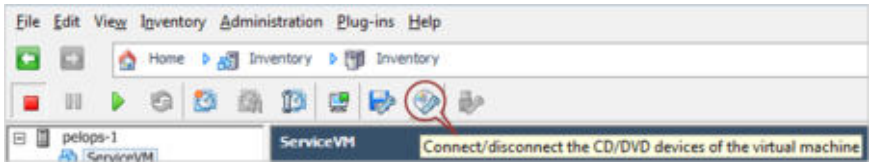
   a. Right-click the Controller VM and select **Edit Settings** /**Hardware** > **CD/DVD Drive** > **Device Type** > **Client Device**.

   b. Select **Options** > **Advanced** > **Boot Options** > **Force BIOS Setup** > **OK**.

   c. Click the **Console** tab of the Controller VM.

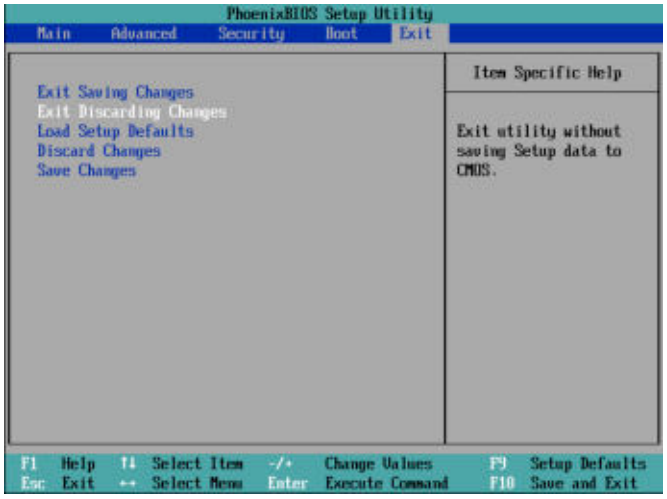
   d. Right-click the Controller VM and select **Power** > **Power On**.

   e. When the Controller VM restarts into BIOS, select the **Connect/Disconnect the CD/DVD devices of the virtual machine** icon at the top of the console and select **CD/DVD Drive** > **Connect to ISO image on local disk**.



   f. Select the `svmrescue.iso` file on your local system and click **OK**.

   g. In the console, press **Esc** and select **Exit Discarding Changes** to exit the BIOS.



4. Choose **Rescue Shell** from the boot menu and press **Enter**.

5. Select from the following choices.

   » **Start Nutanix Controller VM**

   » **Rescue Nutanix Controller VM**.

   > Note: This option reimages the Controller VM, but keeps the metadata (oplog/extent store) and cold data.

   » **Factory Deploy Nutanix Controller VM**.

   > CAUTION: Exercise caution before selecting this option as this option formats and reimages all the disks.

   » **Rescue Shell**. Starts the rescue shell utility.

Rescue Shell Commands

List of rescue shell commands.

Inside the rescue shell you can run the following tools to see if the disk is readable:

`parted`

View the partitions on a drive.

> CAUTION: Only use the `print` command within `parted`. Using other commands could destroy the drive.

As an example, here is what running the `parted` command looks like on an NX-3451:

```
# parted /dev/sda
```

At the resulting prompt, type `print`.

```
(parted) print
```

```
Model: ATA INTEL SSDSC2BA80 (scsi)
Disk /dev/sda: 800GB
Sector size (logical/physical): 512B/512B
Partition Table: msdos

Number  Start   End     Size    Type     File system  Flags
 1      1049kB  10.7GB  10.7GB  primary  ext4
 2      10.7GB  21.5GB  10.7GB  primary  ext4
 3      21.5GB  64.4GB  42.9GB  primary  ext4
 4      64.4GB  800GB   736GB   primary  ext4
```

`lsscsi`

This command lists the SCSI devices presented to the Controller VM. The boot drive is listed as `/dev/sda`.

As an example, here is what running the `lsscsi` command looks like on an NX-3450:

```
# lsscsi
```

```
[2:0:0:0]    disk    ATA      INTEL SSDSC2BA80 0250  /dev/sda
[2:0:1:0]    disk    ATA      INTEL SSDSC2BA80 0250  /dev/sdb
[2:0:2:0]    disk    ATA      ST91000640NS     SN03  /dev/sdc
[2:0:3:0]    disk    ATA      ST91000640NS     SN03  /dev/sdd
[2:0:4:0]    disk    ATA      ST91000640NS     SN03  /dev/sde
[2:0:5:0]    disk    ATA      ST91000640NS     SN03  /dev/sdf
```

On the NX-1000/NX-3050/NX-6000 series, the boot disk is divided into four partitions as follows:

- `sda1`: root partition
- `sda2`: alternate root partition (for upgrades)
- `sda3`: `/home/nutanix`
- `sda4`: Stargate extent data and metadata

Exiting the Rescue Shell

About this task

To exit the rescue shell, perform the following steps:

Procedure

1. Shut down the Controller VM.
2. Right-click the Controller VM and click **Edit Settings**.
3. Select **Hardware** > **CD/DVD Drive**.
4. Choose **Device Type** > **Datastore ISO File**.
5. Click **Browse** and locate the ServiceVM ISO file on the local datastore (for example: `[NTNX-local-ds-nfs-1-4] ServiceVM-1.25_Centos/ServiceVM-1.25_Centos.iso`). Do not select the file or folder or file that starts with .# (period and pound symbols).
6. Select the ServiceVM ISO file and click **OK**.
7. Select **Device Status** > **Connect at power on** and click **OK**.
8. Right-click the Controller VM and select **Power** > **Power On**. It might take a few minutes for the logon prompt to appear.

SATA HDD or SATA SSD Data Drive

Applicable Platforms

All

Failure Indications

All platforms:

An alert on the Nutanix UI that a disk is marked offline.

The Stargate process marks a disk offline and generates an alert when I/O stops for more than 20 seconds to the drive.

Next Steps:

Run the command

```
ncli> disk ls id=xx
```

(where `xx` is the ID shown in the alert) and verify the following:

- Storage tier: SSD-SATA
- Location: [2-6]
- Online: false

If several drives in the same node are marked offline within a short time, the SATA controller on the node may be bad. Verify this by running the `sudo smartctl -a /dev/disk` command on the device and checking for any errors. The most important things to check are the self-assessment results and the SMART error log.

```
nutanix@cvm$ sudo smartctl -a /dev/sdc
```

```
SMART overall-health self-assessment test result: PASSED

SMART Error Log Version: 1
No Errors Logged
```

If there are no errors on the drives that have been marked offline, then the SATA controller may be bad. Nutanix recommends replacing the node first, then marking the disks online with the following commands:

```
nutanix@cvm$ ncli -h true
ncli> disk mark-online id=disk_id
```

Node

Node failure indications.

Applicable Platforms

All platforms

Failure Indications

All platforms:

- When trying to start the node, there is no green light on the power button. A green light indicates power to the node. There may or may not be network/BMC link lights.
- ESXi experiences a PSOD that indicates a CPU or other hardware error.
- One of the on-board NIC ports is not working.
- A diagnosed memory failure turns out to be a memory slot failure.
- Multiple HDDs have gone offline in a single node but the drives do not report any errors.

Next Steps:

All platforms:

- Multi-node platforms: If the node does not start, reseat the node. If that does not resolve the issue, replace the node.
- If the node is on but one or more of the other symptoms are present, replace the node.
- To troubleshoot NIC issues, see KB article 1088.

Chassis or Node Fan

Fan failure indications.

Applicable Platforms

All platforms

Failure Indications

All platforms:

- An alert on the Nutanix UI that a fan has stopped or is running too fast or too slow.
- Running `/ipmitool sel list` from the ESXi host shows fan errors on fan 1 or fan 2.
- Running `/ipmitool sensor list` from the ESXi host shows 0 RPM for either fan 1 or fan 2 OR shows significantly fewer RPM than another fan, AND there are temperature alerts.

- All platforms except NX-6000 and NX-9040: ignore failure reports from fan 3 or fan 4. Each node only sees two fans, labeled on all nodes as fan 1 and fan 2. See KB article 1036 for instructions on how to remove these invalid fans from the sensor output.
- NX-6000 and NX-9040: failure reports from fan 3 or fan 4 refer to the node fan.

Next Steps:

All platforms:

If the fan speed still appears as 0 RPM after you replace the fan, there is probably an issue with the PDB sensor. Replace the chassis.

Memory

Applicable Platforms

All platforms

Failure Indications

All platforms:

- vCenter **Alarms** or **Hardware Status** shows a memory alert.
- NCC returns a correctable-ECC error, such as one of the following:

```
1344   @classmethod

1345   @register_check("15019")

1346   def check_sel_correctable_ecc_1day(cls, host_uptime):

1347     # Check for number of Correctable ECC errors in 1 day.

1348     fail_msg = cls.check_dimm_correctable_ecc_errors(1,
host_uptime,
1349
FLAGS.correctable_ecc_errors_threshold_per_day_critical,
1350
DataExchangeProto.kSelLogCorrectableEccErrorsHighOneDay)
```

```
1353   @classmethod

1354   @register_check("15020")

1355   def check_sel_correctable_ecc_10days(cls, host_uptime):

1356     # Check for number of Correctable ECC errors in 10
days.
1357     fail_msg = cls.check_dimm_correctable_ecc_errors(10,
host_uptime,
1358
FLAGS.correctable_ecc_errors_threshold_ten_days_critical,
1359
DataExchangeProto.kSelLogCorrectableEccErrorsHighTenDays)
```

- IPMI event log (`/ipmitool sel list` from the ESXi shell) for a node shows an uncorrectable ECC memory error for a particular DIMM.
- The host does not detect all memory. For example, the system has 96 GB per node and the host only shows 88 GB.

Next Steps:

All platforms:

- Correctable/Uncorrectable ECC errors

  If the IPMI event log shows errors as described in the preceding section on failure indications, replace the DIMM. View the event log on the IPMI web console page to correctly identify the DIMM slot.

  Each memory replacement guide has the server board schematic for the corresponding platform.

- Undetected memory

  If the IPMI event log shows errors as described in the preceding section on failure indications, replace the DIMM.

  If the host is not detecting all memory, run the following command from the hypervisor to determine the location of the uninstalled DIMM. Use the configuration table to figure out which slots are populated.

  - ESXi:

    ```
    root@esx# smbiosDump | grep -A 12 -B1 'Bank: ' | egrep \
     'Bank|No Memory Installed|Location'
    ```

  - AHV:

    ```
    root@ahv# virsh sysinfo | egrep "size|'locator"
    ```

The specifications tables for each platform describe the supported DIMM configurations.

Power Supply

Power supply failure indications.

Applicable Platforms

All platforms

Failure Indications

All platforms:

- Nutanix UI shows a power supply alert.
- The **Hardware Status** tab in the vSphere client shows a power supply alert.
- IPMI event log for node A shows `Power Supply Failure Detected - Asserted`.
- Alert light on the front of node A flashes at 4-second intervals.
- Alert light on the back of the power supply is orange or unlit.
- All nodes in a block fail (both power supplies fail).

Next Steps:

Run the `/home/nutanix/serviceability/bin/breakfix.py` script to determine which power supply has failed, then replace the PSU.

Chassis

Chassis failure indications.

Applicable Platforms

All platforms

Failure Indications

All platforms:

- Errors or failures on multiple drives, that are not resolved by replacing the drives or the node.
- Errors or a failure on a single drive that are not resolved by replacing the drive.
- PDB sensor failure, after diagnosing a failed PSU or fan indicates that a sensor has failed, not the actual component.
- Physical damage to the chassis.

Next Steps:

All platforms:

Replace the chassis.

5

ADDING A DRIVE

Add a drive to a platform.

About this task

> Note: The process of adding a drive is the same for all platforms (both Nutanix and third-party platforms), assuming the platform is running Nutanix AOS.

What types of drives you can add depends on your platform configuration. For supported drive configurations, see the system specifications for your platform.

- Hybrid: a mixture of SSDs and HDDs. Hybrid configurations fill all available drive slots, so the only case where you would add a drive is if there is a drive missing.
- All-flash: All-flash nodes have both fully populated and partially populated configurations. You can add new drives to the empty slots.
  - All-flash nodes can accept only SSDs.
  - All-flash nodes can contain only even numbers of drives.
- SSD with NVMe: a mixture of SSDs and NVMe drives. Only certain drive slots can contain NVMe drives. Consult the system specifications for your platform for drive configurations.

When adding more than one drive to a node, allow at least one minute between adding each drive.

Procedure

1. Insert the drive in an empty slot.

2. Log on to the web console, go to **Hardware** > **Diagram** , and select the added drive to view the details.

Figure 24: Added Drive (Multi-Node Block)

If the drive is red and shows a label of `Unmounted Disk`, select the drive and click **Repartition and Add** under the diagram.

This message and the button appear only if the replacement drive contains data. Their purpose is to protect you from unintentionally using a drive with data on it.

> CAUTION: This action removes all data on the drive. Do not repartition the drive until you have confirmed that the drive contains no essential data.

3. From the web console **Summary** > **Disk Details** field, verify that the disk has been added to the original storage pool.

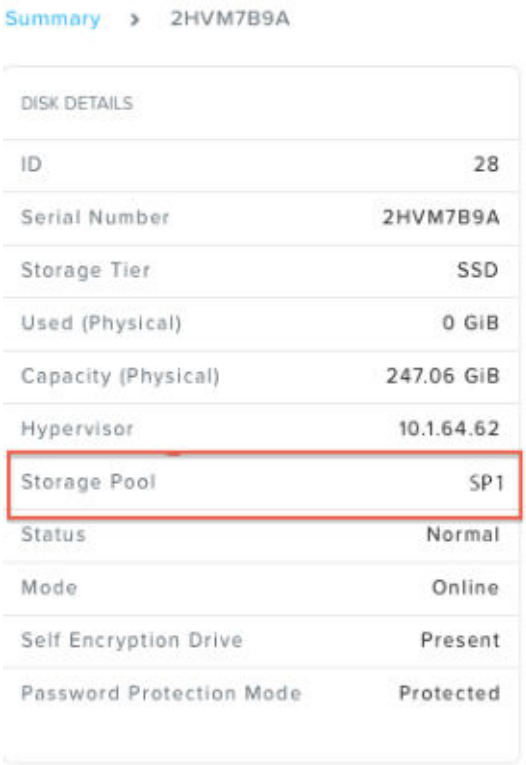


Figure 25: Disk Details

If the cluster has only one storage pool, the disk is automatically added to the storage pool.

4. If the drive is not automatically added to the storage pool (because the cluster has more than one), add it to the desired storage pool.

   a. In the web console, select **Storage** from the pull-down main menu (upper left of screen) and then select the **Table** and **Storage Pool** tabs.

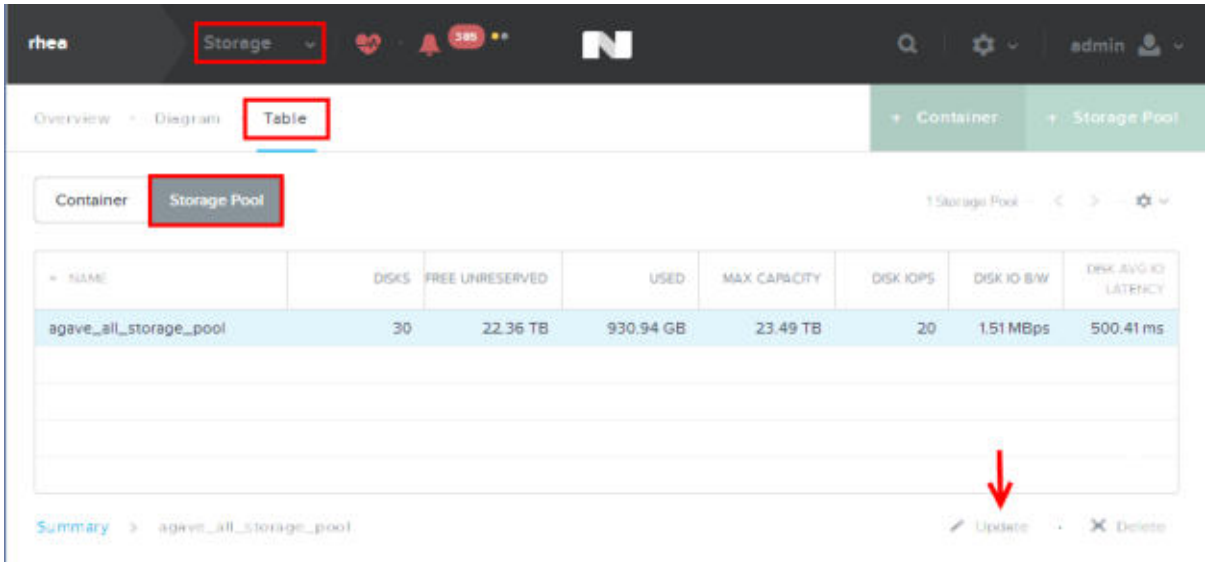


   Figure 26: Storage Pool Table View

   b. Select the target storage pool and then click **Update**.
   The Update Storage Pool window appears.

   c. In the **Capacity** field, check the **Use unallocated capacity** box to add the available unallocated capacity to this storage pool then click **Save**.

   d. Go back to **Hardware** > **Diagram**, select the drive, and confirm that it is in the correct storage pool.

6

REMOTE DIRECT MEMORY ACCESS

Details of RDMA support.

Remote direct memory access (RDMA) gives a node direct access to the memory subsystems of other nodes in the cluster, without involving the CPU-bounded network stack of the operating system. RDMA allows low-latency data transfer between memory subsystems, and so improves network latency and lowers CPU use.

Nutanix currently supports the use of RDMA-enabled network cards for the following platforms.

G5 platforms	NX-9030-G5
G6 platforms	• NX-3060-G6 • NX-3155G-G6 • NX-3170-G6 • NX-8035-G6 • NX-8155-G6
G7 platforms	• NX-3060-G7 • NX-3155G-G7 • NX-3170-G7 • NX-8035-G7 • NX-8150-G7 • NX-8155-G7 • NX-8170-G7
G8 platforms	• NX-1065-G8 • NX-3060-G8 • NX-3155G-G8 • NX-3170-G8 • NX-8150-G8 • NX-8155-G8 • NX-8170-G8

Use of RDMA in a Nutanix platform must meet the following conditions.

- Each node in the cluster must contain two RDMA-enabled network cards.

G5 platforms	Mellanox CX4
G6 platforms	Mellanox CX4
G7 platforms	• Mellanox CX4 • Mellanox CX5
G8 platforms	Mellanox CX5 25 GB (part number MCX512A-ACUT)

Note: Mellanox CX-type NICs are dual-port cards. One card uses a single port dedicated to RDMA traffic. The other card uses its ports for CVM and guest VM traffic.

Note: All network cards in a node must be identical. Nutanix does not support mixing network cards from different manufacturers, or of different capacities.

- RDMA-enabled cards must be installed at the factory. You cannot add them to a node in the field.
- RDMA software support:

G5 platforms	AHV and ESXi

G6 and G7 platforms	• AHV
	• ESXi (starting with AOS 5.11.2)
	  • ESXi 6.5: 6.5U1 or later
	  • ESXi 6.7: 6.7U1 or later
	  • ESXi 7.0

G8 platforms	• Foundation 5.0.4
	• AOS 5.20.1.1
	• AHV 7.2 and later
	• ESXi
	  • 6.7U3b
	  • 7.0 U2a and later

- RDMA nodes cannot mix with non-RDMA nodes in the same cluster.
- AOS does not enable datacenter bridging (DCB) automatically, in order to avoid overwriting any existing switch configuration. Enable DCB on the customer switch manually.

7

CMOS BATTERY REPLACEMENT

Nutanix can not ship batteries. Air freight regulations and international government agencies restrict the shipment of uninstalled lithium batteries. Customers can purchase and replace the CMOS battery if it fails. Nutanix can help install batteries after a customer has purchased the replacement battery.

> Warning: DANGER OF EXPLOSION IF THE BATTERY IS INCORRECTLY REPLACED.
>
> Replace only with the same or equivalent type recommended by the manufacturer. Dispose of used batteries according to manufacturer instructions.

Only use the following batteries for the platforms shown:

- For G5 and earlier platforms: KTS brand CR2032 3V or a reliable equivalent
- For G6 and later platforms: KTS brand BR2032 3V or a reliable equivalent

For platforms that use X10 or later motherboards, which includes all G4 and later platforms, the only function of the CMOS battery is to preserve BIOS settings. A CMOS battery failure does not affect power control.

For older pre-G4 platforms that use X8 or X9 motherboards, the node does not start if the CMOS battery has failed.

Find the battery replacement documentation for each platform on the support portal by selecting **Hardware Replacement Documentation** > **Battery** > **(Platform Generation)** > **(Platform Name)** (for example: **G7** > **NX-3060-G7**).

8

MEMORY CONFIGURATIONS

Supported Memory Configurations (G4 and G5 Platforms)

DIMM installation order for all Nutanix G4 and G5 platforms. When removing, replacing, or adding memory, use the rules and guidelines in this topic.

DIMM Restrictions for G4 and G5 Platforms

For DIMM restrictions see the Product Mixing Restrictions topic in this NX Series Hardware Administration Guide.

Memory Installation Order for All G4 and G5 Platforms

A memory channel is a group of DIMM slots.

Each CPU is associated with four memory channels. Memory channels contain either two or three DIMM slots, depending on the motherboard.

Two DIMMs per channel (2DPC) memory channels have one blue slot and one black slot each, as shown in the following figure.

> Note: DIMM slots on the motherboard are most commonly labeled as A1, A2, and so on. However, some software tools report DIMM slot labels in a different format, such as 1A, 2A, or CPU1, CPU2, or DIMM1, DIMM2.

X10DRT Motherboards

Table 1: X10DRT Platforms with Two DIMMs per Channel (2DPC)

NX-1065-G4/G5	NX-1065S-G5 (Single CPU)	NX-1175S-G5 (Single CPU)
NX-3060-G4/G5	NX-6035-G4/G5	NX-6035C-G5
NX-8035-G4/G5	NX-9030-G5	NX-9060-G4
SX-1065-G5		

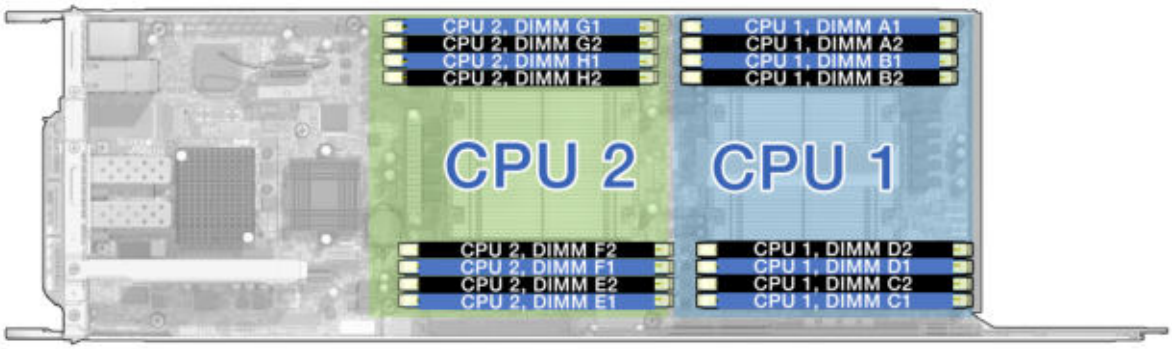


Figure 27: Multi-node Motherboard with 2DPC Memory Channels

Table 2: X10DRT Dual CPU platforms. (2DPC)

Number of CPUs	Number of DIMMs	Slots to use
2	4	A1, B1, E1, F1 (blue slots)
2	6	A1, B1, C1, E1, F1, G1 (blue slots)
2	8	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots)
2	12	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots) A2, B2, E2, F2 (black slots)
2	16	Fill all slots.

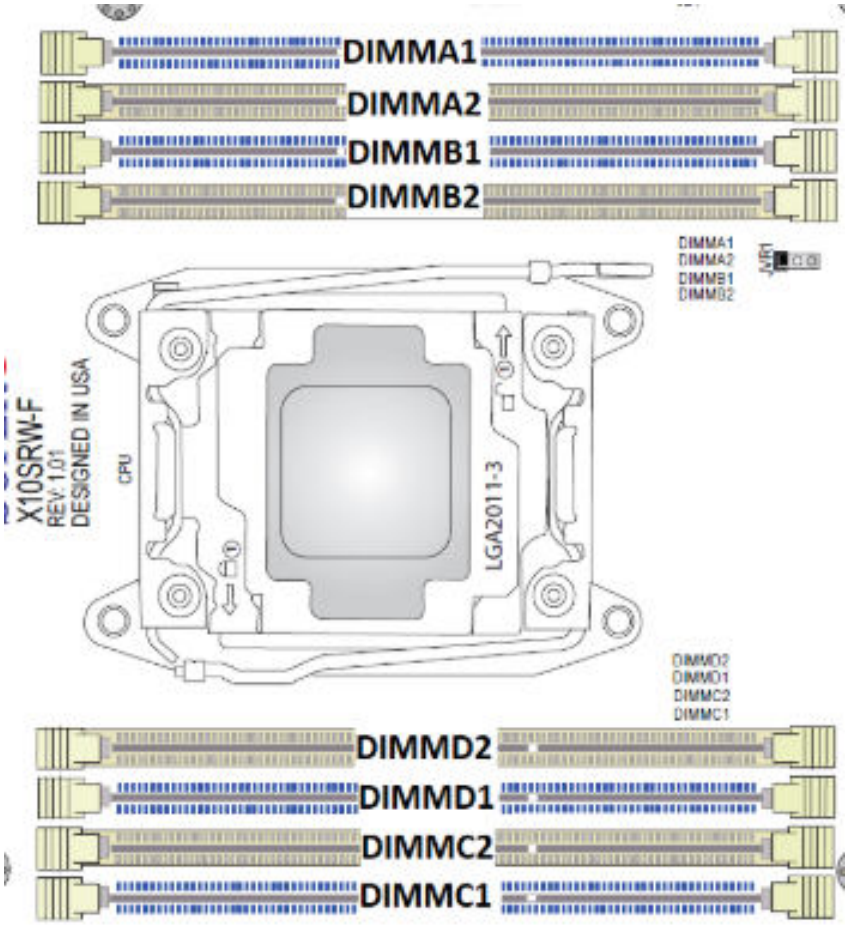


Figure 28: Single-CPU Motherboard with 2DPC Memory Channels

Table 3: X10DRT Single CPU platforms. (2DPC)

Number of CPUs	Number of DIMMs	Slots to use
1	4	A1, B1, C1, D1 (blue slots)
1	6	A1, B1, B2, C1, D1 (blue slots) A2, B2 (black slots)
1	8	Fill all slots.

X10DRU Motherboard

Table 4: Platforms with Three DIMMs Per Channel (3DPC)

NX-1155-G5	NX-3155G-G4/-G5	NX-3175-G4/-G5
NX-6155-G5	NX-8150-G4/-G5	

Three DIMMs per channel (3DPC) memory channels have one blue slot and two black slots each, as shown in the following figure.

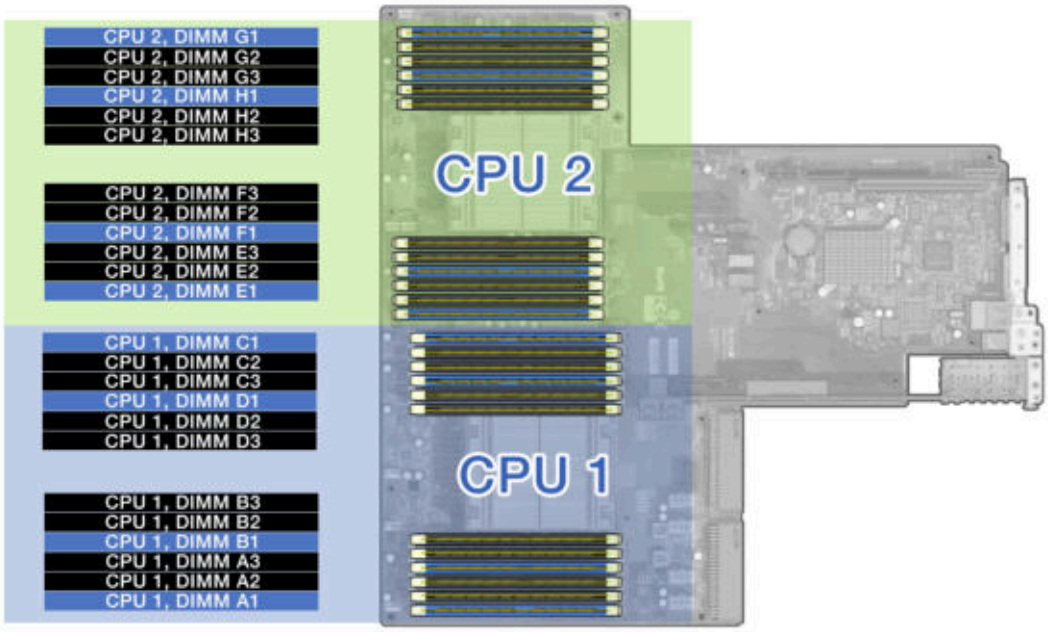


Figure 29: Motherboard with 3DPC Memory Channels

Table 5: Haswell (-G4) X10DRU Dual CPU platforms (3DPC)

Number of CPUs	Number of DIMMs	Slots to use
2	8	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots)
2	12	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots) A2, B2, E2, F2 (black slots)
2	16	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots) A2, B2, C2, D2, E2, F2, G2, H2 (black slots)
2	20	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots) A2, B2, C2, D2, E2, F2, G2, H2 (black slots) A3, B3, E3, F3 (black slots)
2	24	Fill all slots.

Table 6: Broadwell (-G5) X10DRU Dual CPU platforms (3DPC)

Number of CPUs	Number of DIMMs	Slots to use
2	4	A1, B1, E1, F1 (blue slots)
2	6	A1, B1, C1, E1, F1, G1 (blue slots)
2	8	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots)
2	12	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots) A2, B2, E2, F2 (black slots)
2	16	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots) A2, B2, C2, D2, E2, F2, G2, H2 (black slots)
2	20	A1, B1, C1, D1, E1, F1, G1, H1 (blue slots) A2, B2, C2, D2, E2, F2, G2, H2 (black slots) A3, B3, E3, F3 (black slots)
2	24	Fill all slots.

Supported Memory Configurations (G6 and G7 Platforms)

This topic shows DIMM installation order for all Nutanix G6 and G7 platforms. Use the rules and guidelines in this topic to remove, replace, or add memory.

DIMM Restrictions

For DIMM restrictions see the Product Mixing Restrictions topic in this NX Series Hardware Administration Guide.

Memory Installation Order for G6 and G7 Platforms

A memory channel is a group of DIMM slots.

Each CPU is associated with six memory channels. Each memory channel contains two DIMM slots. Memory channels have one blue slot and one black slot each.

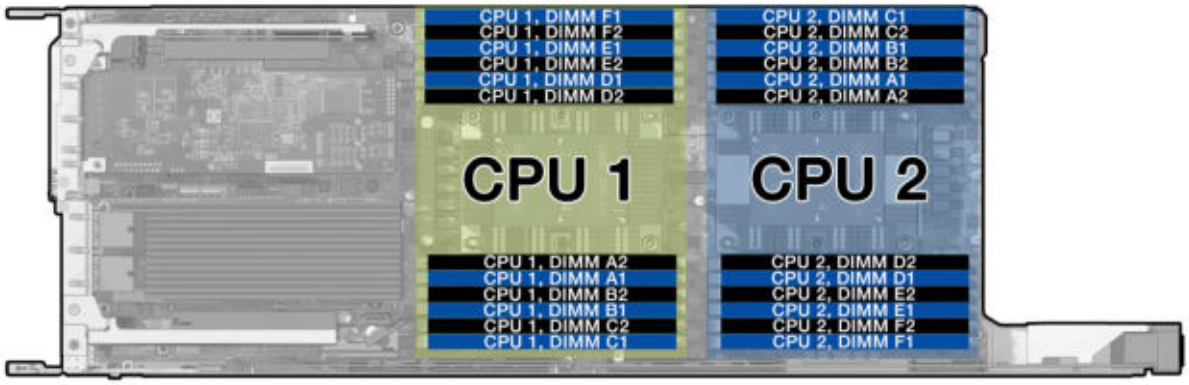


Figure 30: DIMM slots for a G6 and G7 multi-node motherboard

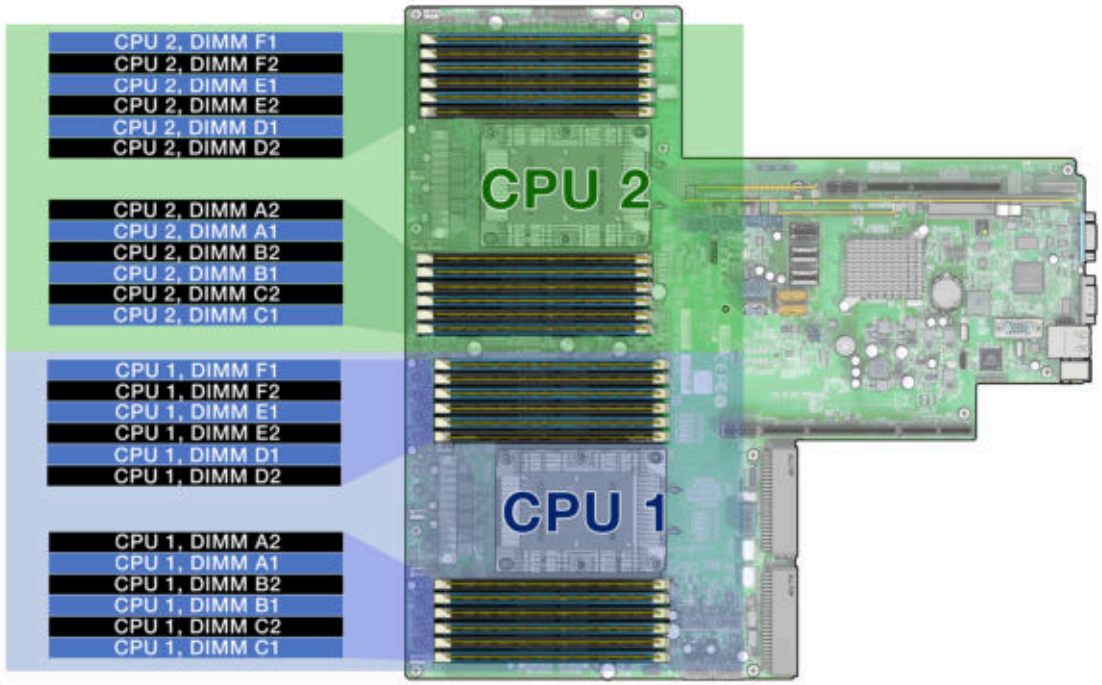


Figure 31: DIMM slots for a G6 and G7 single-node motherboard

Note: DIMM slots on the motherboard are most commonly labeled as A1, A2, and so on. However, some software tools report DIMM slot labels in a different format, such as 1A, 2A, or CPU1, CPU2, or DIMM1, DIMM2.

Table 7: DIMM Installation Order for G6 and G7 Platforms

Number of DIMMs	Slots to use
6	CPU1: A1, B1, C1 (blue slots) CPU2: A1, B1, C1 (blue slots)
8	CPU1: A1, B1, D1, E1 (blue slots) CPU2: A1, B1, D1, E1 (blue slots)
12	CPU1: A1, B1, C1, D1, E1, F1 (blue slots) CPU2: A1, B1, C1, D1, E1, F1 (blue slots)
16	CPU1: A1, B1, D1, E1 (blue slots) CPU1: A2, B2, D2, E2 (black slots) CPU2: A1, B1, D1, E1 (blue slots) CPU2: A2, B2, D2, E2 (black slots)
24	Fill all slots.

Supported Memory Configurations

DIMM installation information for all Nutanix G8 platforms.

DIMM Restrictions

DIMM type

Each G8 node must contain only DIMMs of the same type. So, for example, you cannot mix RDIMM and LRDIMM in the same node.

DIMM capacity

Each G8 node must contain only DIMMs of the same capacity. So, for example, you cannot mix 32 GB DIMMs and 64 GB DIMMs in the same node.

DIMM speed

G8 nodes ship with 3200 MHz DIMMs. 3200 MHz is the highest speed Nutanix currently supports, so you cannot currently mix DIMM speeds in any G8 node.

DIMM manufacturer

You can mix DIMMs from different manufacturers in the same *node*, but not in the same *channel*.

Multi-node platforms

Multi-node G8 platforms contain only one active DIMM slot per channel, so mixing DIMMs in the same channel is not possible.

Single-socket platforms

The single-socket NX-1175S-G8 platform contains only one DIMM slot per channel, so mixing DIMMs in the same channel is not possible.

Single-node platforms

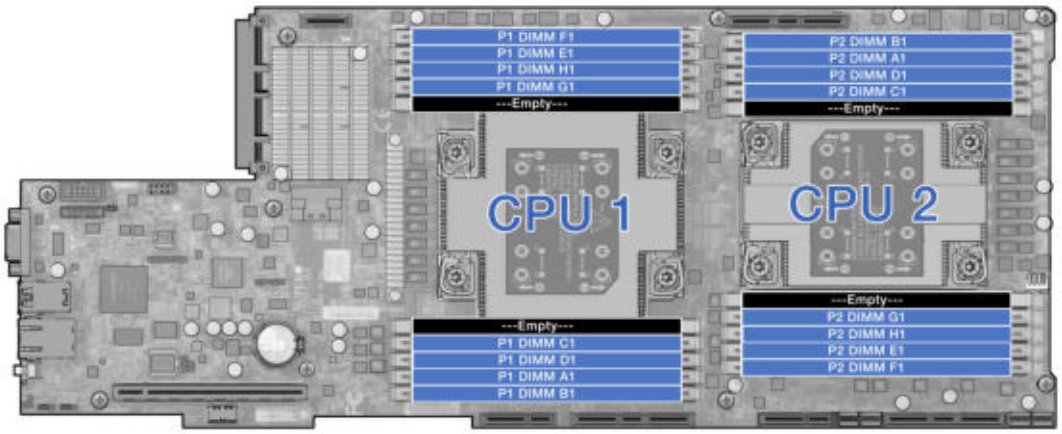
- Single-node G8 platforms contain two DIMM slots per channel. Within a channel, all DIMMs must be from the same manufacturer.
- When replacing a failed DIMM, if there are two DIMMs in the channel, either replace the failed DIMM with a new DIMM form the same manufacturer, or else replace both DIMMs in the channel and make sure that both new DIMMs are from the same manufacturer.
- When adding new DIMMs to a node, if the new DIMMs and the original DIMMs have different manufacturer part numbers, arrange the DIMMs so that the original DIMMs and the new DIMMs are not mixed in the same channel.
  - EXAMPLE: You have an NX-8155-G8 node that has sixteen 32GB DIMMs for a total of 512 GB. You decide to upgrade to thirty-two 32GB DIMMs for a total of 1024 GB. When you remove the node from the chassis and look at the server board, you see that each CPU has eight DIMMs. Remove all DIMMs from one CPU and place them in the empty DIMM slots for the other CPU. Then place all the new DIMMs in the DIMM slots for the first CPU, filling all slots. This way you can ensure that the original DIMMs and the new DIMMs do not share channels.

> Note: You do not need to balance numbers of DIMMs from different manufacturers within a node, so long as you never mix them in the same channel.

Memory Installation Order for Multi-Node G8 Platforms

A memory channel is a group of DIMM slots.

For G8 multi-node platforms, each CPU is associated with eight active memory channels that contain one blue slot each, plus two inactive black slots.



Note: The black slots (C2 and G2 on each CPU) are inactive.

Figure 32: DIMM Slots for a G8 Multinode Serverboard

Table 8: DIMM Installation Order for Multi-node G8 Platforms

Number of DIMMs	Slots to use	Supported capacities (NX-1065-G8)	Supported capacities (NX-3060-G8)
4	CPU1: A1, E1 CPU2: A1, E1	32 GB, 64 GB	32 GB, 64 GB, 128 GB
8	CPU1: A1, C1, E1, G1 CPU2: A1, C1, E1, G1	32 GB, 64 GB	32 GB, 64 GB, 128 GB
12	CPU1: A1, B1, C1, E1, F1, G1 CPU2: A1, B1, C1, E1, F1, G1	32 GB, 64 GB	32 GB, 64 GB, 128 GB
16	Fill all blue slots.	32 GB, 64 GB	32 GB, 64 GB

Memory Installation Order for Single-node G8 Platforms

A memory channel is a group of DIMM slots.

For G8 single-node platforms, each CPU is associated with eight memory channels that contain one blue slot and one black slot each, for a total of 32 DIMM slots.

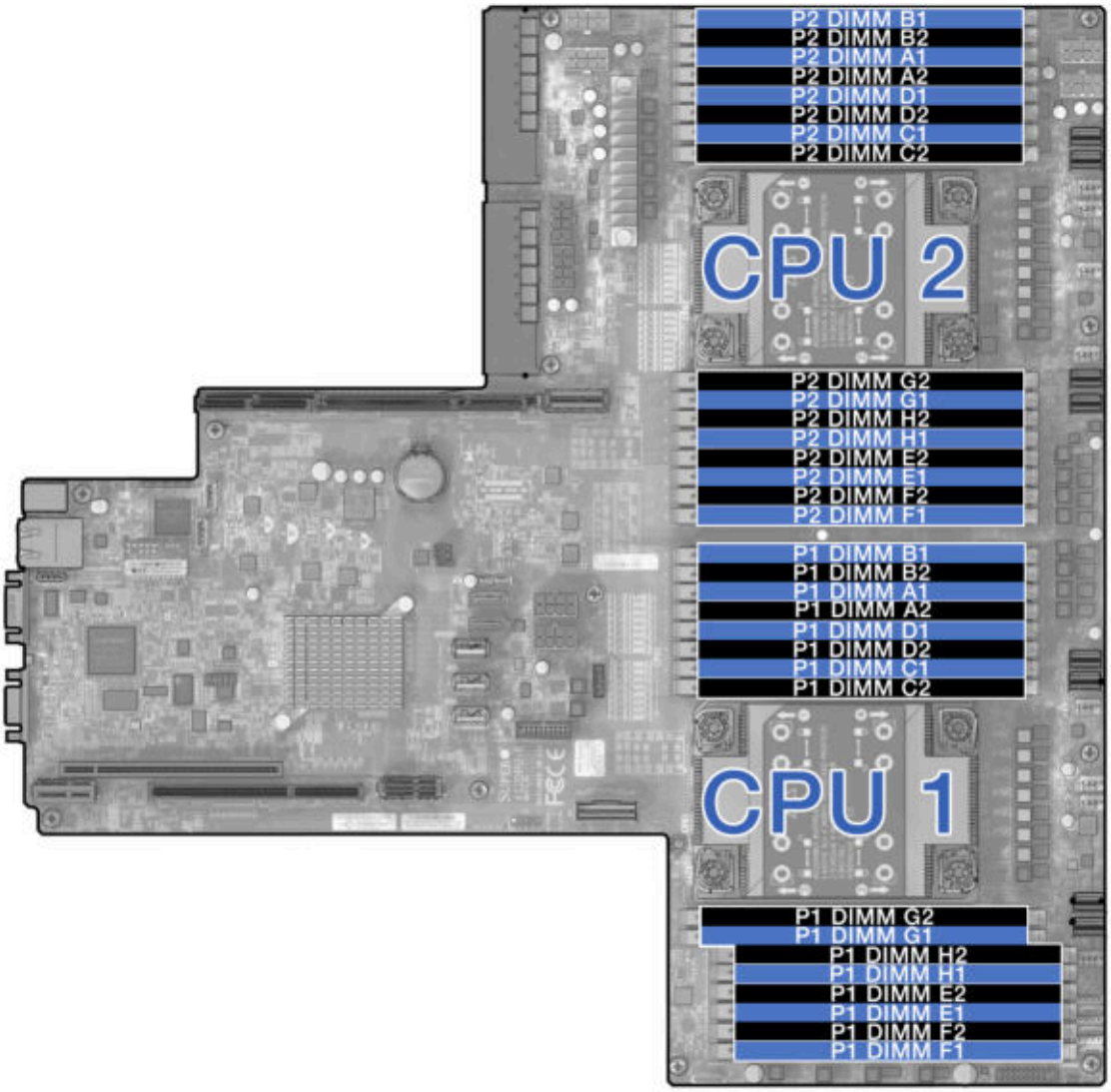


Figure 33: DIMM Slots for a G8 Single-Node Serverboard

Table 9: DIMM Installation Order for Single-Node G8 Platforms

Number of DIMMs	Slots to use	Supported capacities
4	CPU1: A1, E1 (blue slots) CPU2: A1, E1 (blue slots)	32 GB, 64 GB, 128 GB
8	CPU1: A1, C1, E1, G1 (blue slots) CPU2: A1, C1, E1, G1 (blue slots)	32 GB, 64 GB, 128 GB

Number of DIMMs	Slots to use	Supported capacities
12	CPU1: A1, B1, C1, E1, F1, G1 (blue slots) CPU2: A1, B1, C1, E1, F1, G1 (blue slots)	32 GB, 64 GB, 128 GB
16	CPU1: A1, B1, C1, D1, E1, F1, G1, H1 (blue slots) CPU2: A1, B1, C1, D1, E1, F1, G1, H1 (blue slots)	32 GB, 64 GB, 128 GB
24	CPU1: A1, B1, C1, E1, F1, G1 (blue slots) CPU1: A2, B2, C2, E2, F2, G2 (black slots) CPU2: A1, B1, C1, E1, F1, G1 (blue slots) CPU2: A2, B2, C2, E2, F2, G2 (black slots)	32 GB, 64 GB, 128 GB
32	Fill all slots.	32 GB, 64 GB, 128 GB

Memory Installation Order for the Single-socket NX-1175S-G8 Platform

A memory channel is a group of DIMM slots.

On the NX-1175S-G8, the single CPU is associated with eight memory channels that contain one slot each.

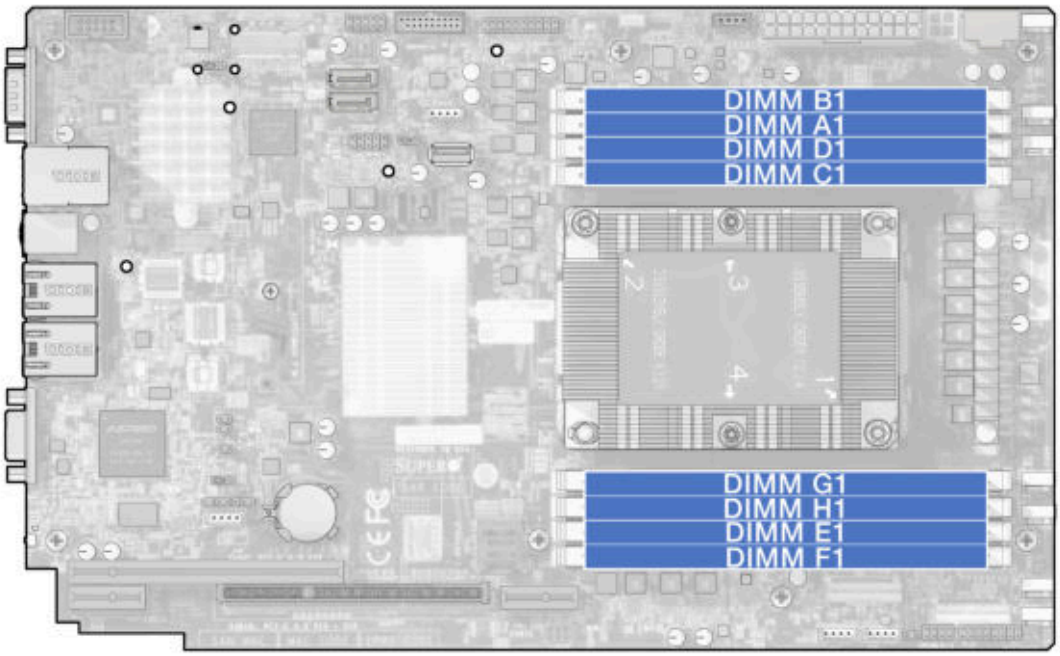


Figure 34: DIMM slots for an NX-1175S-G8 server board

Table 10: DIMM Installation Order for NX-1175S-G8

Number of DIMMs	Slots to use	Supported DIMM capacities
2	A1, E1	32 GB, 64 GB, 128 GB
4	A1, C1, E1, G1	32 GB, 64 GB, 128 GB
8	Fill all slots.	32 GB, 64 GB, 128 GB

9

HOST SECURE BOOT (UEFI)

Host Secure Boot Overview

This topic describes the minimum requirements and host secure boot workflows for the Nutanix NX platforms.

About this task

The Redfish API method can be simpler if you need to enable host secure boot on multiple nodes.

Secure boot support begins with Nutanix G8 platforms.

The following table shows the minimum settings for secure boot.

Components	Minimum version
Foundation	5.0.4
AOS	5.20.1.1
BIOS	W(B/U/W)10.104
BMC	8.0.1
AHV	Bundled AHV
ESXi	ESXI6.7U3b, ESXi7.0U2a
NCC	4.2.0

Summary of the secure boot enablement sequence for NX platforms:

- Confirm the current secure boot status with the Redfish API/ BIOS
- Load all platform keys
- Enable secure boot
- Restart the host
- Verify secure boot status with the Redfish API/ BIOS
- Complete the foundation installation process
- Verify the secure boot status on the CVM/Hypervisor.

The procedures follow this topic in this Hardware Administration Guide.

Enabling Host Secure Boot on Nutanix NX Platforms (Redfish API Method)

Enable host secure boot on Nutanix NX platforms using the Redfish API.

About this task

The Redfish API method can be simpler if you need to enable host secure boot on multiple nodes.

Procedure

1. Verify the secure boot status.

   URL: /redfish/v1/Systems/1/SecureBoot

   Method: GET

   Payload: None

   Response: 200

   Example:

```
nutanix@cvm$ curl --insecure --user <Username>:<Password> \
 "https://<IPMI_IP>/redfish/v1/Systems/1/SecureBoot"

{
 "@odata.type": "#SecureBoot.v1_0_5.SecureBoot",
 "@odata.id": "/redfish/v1/Systems/1/SecureBoot",
 "Id": "SecureBoot",
 "Name": "Security Boot",
 "SecureBootCurrentBoot": "Disabled",
 "SecureBootEnable": false,
 "SecureBootMode": "SetupMode",
       "Actions": {
  "Oem": {},
  "#SecureBoot.ResetKeys": {
   "target":
"/redfish/v1/Systems/1/SecureBoot/Actions/SecureBoot.ResetKeys",
   "@Redfish.ActionInfo":
"/redfish/v1/Systems/1/SecureBoot/ResetKeysActionInfo"
  }
 }
}
```

2. Load all the platform keys.

   URL: /redfish/v1/Systems/1/SecureBoot/Actions/SecureBoot.ResetKeys

   Method: POST

   Payload: {"ResetKeysType": "ResetAllKeysToDefault"}

   Response: 202

   Example:

```
nutanix@cvm$ curl -X POST --insecure --user <Username>:<Password> \
 "https://<IPMI_IP>/redfish/v1/Systems/1/SecureBoot/Actions/SecureBoot. \
ResetKeys" -d "{\"ResetKeysType\":\"ResetAllKeysToDefault\"}"

{
   "Success": {
       "code": "Base.v1_4_0.Success",
       "Message": "Successfully Completed Request."
   }
}
```

3. Enable secure boot.

   URL: /redfish/v1/Systems/1/SecureBoot

   Method: PATCH

   Payload: {"SecureBootEnable": true}

   Response: 202

   Example:

```
nutanix@cvm$ curl -X PATCH --insecure --user <Username>:<Password> \
 "https://<IPMI_IP>/redfish/v1/Systems/1/SecureBoot" -d
 "{\"SecureBootEnable\": true}"
{
 "Success":{
  "code": "Base.v1_4_0.Success",
  "Message": "Successfully Completed Request."
 }
}
```

4. Restart the host.

   > CAUTION: Place the host into maintenance mode and move all VMs to another node before continuing. The node restarts at the end of this step.

   URL: /redfish/v1/Systems/1/Actions/ComputerSystem.Reset

   Method: POST

   Payload: {"ResetType": "ForceRestart"}

   Response: 202

   Example:

```
nutanix@cvm$ curl -X POST --insecure --user <Username>:<Password> \
 "https://<IPMI_IP>/redfish/v1/Systems/1/Actions/ComputerSystem.Reset" \
 -d "{\"ResetType\": \"ForceRestart\"}"
{
   "Success": {
       "code": "Base.v1_4_0.Success",
       "Message": "Successfully Completed Request."
   }
}
```

5. If necessary, refer to the other topics in this *Hardware Administration Guide* for:

   » Secure Boot Error Handling and Failure Conditions

   » Verifying Host Secure Boot Settings

   » Imaging with Foundation

   » Disabling Host Secure Boot

Enabling Host Secure Boot on Nutanix NX Platforms (BIOS Method)

Enable host secure boot on Nutanix NX platforms using BIOS. If you need to enable host secure boot on multiple nodes, the Redfish API method can be simpler.

About this task

This topic describes how to use BIOS to setup host secure boot and verify the host secure boot status.

Procedure

1. Power-on the host: when the boot logo appears, press DELETE to access the BIOS setup utility.
2. Load all the platform keys through the BIOS Setup.
3. Select the **Security Tab** > **Security Boot** > **Key Management** > **Restore Factory Keys**.

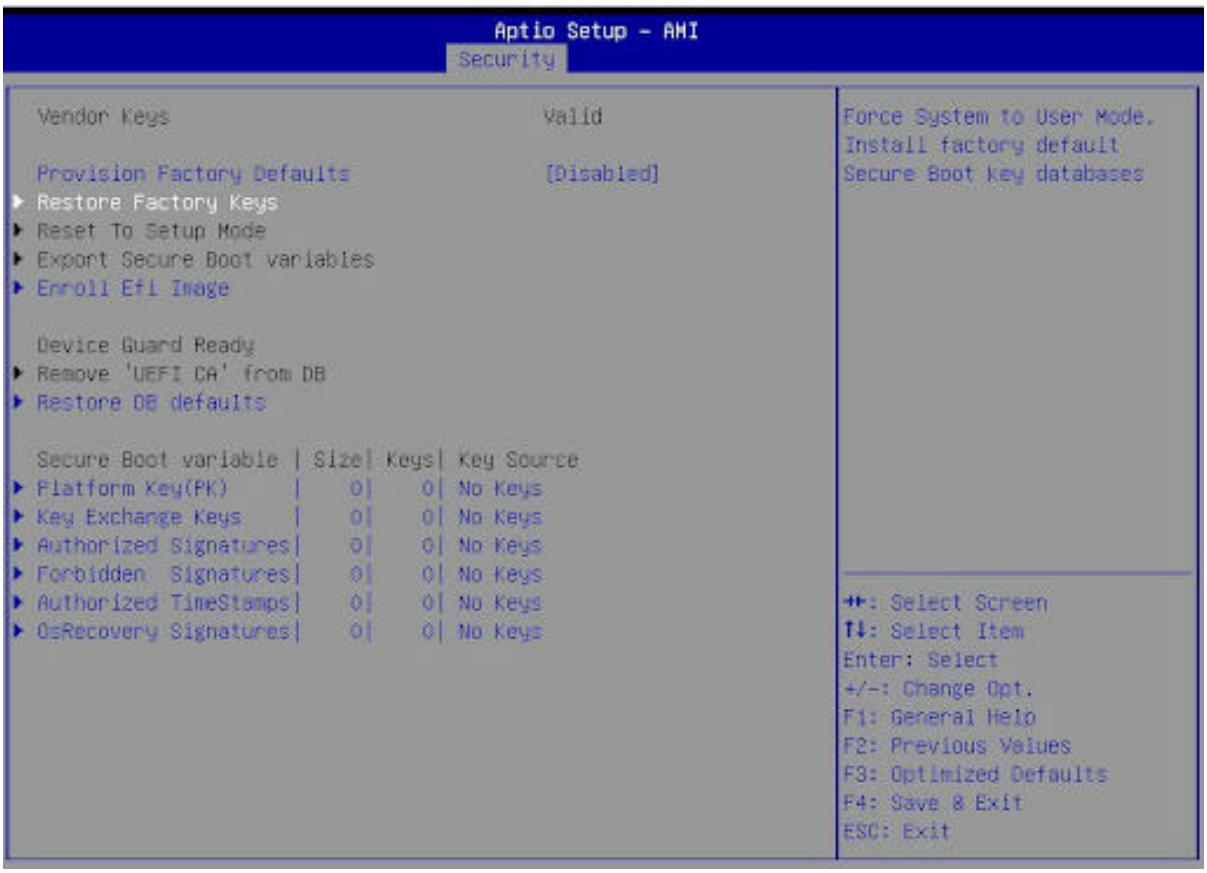


Figure 35: Restore Factory Defaults in BIOS Security

4. Select **Yes**.

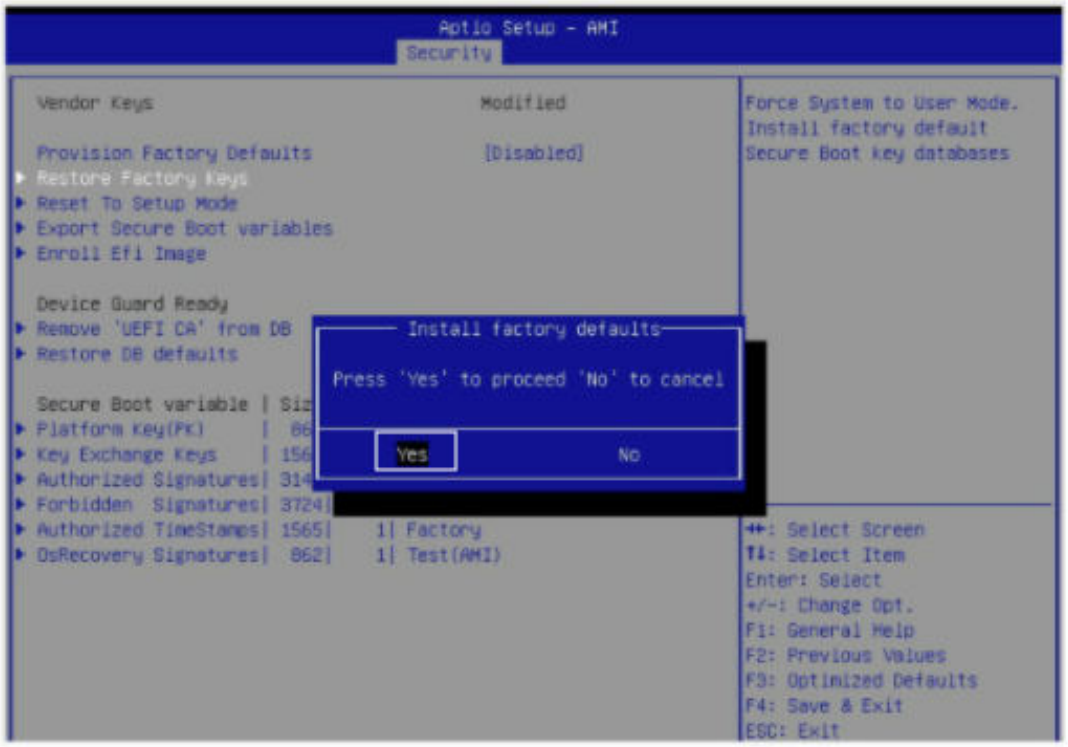


Figure 36: Install Factory Defaults BIOS Menu

The Reset Without Saving prompt is displayed.

5. Select**no** from the Reset without Saving Prompt.

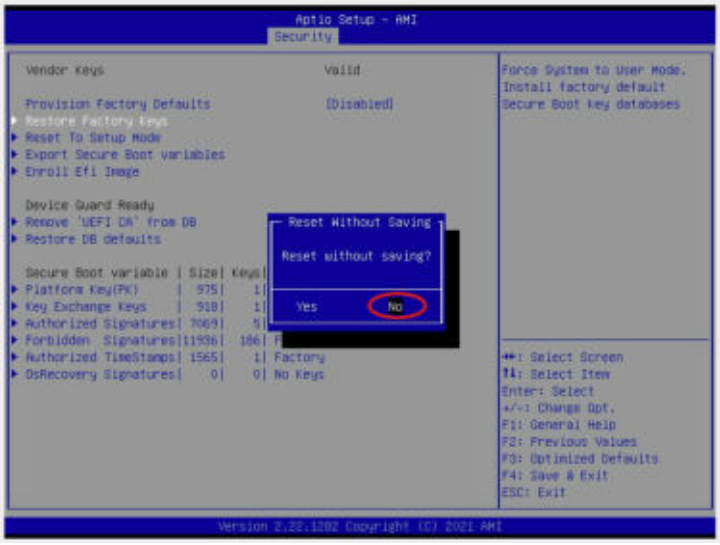


Figure 37: Reset Without Saving Menu

6. From the **Security** tab, select **Security Boot** > **Secure Boot**.

7. Select **Enabled** .

Figure 38: Secure Boot Enabled BIOS Menu

8. Press F4 key to save and exit.
   The host reboots.

9. If necessary, refer to the other topics in this *Hardware Administration Guide* for:

   » Secure Boot Error Handling and Failure Conditions

   » Verifying Host Secure Boot Settings

   » Imaging with Foundation

   » Disabling Host Secure Boot

Secure Boot Error Handling and Failure Conditions

This topic reviews methods for checking on error messages or failures.

About this task

If secure boot is not properly enabled, the node will not boot. The IPMI Remote Control Console will show an error message.

Procedure

1. You can use any of the following alerts to check UEFI secure boot errors or failure conditions.

   » Check the iKVM console for the message: "`Secure Boot violation: Invalid signature detected. Check secure boot policy in setup.`"

Figure 39: iKVM Secure Boot Violation Message

   This message confirms the Node is down due to the Secure Boot violation. The cause could be a mismatch of the secure boot keys between the BIOS and the Hypervisor.

2. If you discover errors, verify if the compatible hypervisor image has been installed correctly.
3. If you want to continue to boot the node with non-secure booting, you can disable the secure boot feature by following the topic *Disabling Host Secure Boot on Nutanix NX Platforms* in this Hardware Administration Guide.
4. Contact Nutanix support if you need further assistance.

Verifying Enabled Host Secure Boot Settings With the Redfish API or BIOS

You can verify the host secure boot settings in the Redfish API or the UEFI console.

About this task

The Redfish API method can be simpler if you need to enable host secure boot on multiple nodes.

Verifying Enabled Host Secure Boot Settings in the Redfish API

Procedure

Verify secure boot status using the Redfish API.

Redfish API method:

URL: /redfish/v1/Systems/1/SecureBoot

Method: GET

Payload: None

Response: 200

Example:

```
nutanix@cvm$ curl --insecure --user <Username>:<Password> \
 "https://<IPMI_IP>/redfish/v1/Systems/1/SecureBoot"
{
   "@odata.type": "#SecureBoot.v1_0_5.SecureBoot",
   "@odata.id": "/redfish/v1/Systems/1/SecureBoot",
   "Id": "SecureBoot",
   "Name": "Security Boot",
   "SecureBootCurrentBoot": "Enabled",
   "SecureBootEnable": true,
   "SecureBootMode": "UserMode",
   "Actions": {
       "Oem": {},
       "#SecureBoot.ResetKeys": {
           "target":
"/redfish/v1/Systems/1/SecureBoot/Actions/SecureBoot.ResetKeys",
           "@Redfish.ActionInfo":
"/redfish/v1/Systems/1/SecureBoot/ResetKeysActionInfo"
       }
   }
}
```

Verifying Enabled Host Secure Boot Settings with BIOS

Procedure

1. Power-on the host: when the boot logo appears, press DELETE to access the BIOS setup utility.
2. Enter the **BIOS Setup** screen.

3. Browse to the **Security** > **Secure Boot** page.

Figure 40: BIOS Security UEFI Security Enabled

System mode shows the following details:

- **System Mode: User**
- **Vendor Keys: Active**
- **Secure Boot: Active**
- **Secure Boot: Enabled**

4. Enter the **Key Management** window and verify that all the factory default keys have been loaded.

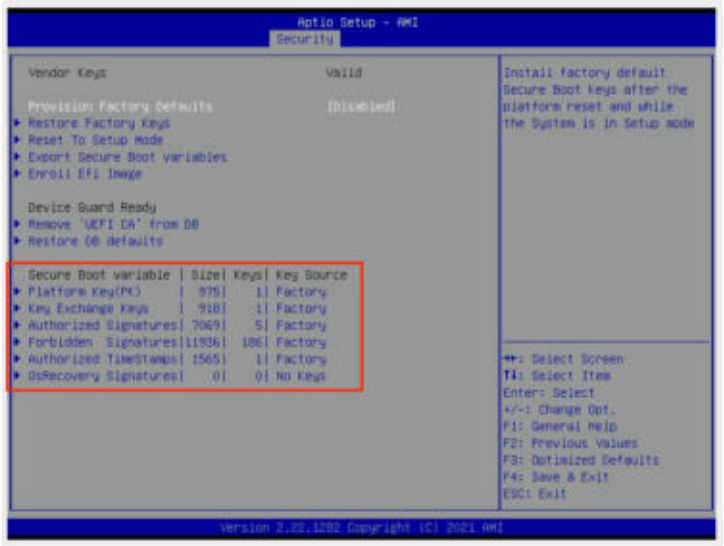


Figure 41: BIOS Security: UEFI Key Management Secure Boot Variables

Imaging with Foundation on a Host Secure Boot Enabled Node

Foundation instructions after enabling host secure boot.

About this task

After the host secure boot is enabled, you can image the node. Ensure that you use the newest *Field Installation Guide* on the Nutanix Support Portal. The following link directs you to version 5.0, but check to see if there is a newer version.

Procedure

Follow the regular Foundation imaging workflow for imaging the host in the Field Installation Guide.

Verifying Enabled Host Secure Boot Settings With the Hypervisor or CVM

After Foundation imaging, you can verify the host secure boot settings from the Hypervisor or CVM.

About this task

Wait until all the cluster services are up before running the following commands.

Procedure

1. For AHV, use `mokutil` from the command line.

```
root@ahv# mokutil --sb-state
SecureBoot enabled
root@ahv#
```

2. For vSphere, use the Python utility.

```
root@esxi# cd usr/lib/vmware/secureboot/bin/
SecureBoot enabled
root@ahv#
```

3. From the Controller VM, use `grep` to determine if secure boot is true.

```
nutanix@cvm$ 10.x.x.x.:~$ zeus_config_printer | grep "is_secure"
is_secure_booted: true
nutanix@cvm:10.x.x.x.:~$
```

Disabling Host Secure Boot on Nutanix NX Platforms

Disable host secure boot on Nutanix NX platforms using the RedFish API.

About this task

Secure boot support begins with Nutanix G8 platforms.

Summary of the secure boot disablement host reboot sequence for NX platforms:

- Verify secure boot status
- Delete all platform keys
- Disable the secure boot
- Restart the host
- Optionally, verify secure boot status from Redfish API or UEFI
- Optionally, confirm the secure boot status from the hypervisor and CVM

- Optionally, verify the secure boot status from the Redfish API or UEFI

Some tools can send API calls that include **curl** commands that you can run on the CVM or on other systems or tools such as Cygwin or Mobaxterm.

Procedure

1. Verify the secure boot status.

   URL: /redfish/v1/Systems/1/SecureBoot

   Method: GET

   Payload: None

   Response: 200

   Example:

```
nutanix@cvm$ curl --insecure --user <Username>:<Password> \
"https://<IPMI_IP>/redfish/v1/Systems/1/SecureBoot"

{
 "@odata.type": "#SecureBoot.v1_0_5.SecureBoot",
 "@odata.id": "/redfish/v1/Systems/1/SecureBoot",
 "Id": "SecureBoot",
 "Name": "Security Boot",
 "SecureBootCurrentBoot": "Enabled",
 "SecureBootEnable": true,
 "SecureBootMode": "UserMode",
       "Actions": {
  "Oem": {},
  "#SecureBoot.ResetKeys": {
   "target":
"/redfish/v1/Systems/1/SecureBoot/Actions/SecureBoot.ResetKeys",
   "@Redfish.ActionInfo":
"/redfish/v1/Systems/1/SecureBoot/ResetKeysActionInfo"
  }
 }
}
```

2. Delete all the platform keys.

   URL: /redfish/v1/Systems/1/SecureBoot/Actions/SecureBoot.ResetKeys

   Method: POST

   Payload: {"ResetKeysType": "DeleteAllKeys"}

   Response: 202

   Example:

```
nutanix@cvm$ curl -X POST --insecure --user <Username>:<Password>
"https://<IPMI_IP>/redfish/v1/Systems/1/SecureBoot/Actions/SecureBoot.
ResetKeys" -d "{\"ResetKeysType\":\"DeleteAllKeys\"}"
{
   "Success": {
       "code": "Base.v1_4_0.Success",
       "Message": "Successfully Completed Request."
   }
}
```

3. Disable secure boot.

   URL: /redfish/v1/Systems/1/SecureBoot

   Method: PATCH

   Payload: {"SecureBootEnable": false}

   Response: 202

   Example:

```
nutanix@cvm$ curl -X PATCH --insecure --user <Username>:<Password> \
"https://<IPMI_IP>/redfish/v1/Systems/1/SecureBoot" -d
"{\"SecureBootEnable\": false}"
{
 "Success":{
  "code": "Base.v1_4_0.Success",
  "Message": "Successfully Completed Request."
 }
}
```

4. Restart the host.

   > CAUTION: Place the host into maintenance mode and move all VMs to another node before continuing. The node restarts at the end of this step.

   URL: /redfish/v1/Systems/1/Actions/ComputerSystem.Reset

   Method: POST

   Payload: {"ResetType": "ForceRestart"}

   Response: 202

   Example:

```
nutanix@cvm$ curl -X POST --insecure --user <Username>:<Password>
 "https://<IPMI_IP>/redfish/v1/Systems/1/Actions/ComputerSystem.Reset"
 -d "{\"ResetType\": \"ForceRestart\"}"
{
   "Success": {
       "code": "Base.v1_4_0.Success",
       "Message": "Successfully Completed Request."
   }
}
```

Verifying the Disabled Host Secure Boot Settings in the Redfish API (Optional)

About this task

Verify the disabled host secure boot settings in the Redfish API or the UEFI console.

Procedure

Verify the secure boot status using the Redfish API or the UEFI console.

Redfish API method:

URL: /redfish/v1/Systems/1/SecureBoot

Method: GET

Payload: None

Response: 200

Example:

```
nutanix@cvm$ curl --insecure --user <Username>:<Password>
"https://<IPMI_IP>/redfish/v1/Systems/1/SecureBoot"
{
   "@odata.type": "#SecureBoot.v1_0_5.SecureBoot",
   "@odata.id": "/redfish/v1/Systems/1/SecureBoot",
   "Id": "SecureBoot",
   "Name": "Security Boot",
   "SecureBootCurrentBoot": "Disabled",
   "SecureBootEnable": false,
   "SecureBootMode": "SetupMode",
   "Actions": {
       "Oem": {},
       "#SecureBoot.ResetKeys": {
           "target":
"/redfish/v1/Systems/1/SecureBoot/Actions/SecureBoot.ResetKeys",
           "@Redfish.ActionInfo":
"/redfish/v1/Systems/1/SecureBoot/ResetKeysActionInfo"
       }
   }
}
```

Verifying the Disabled Host Secure Boot Settings with UEFI (Optional)

About this task

Verify the disabled host secure boot settings in the UEFI console or the Redfish API.

Procedure

1. Restart the host.
2. Enter the BIOS/UEFI setup screen.

3. Browse to the **Security** > **Secure Boot** page.

Figure 42: Disabled: BIOS/UEFI Security

System mode shows the following details:

- **System Mode: User**
- **Vendor Keys: Active**
- **Secure Boot: Active**
- **Secure Boot: Enabled**

4. Enter the **Key Management** window and verify that all the factory default keys are loaded.

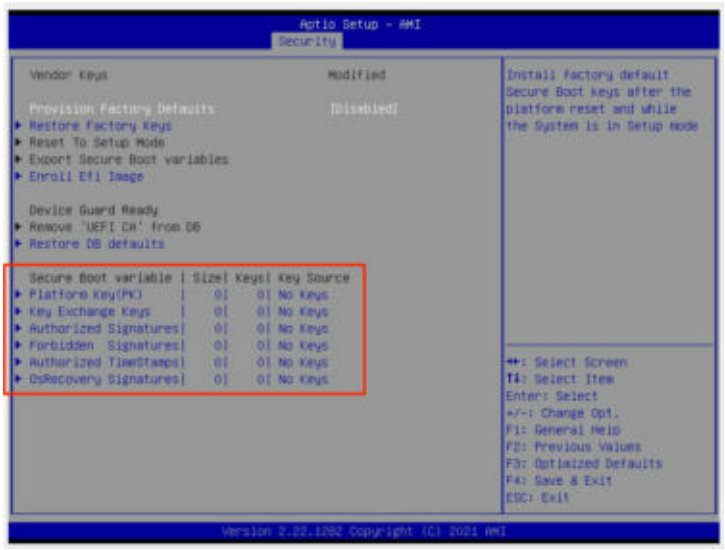


Figure 43: Disabled BIOS Security: UEFI Key Management Secure Boot Variables

Confirming the Disabled Secure Boot Status from the Hypervisor and CVM

About this task

Wait until all the cluster services are up before running the following commands.

Procedure

1. For AHV, use `mokutil` from the command line,

```
root@ahv# mokutil --sb-state
SecureBoot disabled
root@ahv
```

2. For vSphere, use the Python utility.

```
root@esxi# cd usr/lib/vmware/secureboot/bin/
root@esxi#/usr/lib/vmware/secureboot/bin] python secureBoot.py -s
disabled
root@esxi#/usr/lib/vmware/secureboot/bin
```

3. From the Controller VM, use `grep` to determine if secure boot is true.

```
nutanix@cvm$ 10.x.x.x.:~$ zeus_config_printer | grep "is_secure"
is_secure_booted: false
nutanix@cvm:10.x.x.x.$
```

COPYRIGHT

Copyright 2022 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

License

The provision of this software to you does not grant any licenses or other rights under any Microsoft patents with respect to anything other than the file server implementation portion of the binaries for this software, including no licenses or any other rights in any hardware or any devices or software that are used to communicate with or in connection with this software.

Conventions

Convention	Description
`variable_value`	The action depends on a value that is unique to your environment.
`ncli> command`	The commands are executed in the Nutanix nCLI.
`user@host$ command`	The commands are executed as a non-privileged user (such as nutanix) in the system shell.
`root@host# command`	The commands are executed as the root user in the vSphere or Acropolis host shell.
`> command`	The commands are executed in the Hyper-V host shell.
`output`	The information is displayed as output from a command or in a log file.

Default Cluster Credentials

Interface	Target	Username	Password
Nutanix web console	Nutanix Controller VM	admin	Nutanix/4u
vSphere Web Client	ESXi host	root	nutanix/4u
vSphere Client	ESXi host	root	nutanix/4u
SSH client or console	ESXi host	root	nutanix/4u
SSH client or console	AHV host	root	nutanix/4u

Interface	Target	Username	Password
SSH client or console	Hyper-V host	Administrator	nutanix/4u
SSH client	Nutanix Controller VM	nutanix	nutanix/4u
SSH client	Nutanix Controller VM	admin	Nutanix/4u
IPMI web interface or ipmitool	Nutanix node	ADMIN	ADMIN
SSH client or console	Acropolis OpenStack Services VM (Nutanix OVM)	root	admin
SSH client or console	Xtract VM	nutanix	nutanix/4u
SSH client or console	Xplorer VM	nutanix	nutanix/4u

Version

Last modified: March 3, 2022 (2022-03-03T09:56:00+05:30)

Intel® Xeon® Silver 4310T Processor

15M Cache, 2.30 GHz

Add to Compare

Specifications
- Essentials
- CPU Specifications
- Supplemental Information
- Memory Specifications
- Expansion Options
- Package Specifications
- Advanced Technologies
- Security & Reliability

Ordering and Compliance

Essentials

Export specifications

Product Collection	3rd Generation Intel® Xeon® Scalable Processors
Code Name	Products formerly Ice Lake
Vertical Segment	Server
Processor Number	4310T
Status	Launched
Launch Date	Q2'21
Lithography	10 nm
Recommended Customer Price	$594.00

CPU Specifications

Total Cores	10
Total Threads	20
Max Turbo Frequency	3.40 GHz
Processor Base Frequency	2.30 GHz
Cache	15 MB
Intel® UPI Speed	10.4 GT/s
# of UPI Links	2
TDP	105 W

Supplemental Information

Embedded Options Available	Yes

Memory Specifications

Max Memory Size (dependent on memory type)	6 TB
Memory Types	DDR4-2667
Maximum Memory Speed	2667 MHz
Max # of Memory Channels	8
ECC Memory Supported ‡	Yes
Intel® Optane™ Persistent Memory Supported	No

Expansion Options

Scalability	2S
PCI Express Revision	4.0
Max # of PCI Express Lanes	64

Package Specifications

Sockets Supported	FCLGA4189
T_{CASE}	88°C
Package Size	77.5mm x 56.5mm

Advanced Technologies

Intel® Deep Learning Boost (Intel® DL Boost)	Yes
Intel® Resource Director Technology (Intel® RDT)	Yes
Intel® Speed Shift Technology	Yes
Intel® Turbo Boost Technology ‡	2.0
Intel® Hyper-Threading Technology ‡	Yes
Intel® Virtualization Technology (VT-x) ‡	Yes
Intel® Virtualization Technology for Directed I/O (VT-d) ‡	Yes
Intel® VT-x with Extended Page Tables (EPT) ‡	Yes
Intel® Transactional Synchronization Extensions	Yes
Intel® 64 ‡	Yes
Instruction Set Extensions	Intel® SSE4.2, Intel® AVX, Intel® AVX2, Intel® AVX-512
# of AVX-512 FMA Units	2
Intel® Volume Management Device (VMD)	Yes

Security & Reliability

Intel® AES New Instructions	Yes
Intel® Software Guard Extensions (Intel® SGX)	Yes with Intel® SPS
Intel® Trusted Execution Technology ‡	Yes
Execute Disable Bit ‡	Yes
Mode-based Execute Control (MBEC)	Yes
Intel® Total Memory Encryption	Yes
Intel® Platform Firmware Resilience Support	Yes
Maximum Enclave Page Cache (EPC) Size for Intel® SGX	8 GB
Intel® Crypto Acceleration	Yes

More support options for Intel® Xeon® Silver 4310T Processor (15M Cache, 2.30 GHz)

Product Support | Downloads and Software | Support Community | Warranty and Replacement

Need more help?

Contact support

Give Feedback

All information provided is subject to change at any time, without notice. Intel may make changes to manufacturing life cycle, specifications, and product descriptions at any time, without notice. The information herein is provided "as-is" and Intel does not make any representations or warranties whatsoever regarding accuracy of the information, nor on the product features, availability, functionality, or compatibility of the products listed. Please contact system vendor for more information on specific products or systems.

Intel classifications are for informational purposes only and consist of Export Control Classification Numbers (ECCN) and Harmonized Tariff Schedule (HTS) numbers. Any use made of Intel classifications are without recourse to Intel and shall not be construed as a representation or warranty regarding the proper ECCN or HTS. Your company as an importer and/or exporter is responsible for determining the correct classification of your transaction.

Refer to Datasheet for formal definitions of product properties and features.

‡ This feature may not be available on all computing systems. Please check with the system vendor to determine if your system delivers this feature, or reference the system specifications (motherboard, processor, chipset, power supply, HDD, graphics controller, memory, BIOS, drivers, virtual machine monitor-VMM, platform software, and/or operating system) for feature compatibility. Functionality, performance, and other benefits of this feature may vary depending on system configuration.

Some products can support AES New Instructions with a Processor Configuration update, in particular, i7-2630QM/i7-2635QM, i7-2670QM/i7-2675QM, i5-2430M/i5-2435M, i5-2410M/i5-2415M. Please contact OEM for the BIOS that includes the latest Processor configuration update.

"Announced" SKUs are not yet available. Please refer to the Launch Date for market availability.

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. See http://www.intel.com/content/www/us/en/processors/processor-numbers.html for details.

See http://www.intel.com/content/www/us/en/architecture-and-technology/hyper-threading/hyper-threading-technology.html?wapkw=hyper+threading for more information including details on which processors support Intel® HT Technology.

Max Turbo Frequency refers to the maximum single-core processor frequency that can be achieved with Intel® Turbo Boost Technology. See www.intel.com/technology/turboboost/ for more information and applicability of this technology.

Processors that support 64-bit computing on Intel® architecture require an Intel 64 architecture-enabled BIOS.

System and Maximum TDP is based on worst case scenarios. Actual TDP may be lower if not all I/Os for chipsets are used.

Intel technologies may require enabled hardware, software or service activation. // No product or component can be absolutely secure. // Your costs and results may vary. // Performance varies by use, configuration and other factors. // See our complete legal Notices and Disclaimers. // Intel is committed to respecting human rights and avoiding complicity in human rights abuses. See Intel's Global Human Rights Principles. Intel's products and software are intended only to be used in applications that do not cause or contribute to a violation of an internationally recognized human right.

Intel® Xeon® Gold 6354 Processor

39M Cache, 3.00 GHz

Add to Compare

Specifications
- Essentials
- CPU Specifications
- Supplemental Information
- Memory Specifications
- Expansion Options
- Package Specifications
- Advanced Technologies
- Security & Reliability

Ordering and Compliance

Export specifications

Essentials

Product Collection	3rd Generation Intel® Xeon® Scalable Processors
Code Name	Products formerly Ice Lake
Vertical Segment	Server
Processor Number	6354
Status	Launched
Launch Date	Q2'21
Lithography	10 nm
Recommended Customer Price	$2616.00

CPU Specifications

Total Cores	18
Total Threads	36
Max Turbo Frequency	3.60 GHz
Processor Base Frequency	3.00 GHz
Cache	39 MB
Intel® UPI Speed	11.2 GT/s
# of UPI Links	3
TDP	205 W

Supplemental Information

Embedded Options Available	No

Memory Specifications

Max Memory Size (dependent on memory type)	6 TB
Memory Types	DDR4-3200
Maximum Memory Speed	3200 MHz
Max # of Memory Channels	8
ECC Memory Supported ‡	Yes
Intel® Optane™ Persistent Memory Supported	Yes

Expansion Options

Scalability	2S
PCI Express Revision	4.0
Max # of PCI Express Lanes	64

Package Specifications

Sockets Supported	FCLGA4189
T_{CASE}	77°C
Package Size	77.5mm x 56.5mm

Advanced Technologies

Intel® Speed Select Technology – Core Power	Yes
Intel® Speed Select Technology – Turbo Frequency	Yes
Intel® Deep Learning Boost (Intel® DL Boost)	Yes
Intel® Speed Select Technology - Base Frequency	Yes
Intel® Resource Director Technology (Intel® RDT)	Yes
Intel® Speed Shift Technology	Yes
Intel® Turbo Boost Technology ‡	2.0
Intel® Hyper-Threading Technology ‡	Yes
Intel® Virtualization Technology (VT-x) ‡	Yes
Intel® Virtualization Technology for Directed I/O (VT-d) ‡	Yes
Intel® VT-x with Extended Page Tables (EPT) ‡	Yes
Intel® Transactional Synchronization Extensions	Yes
Intel® 64 ‡	Yes
Instruction Set Extensions	Intel® SSE4.2, Intel® AVX, Intel® AVX2, Intel® AVX-512
# of AVX-512 FMA Units	2
Intel® Volume Management Device (VMD)	Yes

Security & Reliability

Intel® AES New Instructions	Yes
Intel® Software Guard Extensions (Intel® SGX)	Yes with Intel® SPS
Intel® Trusted Execution Technology ‡	Yes
Execute Disable Bit ‡	Yes
Intel® Run Sure Technology	Yes
Mode-based Execute Control (MBEC)	Yes
Intel® Total Memory Encryption	Yes
Intel® Platform Firmware Resilience Support	Yes
Maximum Enclave Page Cache (EPC) Size for Intel® SGX	64 GB
Intel® Crypto Acceleration	Yes

More support options for Intel® Xeon® Gold 6354 Processor (39M Cache, 3.00 GHz)

Product Support | Downloads and Software | Support Community | Warranty and Replacement

Need more help?

Contact support

Give Feedback

All information provided is subject to change at any time, without notice. Intel may make changes to manufacturing life cycle, specifications, and product descriptions at any time, without notice. The information herein is provided "as-is" and Intel does not make any representations or warranties whatsoever regarding accuracy of the information, nor on the product features, availability, functionality, or compatibility of the products listed. Please contact system vendor for more information on specific products or systems.

Intel classifications are for informational purposes only and consist of Export Control Classification Numbers (ECCN) and Harmonized Tariff Schedule (HTS) numbers. Any use made of Intel classifications are without recourse to Intel and shall not be construed as a representation or warranty regarding the proper ECCN or HTS. Your company as an importer and/or exporter is responsible for determining the correct classification of your transaction.

Refer to Datasheet for formal definitions of product properties and features.

‡ This feature may not be available on all computing systems. Please check with the system vendor to determine if your system delivers this feature, or reference the system specifications (motherboard, processor, chipset, power supply, HDD, graphics controller, memory, BIOS, drivers, virtual machine monitor-VMM, platform software, and/or operating system) for feature compatibility. Functionality, performance, and other benefits of this feature may vary depending on system configuration.

Some products can support AES New Instructions with a Processor Configuration update, in particular, i7-2630QM/i7-2635QM, i7-2670QM/i7-2675QM, i5-2430M/i5-2435M, i5-2410M/i5-2415M. Please contact OEM for the BIOS that includes the latest Processor configuration update.

"Announced" SKUs are not yet available. Please refer to the Launch Date for market availability.

Intel processor numbers are not a measure of performance. Processor numbers differentiate features within each processor family, not across different processor families. See http://www.intel.com/content/www/us/en/processors/processor-numbers.html for details.

See http://www.intel.com/content/www/us/en/architecture-and-technology/hyper-threading/hyper-threading-technology.html?wapkw=hyper+threading for more information including details on which processors support Intel® HT Technology.

Max Turbo Frequency refers to the maximum single-core processor frequency that can be achieved with Intel® Turbo Boost Technology. See www.intel.com/technology/turboboost/ for more information and applicability of this technology.

Processors that support 64-bit computing on Intel® architecture require an Intel 64 architecture-enabled BIOS.

System and Maximum TDP is based on worst case scenarios. Actual TDP may be lower if not all I/Os for chipsets are used.

Intel technologies may require enabled hardware, software or service activation. // No product or component can be absolutely secure. // Your costs and results may vary. // Performance varies by use, configuration and other factors. // See our complete legal Notices and Disclaimers. // Intel is committed to respecting human rights and avoiding complicity in human rights abuses. See Intel's Global Human Rights Principles. Intel's products and software are intended only to be used in applications that do not cause or contribute to a violation of an internationally recognized human right.

Flash CBL licenses in TiB, it ensures that a single unit of measure is utilized from quote all the way through to license provisioning and compliance checks. Imagine the confusion if a customer ordered 100 TB of flash licenses but we delivered only 91 TiB licenses to the portal, or if a customer provisioned 100 TB of flash licenses from the portal but only 91 TiB of licensing shows up in Prism.

Q: Is mixing server hardware brands within a cluster allowed?
A: No. Mixing server hardware brands within a cluster violates our support policies.

Q: Will license type and relevant features be enforced?
A: License noncompliance triggers nagware. Alerts will show up if cluster capacity exceeds licensed capacity. Alerts will also show up if licenses expire or if there is a feature violation.

Q: What happens when a Nutanix software license expires at the end of the term?
A: When a Nutanix software license expires, the customer's legal entitlement to use the software and software support also expires. To avoid this, the customer is able to renew their software license in a similar manner to how they renew support today. Once procured, the renewed license entitlements will be reflected in the customer's account on the Nutanix Support Portal and the licenses should be refreshed on the cluster. Renewed licenses may be co-terminated with hardware support, other software subscriptions, or a target date.

If a customer doesn't renew their licenses the following actions occur:

- The customer will see license non-compliance messaging in their Prism interface
- The customer no longer has legal entitlement to use the software
- The customer no longer has legal entitlement for support for the non-compliant software
- The cluster continues to operate with full performance, functionality, data access, and management.

Q: Can a SWO node license be traded in for CBL licenses?
A: Systemic trade-in of SWO node licenses for CBL licenses is not currently supported. If you have a specific customer situation, please reach out with opportunity details on the slack channel.

Q: How does a customer upgrade their Acropolis software edition (e.g., Starter → Pro)?
A: The customer would purchase an upgrade SKU. Once this upgrade transaction is complete, license entitlements will be reflected in the customer's account on the Nutanix Support Portal. Customer would initiate a licensing action to upgrade.

Q: Starter is now available for CBL. What does this mean?
A:

- Starter license level is available (unlike node based license for third party platform). This will lower price point to enter. STR licenses will be deposited in the Nutanix Support Portal when purchased.
- To be compliant running STR license, licensing actions similar to PRO and ULT will be required.
- Single license level for a cluster will continue to apply
- No change to existing SWO node licenses. STR will not be available for SWO node licenses.

Leap 6.1

Leap Administration Guide

March 17, 2022

Contents

1. Leap Overview....4
   - Leap Terminology....8
   - Nutanix Disaster Recovery Solutions....11
   - Leap Deployment Workflow....12
   - On-Prem Hardware Resource Requirements....15

2. Protection and DR between On-Prem Sites (Leap)....18
   - Leap Requirements....20
   - Leap Limitations....24
   - Leap Recommendations....25
   - Leap Service-Level Agreements (SLAs)....26
   - Leap Views....26
     - Availability Zones View....26
     - Protection Summary View....28
     - Protection Policies View....31
     - Recovery Plans View....33
     - VM Recovery Points....34
     - Dashboard Widgets....36
   - Enabling Leap for On-Prem Site....36
   - Pairing Availability Zones (Leap)....37
   - Protection and Automated DR (Leap)....40
     - Protection with Asynchronous Replication Schedule and DR (Leap)....40
     - Protection with NearSync Replication Schedule and DR (Leap)....100
     - Protection with Synchronous Replication Schedule (0 RPO) and DR....116
     - Protection Policy Management....135
     - Recovery Plan Management....141
   - Protection and Manual DR (Leap)....143
     - Creating Recovery Points Manually (Out-of-Band Snapshots)....144
     - Replicating Recovery Points Manually....144
     - Manual Recovery of Guest VMs....144
   - Entity Synchronization Between Paired Availability Zones....151
     - Entity Synchronization Recommendations (Leap)....152
     - Forcing Entity Synchronization (Leap)....152

3. Protection and DR between On-Prem Site and Xi Cloud Service (Xi Leap)....154
   - Xi Leap Requirements....156
   - Xi Leap Limitations....159
   - Xi Leap Recommendations....160
   - Xi Leap Service-Level Agreements (SLAs)....161
   - Xi Leap Views....161
     - Availability Zones View in Xi Cloud Services....162
     - Protection Policies View in Xi Cloud Services....163
     - Recovery Plans View in Xi Cloud Services....164
     - Dashboard Widgets in Xi Cloud Services....165
   - Enabling Leap for On-Prem Site....166
   - Xi Leap Environment Setup....166

Pairing Availability Zones (Xi Leap)....167
VPN Configuration (On-prem and Xi Cloud Services)....167
Nutanix Virtual Networks....191
Xi Leap RPO Sizer....193
Protection and Automated DR (Xi Leap)....196
Protection with Asynchronous Replication and DR (Xi Leap)....196
Protection with NearSync Replication and DR (Xi Leap)....228
Protection Policy Management....237
Recovery Plan Management....241
Protection and Manual DR (Xi Leap)....244
Creating Recovery Points Manually (Out-of-Band Snapshots)....244
Replicating Recovery Points Manually....244
Recovering a Guest VM from a Recovery Point Manually....245
Entity Synchronization Between Paired Availability Zones....245
Entity Synchronization Recommendations (Xi Leap)....246
Forcing Entity Synchronization (Xi Leap)....247

4. Migrating Guest VMs from a Protection Domain to a Protection Policy....248

Copyright....249
License....249
Conventions....249
Version....249

1

LEAP OVERVIEW

Legacy disaster recovery (DR) configurations use protection domains (PDs) and third-party integrations to protect your applications. These DR configurations replicate data between on-prem Nutanix clusters. Protection domains provide limited flexibility in terms of supporting complex operations (for example, VM boot order, network mapping). With protection domains, you have to perform manual tasks to protect new guest VMs as and when your application scales up.

Leap offers an entity-centric automated approach to protect and recover applications. It uses categories to group the guest VMs and automate the protection of the guest VMs as the application scales. Application recovery is more flexible with network mappings, an enforceable VM start sequence, and inter-stage delays. Application recovery can also be validated and tested without affecting your production workloads. Asynchronous, NearSync, and Synchronous replication schedules ensure that an application and its configuration details synchronize to one or more recovery locations for a smoother recovery.

> Note: You can protect a guest VM either with legacy DR solution (protection domain-based) or with new Leap. To see various Nutanix DR solutions, refer Nutanix Disaster Recovery Solutions on page 11.

Leap works with sets of physically isolated locations called availability zones. An instance of Prism Central represents an availability zone. One availability zone serves as the primary site for an application while one or more paired availability zones serve as the recovery sites.

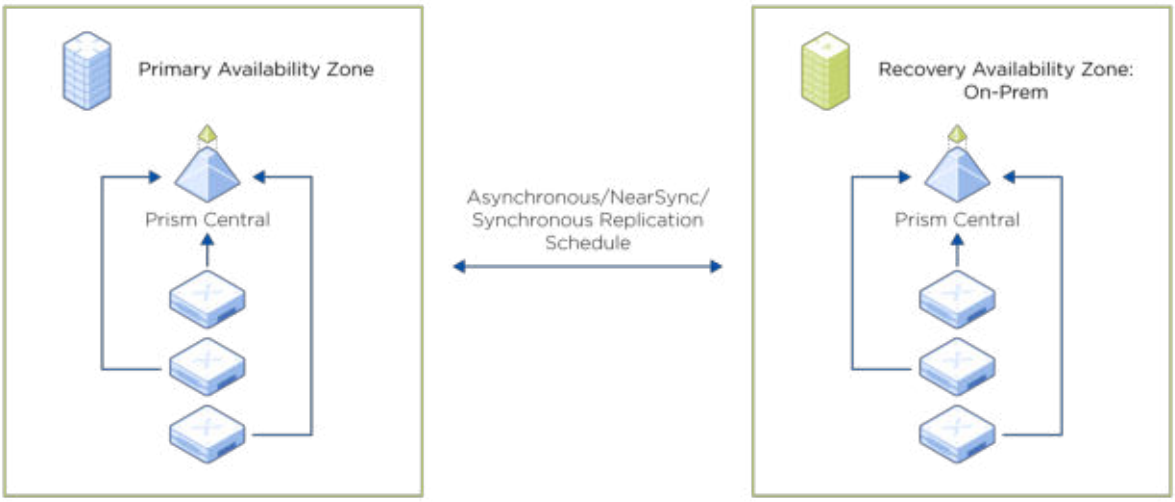


Figure 1: A primary on-prem site and one recovery on-prem site

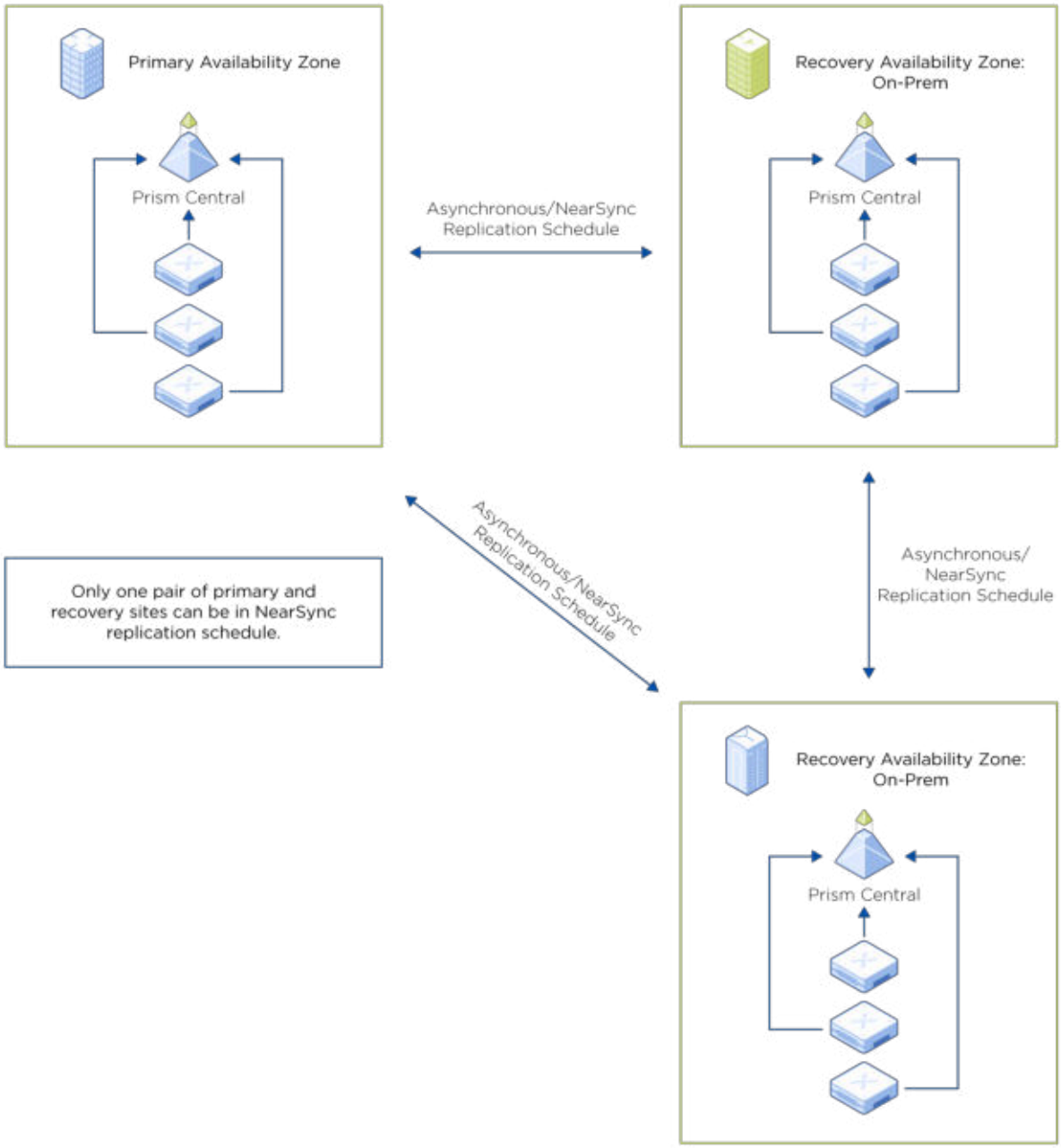


Figure 2: A primary on-prem site and two recovery on-prem sites

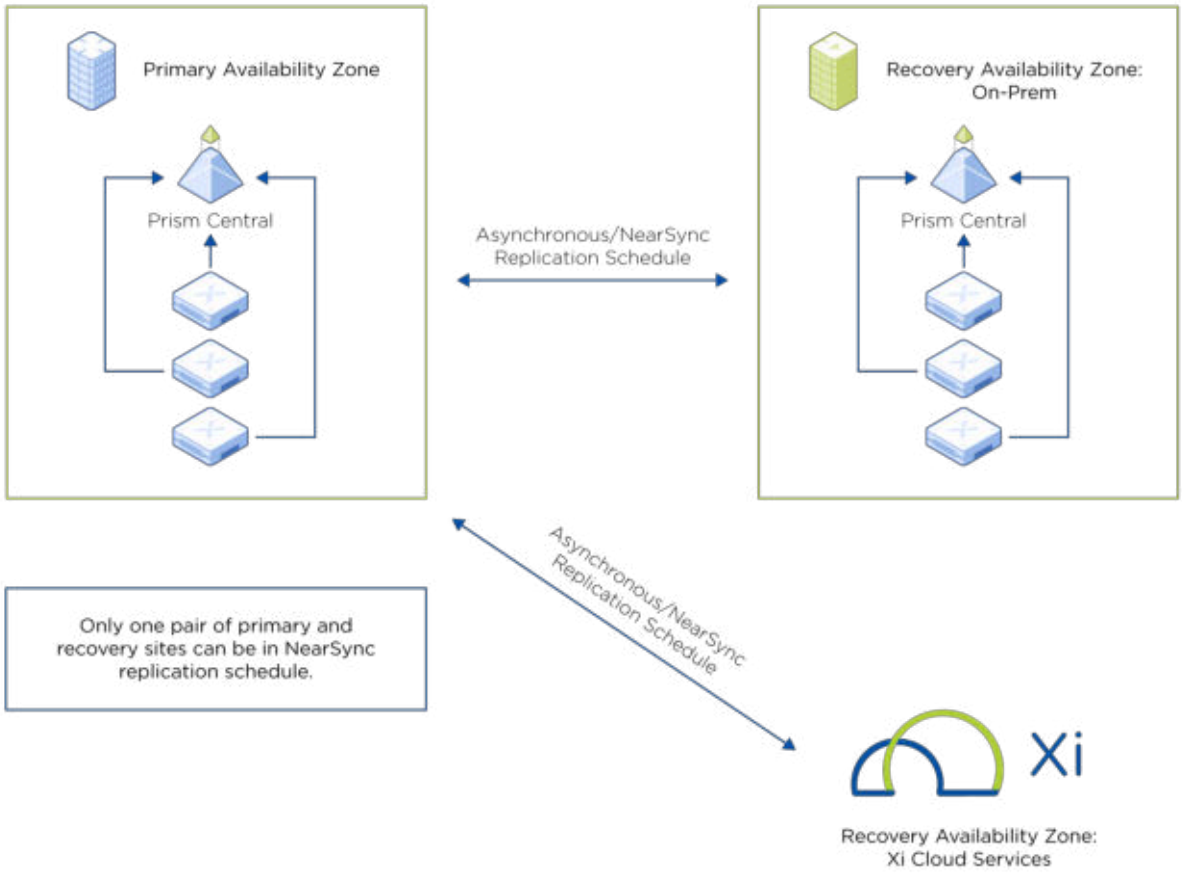


Figure 3: A primary on-prem site and two recovery sites: one on-prem recovery site and one recovery site in Cloud (Xi Cloud Services)

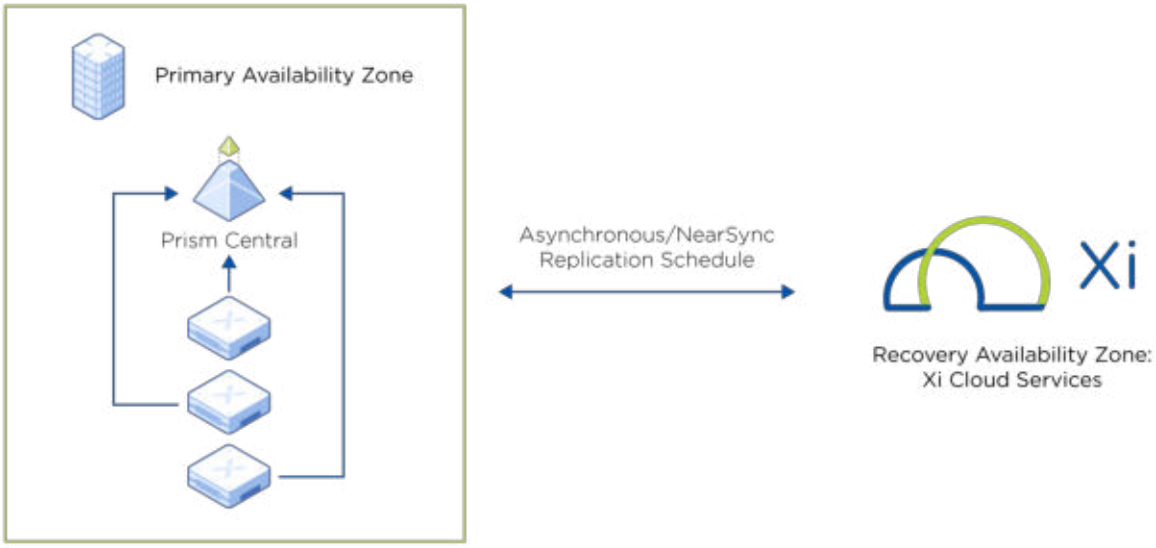


Figure 4: A primary on-prem site and one recovery site at Xi Cloud Services

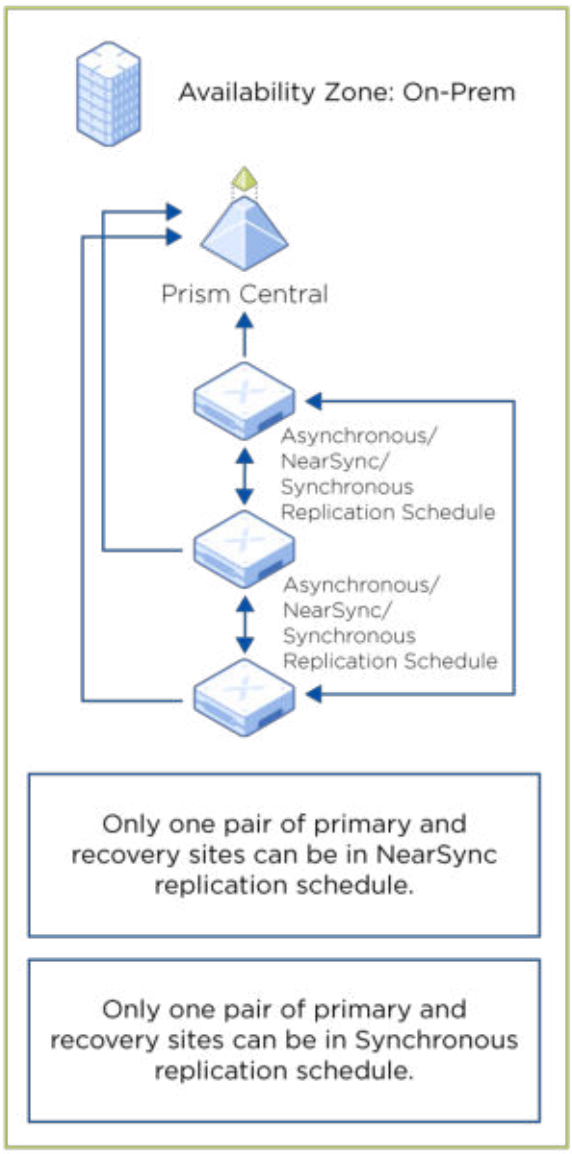


Figure 5: A primary Nutanix cluster and at most two recovery Nutanix clusters at the same on-prem site

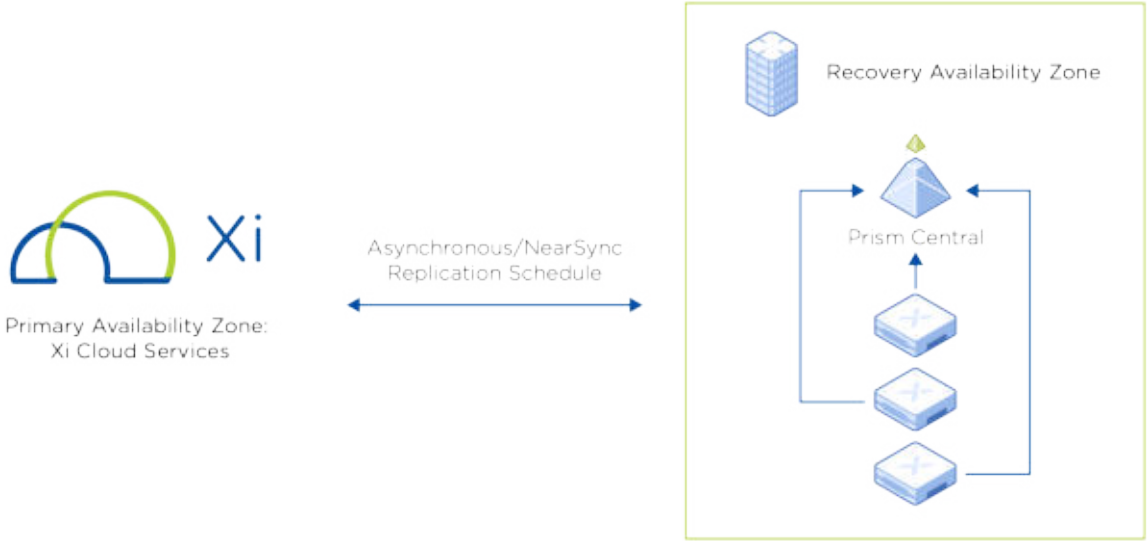


Figure 6: A primary site at Xi Cloud Services and recovery on-prem site

When paired, the primary site replicates the entities (protection policies, recovery plans, and recovery points) to the recovery sites in the specified time intervals (RPO). The approach helps application recovery at any of the recovery sites when there is a service disruption at the primary site (For example, natural disasters or scheduled maintenance). The entities start replicating back to the primary site when the primary site is up and running to ensure High Availability of applications. The entities you create or update synchronize continuously between the primary and recovery sites. The reverse synchronization enables you to create or update entities (protection policies, recovery plans, or guest VMs) at either the primary or the recovery sites.

This guide is primarily divided into the following two parts.

- Protection and DR between On-Prem Sites (Leap) on page 18

  The section walks you through the procedure of application protection and DR to other Nutanix clusters at the same or different on-prem sites. The procedure also applies to protection and DR to other Nutanix clusters in supported public cloud.

- Protection and DR between On-Prem Site and Xi Cloud Service (Xi Leap) on page 154

  Xi Leap is essentially an extension of Leap to Xi Cloud Services. You can protect applications and perform DR to Xi Cloud Services or from Xi Cloud Services to an on-prem availability zone. The section describes application protection and DR from Xi Cloud Services to an on-prem Nutanix cluster. For application protection and DR to Xi Cloud Services, refer the supported capabilities in Protection and DR between On-Prem Sites (Leap) on page 18 because the protection procedure remains the same when the primary site is an on-prem availability zone.

Configuration tasks and DR workflows are largely the same regardless of the type of recovery site. For more information about the protection and DR workflow, see Leap Deployment Workflow on page 12.

Leap Terminology

The following section describes the terms and concepts used throughout the guide. Nutanix recommends gaining familiarity with these terms before you begin configuring protection and Leap or Xi Leap disaster recovery (DR).

Availability Zone

A zone that can have one or more independent datacenters inter-connected by low latency links. An availability zone can either be in your office premises (on-prem) or in Xi Cloud Services. Availability zones are physically isolated from each other to ensure that a disaster at one availability zone does not affect another availability zone. An instance of Prism Central represents an on-prem availability zone.

> Note: An availability zone is referred to as a site throughout this document.

On-Prem Availability Zone

An availability zone (site) in your premises.

Xi Cloud Services

A site in the Nutanix Enterprise Cloud Platform (Xi Cloud Services).

Primary Availability Zone

A site that initially hosts guest VMs you want to protect.

Recovery Availability Zone

A site where you can recover the protected guest VMs when a planned or an unplanned event occurs at the primary site causing its downtime. You can configure at most two recovery sites for a guest VM.

Nutanix Cluster

A cluster running AHV or ESXi nodes on an on-prem availability zone, Xi Cloud Services, or any supported public cloud. Leap does not support guest VMs from Hyper-V clusters.

Prism Element

The GUI that provides you the ability to configure, manage, and monitor a single Nutanix cluster. It is a service built into the platform for every Nutanix cluster deployed.

Prism Central

The GUI that allows you to monitor and manage many Nutanix clusters (Prism Element running on those clusters). Prism Starter, Prism Pro, and Prism Ultimate are the three flavors of Prism Central.

Prism Central essentially is a VM that you deploy (host) in a Nutanix cluster (Prism Element). For more information about Prism Central, see *Prism Central Guide*. You can set up the following configurations of Prism Central VM.

Small Prism Central

A Prism Central VM with configuration equal to or less than 8 vCPU and 32 GB memory. The VM hot adds extra 4 GB and 1 GB memory when you enable Leap and Flow respectively in small Prism Central.

Small Prism Central (Single node)

A small Prism Central deployed in a single VM.

Small Prism Central (Scaleout)

Three small Prism Centrals deployed in three VMs in the same availability zone (site).

Large Prism Central

A Prism Central VM with configuration more than 8 vCPU and 32 GB memory. The VM hot adds extra 8 GB and 1 GB memory when you enable Leap and Flow respectively in large Prism Central.

Large Prism Central (Single node)

A large Prism Central deployed in a single VM.

Large Prism Central (Scaleout)

Three large Prism Centrals deployed in three VMs in the same availability zone (site).

> Note: A scaleout Prism Central works like a single node Prism Central in the availability zone (site). You can upgrade a single node Prism Central to scaleout Prism Central to increase the capacity, resiliency, and redundancy of Prism Central VM. For detailed information about the available configurations of Prism Central, see *Prism Central Scalability* in *Prism Central Release Notes* .

Virtual Private Cloud (VPC)

A logically isolated network service in Xi Cloud Services. A VPC provides the complete IP address space for hosting user-configured VPNs. A VPC allows creating workloads manually or by failover from a paired primary site.

The following VPCs are available in each Xi Cloud Services account. You cannot create more VPCs in Xi Cloud Services.

Production VPC

Used to host production workloads.

Test VPC

Used to test failover from a paired site.

Source Virtual Network

The virtual network from which guest VMs migrate during a failover or failback.

Recovery Virtual Network

The virtual network to which guest VMs migrate during a failover or failback operation.

Network Mapping

A mapping between two virtual networks in paired sites. A network mapping specifies a recovery network for all guest VMs of the source network. When you perform a failover or failback, the guest VMs in the source network recover in the corresponding (mapped) recovery network.

Category

A VM category is a key-value pair that groups similar guest VMs. Associating a protection policy with a VM category ensures that the protection policy applies to all the guest VMs in the group regardless of how the group scales with time. For example, you can associate a group of guest VMs with the `Department: Marketing` category, where `Department` is a category that includes a value `Marketing` along with other values such as `Engineering` and `Sales`.

VM categories remain the same way on on-prem sites and Xi Cloud Services. For more information about VM categories, see *Category Management* in *Prism Central Guide*.

Recovery Point

A copy of the state of a system at a particular point in time.

Crash-consistent Snapshots

A snapshot is crash-consistent if it captures all of the data components (write order consistent) at the instant of the crash. VM snapshots are crash-consistent (by default), which means that the vDisks that the snapshot captures are consistent with a single point in time. Crash-consistent snapshots are more suited for non-database operating systems and applications which may not support quiescence (freezing) and un-quiescence (thawing) and such as file servers, DHCP servers, print servers.

Application-consistent Snapshots

A snapshot is application-consistent if, in addition to capturing all of the data components (write order consistent) at the instant of the crash, the running applications have completed all their operations and flushed their buffers to disk (in other words, the application is quiesced). Application-consistent snapshots capture the same data as crash-consistent snapshots, with the addition of all data in memory and all transactions in process. Therefore, application-consistent snapshots may take longer to complete.

Application-consistent snapshots are more suited for systems and applications that can be quiesced and un-quiesced or thawed, such as database operating systems and applications such as SQL, Oracle, and Exchange.

Recoverable Entity

A guest VM that you can recover from a recovery point.

Protection Policy

A configurable policy that takes recovery points of the protected guest VMs in equal time intervals, and replicates those recovery points to the recovery sites.

Recovery Plan

A configurable policy that orchestrates the recovery of protected guest VMs at the recovery site.

Recovery Point Objective (RPO)

The time interval that refers to the acceptable data loss if there is a failure. For example, if the RPO is 1 hour, the system creates a recovery point every 1 hour. On recovery, you can recover the guest VMs with data as of up to 1 hour ago. **Take Snapshot Every** in the **Create Protection Policy** GUI represents RPO.

Recovery Time Objective (RTO)

The time period from failure event to the restored service. For example, an RTO of 30 minutes enables you to back up and run the protected guest VMs in 30 minutes after the failure event.

Nutanix Disaster Recovery Solutions

The following flowchart provides you with the detailed representation of the disaster recovery (DR) solutions of Nutanix. This decision tree covers both the DR solutions—protection domain-based DR and Leap helping you to make quick decisions on which DR strategy will best suit your environment.

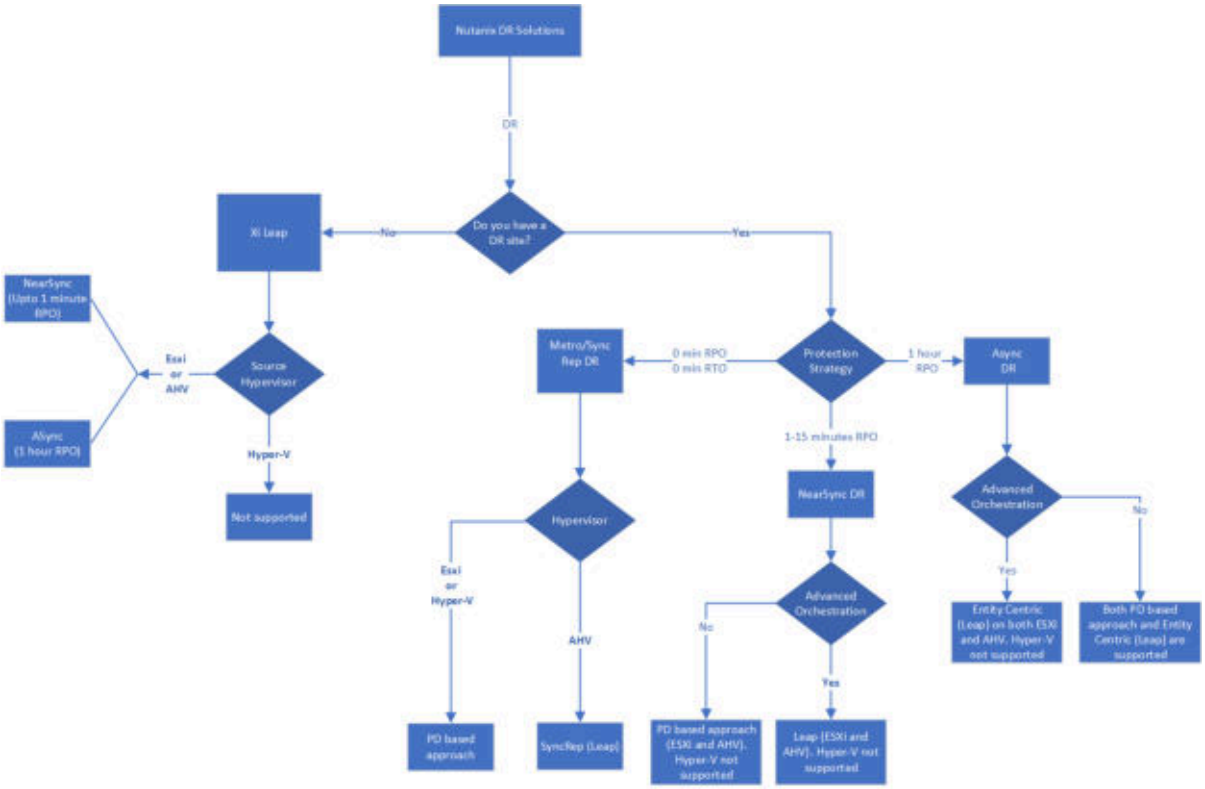


Figure 7: Decision Tree for Nutanix DR Solutions

For information about protection domain-based (legacy) DR, see *Data Protection and Recovery with Prism Element* guide. With Leap, you can protect your guest VMs and perform DR to on-prem availability zones (sites) or to Xi Cloud Services. A Leap deployment for DR from Xi Cloud Services to an on-prem Nutanix cluster is Xi Leap. The detailed information about Leap and Xi Leap DR configuration is available in the following sections of this guide.

Protection and DR between On-Prem Sites (Leap) on page 18

- For information about protection with Asynchronous replication schedule and DR, see Protection with Asynchronous Replication Schedule and DR (Leap) on page 40.
- For information about protection with NearSync replication schedule and DR, see Protection with NearSync Replication Schedule and DR (Leap) on page 100.
- For information about protection with Synchronous replication schedule and DR, see Protection with Synchronous Replication Schedule (0 RPO) and DR on page 116.

Protection and DR between On-Prem Site and Xi Cloud Service (Xi Leap) on page 154

- For information about protection with Asynchronous replication schedule and DR, see Protection with Asynchronous Replication and DR (Xi Leap) on page 196.
- For information about protection with NearSync replication schedule and DR, see Protection with NearSync Replication and DR (Xi Leap) on page 228.

Leap Deployment Workflow

The workflow for entity-centric protection and disaster recovery (DR) configuration is as follows. The workflow is largely the same for both Leap and Xi Leap configurations except a few extra steps you must perform while configuring Xi Leap.

Procedure

1. Enable Leap at the primary and recovery on-prem availability zones (Prism Central).

   Enable Leap at the on-prem availability zones (sites) only. For more information about enabling Leap, see Enabling Leap for On-Prem Site on page 36.

2. Pair the primary and recovery sites with each other.

   Only when you pair a site, the site lists for recovery sites while configuring protection policies and recovery plans (see step 6 and step 7). For more information about pairing the sites, see Pairing Availability Zones (Leap) on page 37.

3. (only for Xi Leap configuration) Set up your environment to proceed with replicating to Xi Cloud Services.

   For more information about environment setup, see Xi Leap Environment Setup on page 166.

4. (only for Xi Leap configuration) Reserve floating IP addresses.

   For more information about floating IP addresses, see *Floating IP Address Management* in *Xi Infrastructure Service Management Guide*.

5. Create production and test virtual networks at the primary and recovery sites.

   Create production and test virtual networks only at the on-prem sites. Xi Cloud Services create production and test virtual networks dynamically for you. However, Xi Cloud Services provides floating IP addresses (step 4), a feature that is not available for on-prem sites. For more information about production and test virtual networks, see Nutanix Virtual Networks on page 191.

6. Create a protection policy with replication schedules at the primary site.

   A protection policy can replicate recovery points to at most two other Nutanix clusters at the same or different sites. To replicate the recovery points, add a replication schedule between the primary site and each recovery site.

   - To create a protection policy with an Asynchronous replication schedule, see:
     - Creating a Protection Policy with an Asynchronous Replication Schedule (Leap) on page 42
     - Creating a Protection Policy with Asynchronous Replication Schedule (Xi Leap) on page 198
   - To create a protection policy with a NearSync replication schedule, see:
     - Creating a Protection Policy with a NearSync Replication Schedule (Leap) on page 104
     - Creating a Protection Policy with NearSync Replication Schedule (Xi Leap) on page 233

     > Note: To maintain the efficiency of minutely replication, protection policies allow you to add NearSync replication schedule between the primary site and only one recovery site.

   - To create a protection policy with the Synchronous replication schedule, see Creating a Protection Policy with the Synchronous Replication Schedule (Leap) on page 119.

     > Note: To maintain the efficiency of synchronous replication, protection policies allow you to add only one recovery site when you add Synchronous replication schedule. If you already have an Asynchronous or a NearSync replication schedule in the protection

policy, you cannot add another recovery site to protect the guest VMs with Synchronous replication schedule.

You can also create a protection policy at a recovery site. Protection policies you create or update at a recovery site synchronize back to the primary site. The reverse synchronization helps when you protect more guest VMs in the same protection policy at the recovery site.

7. Create a recovery plan at the primary site.

   A recovery plan orchestrates the failover of the protected guest VMs (step 6) to a recovery site. For two recovery sites, create two discrete recovery plans at the primary site—one for DR to each recovery site.

   - To create a recovery plan for DR to another Nutanix cluster at the same or different on-prem sites, see Creating a Recovery Plan (Leap) on page 60.
   - To create a recovery plan for DR to Xi Cloud Services, see Creating a Recovery Plan (Xi Leap) on page 209.

   You can also create a recovery plan at a recovery site. The recovery plan you create or update at a recovery site synchronizes back to the primary site. The reverse synchronization helps in scenarios where you add more guest VMs to the same recovery plan at the recovery site.

8. Validate or test the recovery plan you create in step 7.

   To test a recovery plan, perform a test failover to a recovery site.

   - To perform test failover to another Nutanix cluster at the same or different on-prem sites, see Performing a Test Failover (Leap) on page 77.
   - To perform test failover to Xi Cloud Services, see Failover and Failback Operations (Xi Leap) on page 216.

9. (only for Xi Leap configuration) After the failover to recovery site, enable external connectivity. To enable external connectivity, perform the following.

   - 1. After a planned failover, shut down the VLAN interface on the on-prem Top-of-Rack (TOR) switch.
     2. To access the Internet from Xi Cloud Services, create both inbound and outbound policy-based routing (PBR) policies on the virtual private cloud (VPC). For more information, see *Policy Configuration* in *Xi Infrastructure Service Administration Guide*.

10. (only for Xi Leap configuration) Perform the following procedure to access the recovered guest VMs through the Internet.

    - 1. Assign a floating IP address to the guest VMs failed over to Xi Cloud Services. For more information, see *Floating IP Addresses Management* in *Xi Infrastructure Service Administration Guide*
      2. Create PBR policies and specify the internal or private IP address of the guest VMs. For more information, see *Policy Configuration* in *Xi Infrastructure Service Administration Guide*.

    Note: If a guest VM (that hosts a publicly accessible website) fails over, update the authoritative DNS server (for example, Amazon Route 53, GoDaddy, DNSmadeEasy) with the primary failover record (on-prem public IP address) and the secondary failover record (Xi floating IP address). For example, if your authoritative DNS server is Amazon Route53, configure the primary and the secondary failover records. Amazon Route53 performs the health checks on the primary failover record and returns the secondary failover record when the primary is down.

On-Prem Hardware Resource Requirements

For DR solutions with Asynchronous, NearSync, and Synchronous replication schedules to succeed, the nodes in the on-prem Availability Zones (AZs or sites) must have certain resources. This section provides information about the node, disk and Foundation configurations necessary to support the RPO-based recovery point frequencies.

- The conditions and configurations provided in this section apply to Local and Remote recovery points.
- Any node configuration with two or more SSDs, each SSD being 1.2 TB or greater capacity, supports recovery point frequency for NearSync.
- Any node configuration that supports recovery point frequency of six (6) hours also supports AHV-based Synchronous replication schedules because a protection policy with Synchronous replication schedule takes recovery points of the protected VMs every 6 hours. See Protection with Synchronous Replication Schedule (0 RPO) and DR on page 116 for more details about Synchronous replication.
- Both the primary cluster and replication target cluster must fulfill the same minimum resource requirements.
- Ensure that any new node or disk additions made to the on-prem sites (Availability Zones) meet the minimum requirements.
- Features such as Deduplication and RF3 may require additional memory depending on the DR schedules and other workloads run on the cluster.

> Note: In case of on-prem deployments, the default minimum recovery point frequency using the default Foundation configuration is 6 hours. To increase the recovery point frequency, you must also modify the Foundation configuration (SSD and CVM) accordingly. For example, an all-flash setup with a capacity between 48 TB to 92 TB has the default recovery point frequency is 6 hours. If you want to decrease the recovery point interval to one (1) hour, you must modify the default Foundation configuration to:
>
> - 14 vCPUs for CVM
> - 40 GB for CVM

The table lists the supported frequency for the recovery points across various hardware configurations.

Table 1: Recovery Point Frequency

Type of disk	Capacity per node	Minimum recovery point frequency	Foundation Configuration - SSD and CVM requirements
Hybrid	Total HDD tier capacity of 32 TB or lower. Total capacity (HDD + SSD) of 40 TB or lower.	• NearSync • Async (Hourly)	No change required —Default Foundation configuration. • 2 x SSDs • Each SSD must be minimum 1.2 TB or more for NearSync.
	Total HDD tier capacity between 32-64 TB. Total capacity (HDD + SSD) of 92 TB or lower. Up to 64 TB HDD Up to 32 TB SSD (4 x 7.68 TB SSDs)	• NearSync • Async (Hourly)	Modify Foundation configurations to minimum: • 4 x SSDs • Each SSD must be minimum 1.2 TB or more for NearSync. • 14 vCPU for CVM • 40 GB for CVM
	Total HDD tier capacity between 32-64 TB. Total capacity (HDD + SSD) of 92 TB or lower. Up to 64 TB HDD Up to 32 TB SSD	Async (Every 6 hours)	No change required —Default Foundation configuration.
	Total HDD tier capacity between 64-80 TB. Total capacity (HDD + SSD) of 96 TB or lower.	Async (Every 6 hours)	No change required —Default Foundation configuration.
	Total HDD tier capacity greater than 80 TB. Total capacity (HDD + SSD) of 136 TB or lower.	Async (Every 6 hours)	Modify Foundation configurations to minimum: • 12 vCPU for CVM • 36 GB for CVM

Type of disk	Capacity per node	Minimum recovery point frequency	Foundation Configuration - SSD and CVM requirements
All Flash	Total capacity of 48 TB or lower	• NearSync • Async (Hourly)	No change required —Default Foundation configuration. Each SSD must be minimum 1.2 TB or more for NearSync.
	Total capacity between 48-92 TB	• NearSync • Async (Hourly)	Modify Foundation configurations to minimum: • 14 vCPU for CVM • Each SSD must be minimum 1.2 TB or more for NearSync. • 40 GB for CVM
	Total capacity between 48-92 TB	Async (Every 6 hours)	No change required —Default Foundation configuration.
	Total capacity greater than 92 TB	Async (Every 6 hours)	Modify Foundation configurations to minimum: • 12 vCPU for CVM • 36 GB for CVM

2

PROTECTION AND DR BETWEEN ON-PREM SITES (LEAP)

Leap protects your guest VMs and orchestrates their disaster recovery (DR) to other Nutanix clusters when events causing service disruption occur at the primary availability zone (site). For protection of your guest VMs, protection policies with Asynchronous, NearSync, or Synchronous replication schedules generate and replicate recovery points to other on-prem availability zones (sites). Recovery plans orchestrate DR from the replicated recovery points to other Nutanix clusters at the same or different on-prem sites.

Protection policies create a recovery point—and set its expiry time—in every iteration of the specified time period (RPO). For example, the policy creates a recovery point every 1 hour for an RPO schedule of 1 hour. The recovery point expires at its designated expiry time based on the retention policy—see step 3 in Creating a Protection Policy with an Asynchronous Replication Schedule (Leap) on page 42. If there is a prolonged outage at a site, the Nutanix cluster retains the last recovery point to ensure you do not lose all the recovery points. For NearSync replication (lightweight snapshot), the Nutanix cluster retains the last full hourly snapshot. During the outage, the Nutanix cluster does not clean up the recovery points due to expiry. When the Nutanix cluster comes online, it cleans up the recovery points that are past expiry immediately.

For High Availability of a guest VM, Leap enables replication of its recovery points to one or more on-prem sites. A protection policy can replicate recovery points to maximum two on-prem sites. For replication, you must add a replication schedule between sites. You can set up the on-prem sites for protection and DR in the following arrangements.

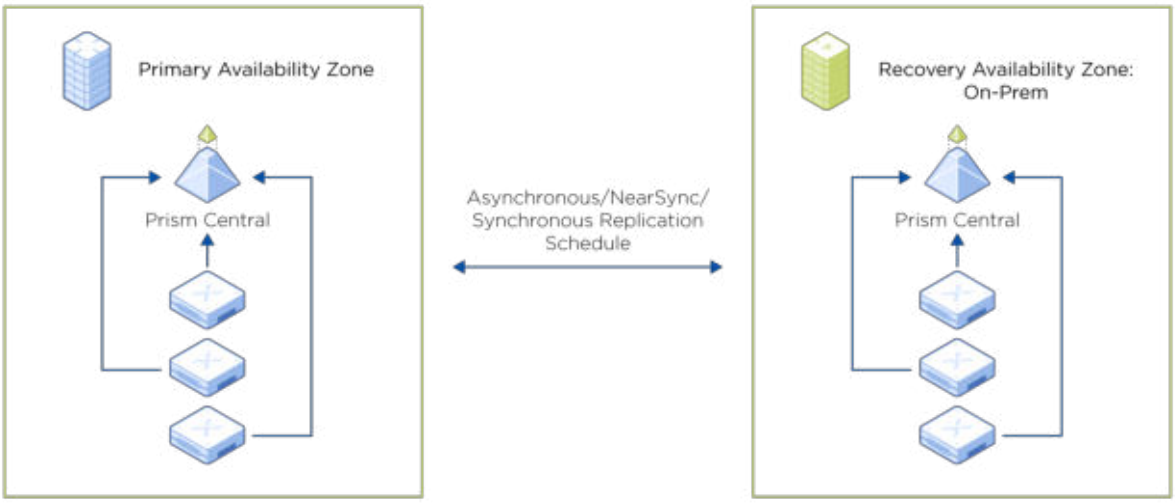


Figure 8: The Primary and recovery Nutanix clusters at the different on-prem sites

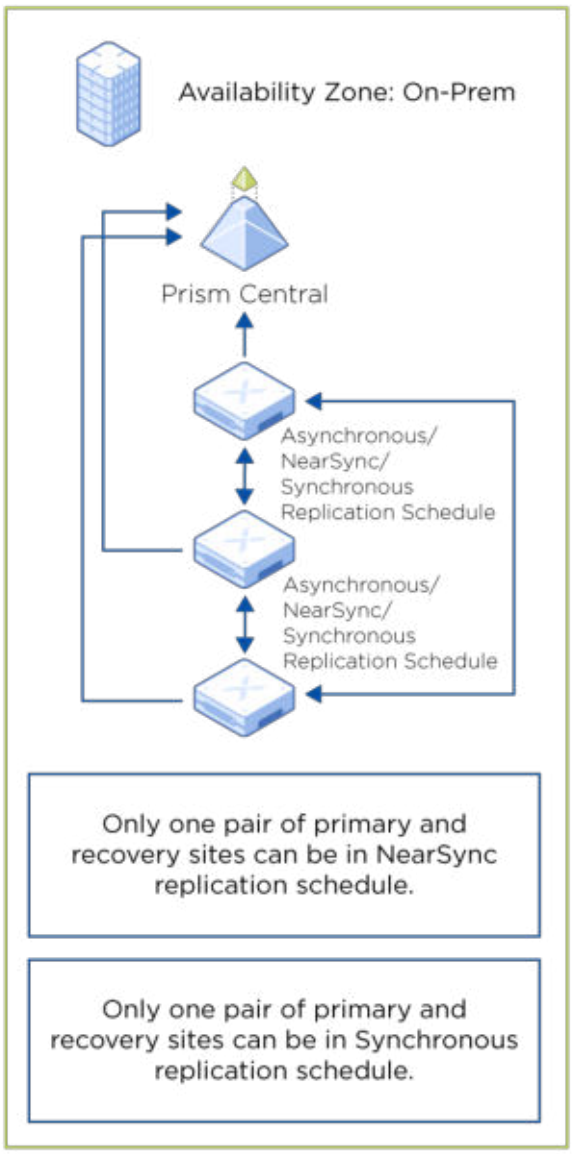


Figure 9: The Primary and recovery Nutanix clusters at the same on-prem site

The replication to multiple sites enables DR to Nutanix clusters at all the sites where the recovery points replicate or exist. To enable performing DR to a Nutanix cluster at the same or different site (recovery site), you must create a recovery plan. To enable performing DR to two different Nutanix clusters at the same or different recovery sites, you must create two discrete recovery plans—one for each recovery site. In addition to performing DR to Nutanix clusters running the same hypervisor type, you can also perform cross-hypervisor disaster recovery (CHDR)—DR from AHV clusters to ESXi clusters, or from ESXi clusters to AHV clusters.

The protection policies and recovery plans you create or update synchronize continuously between the primary and recovery on-prem sites. The reverse synchronization enables you to create or update entities (protection policies, recovery plans, and guest VMs) at either the primary or the recovery sites.

The following section describes protection of your guest VMs and DR to a Nutanix cluster at the same or different on-prem sites. The workflow is the same for protection and DR to a Nutanix cluster in supported public cloud platforms. For information about protection of your guest

VMs and DR from Xi Cloud Services to an on-prem Nutanix cluster (Xi Leap), see Protection and DR between On-Prem Site and Xi Cloud Service (Xi Leap) on page 154.

Leap Requirements

The following are the general requirements of Leap. Along with the general requirements, there are specific requirements for protection with the following supported replication schedules.

- For information about the on-prem node, disk and Foundation configurations required to support Asynchronous, NearSync, and Synchronous replication schedules, see On-Prem Hardware Resource Requirements on page 15.
- For specific requirements of protection with Asynchronous replication schedule (1 hour or greater RPO), see Asynchronous Replication Requirements (Leap) on page 40.
- For specific requirements of protection with NearSync replication schedule (1–15 minutes RPO), see NearSync Replication Requirements (Leap) on page 102.
- For specific requirements of protection with Synchronous replication schedule (0 RPO), see Synchronous Replication Requirements on page 117.

License Requirements

The AOS license required depends on the features that you want to use. For information about the features that are available with an AOS license, see Software Options.

Hypervisor Requirements

The underlying hypervisors required differ in all the supported replication schedules. For more information about underlying hypervisor requirements for the supported replication schedules, see:

- Asynchronous Replication Requirements (Leap) on page 40
- NearSync Replication Requirements (Leap) on page 102
- Synchronous Replication Requirements on page 117

Nutanix Software Requirements

- Each on-prem availability zone (site) must have a Leap enabled Prism Central instance. To enable Leap in Prism Central, see Enabling Leap for On-Prem Site on page 36.

> Note: If you are using ESXi, register at least one vCenter Server to Prism Central. You can also register two vCenter Servers, each to a Prism Central at different sites. If you register both the Prism Central to the single vCenter Server, ensure that each ESXi cluster is part of different datacenter object in vCenter.

- The primary and recovery Prism Central and Prism Element on the Nutanix clusters must be running on the supported AOS versions. For more information about the required versions for the supported replication schedules, see:
  - Asynchronous Replication Requirements (Leap) on page 40
  - NearSync Replication Requirements (Leap) on page 102
  - Synchronous Replication Requirements on page 117

  > Tip:
  >
  > Nutanix supports replications between the all the latest supported LTS and STS released AOS versions. To check the list of the latest supported AOS versions, see KB-5505. To determine if the AOS versions currently running on your clusters are EOL, see the EOL document .
  >
  > Upgrade the AOS version to the next available supported LTS/STS release. To determine if an upgrade path is supported, check the Upgrade Paths page before you upgrade the AOS.
  >
  > > Note: If both clusters have different AOS versions that are EOL, upgrade the cluster with lower AOS version to match the cluster with higher AOS version and then perform the upgrade to the next supported LTS version.
  >
  > For example, the clusters are running AOS versions 5.5.x and 5.10.x respectively. Upgrade the cluster on 5.5.x to 5.10.x. After both the clusters are on 5.10.x, proceed to upgrade each cluster to 5.15.x (supported LTS). Once both clusters are on 5.15.x you can upgrade the clusters to 5.20.x or newer.
  >
  > Nutanix recommends that both the primary and the replication clusters or sites run the same AOS version.

User Requirements

You must have one of the following roles in Prism Central.

- User admin
- Prism Central admin
- Prism Self Service admin
- Xi admin

To view the available roles or create a role, click the hamburger icon at the top-left corner of the window and go to **Administration** > **Roles** in the left pane.

Firewall Port Requirements

To allow two-way replication between Nutanix clusters at the same or different sites, you must enable certain ports in your external firewall. To know about the required ports, see *Disaster Recovery - Leap* in Port Reference.

Networking Requirements

Requirements for static IP address preservation after failover

You can preserve one IP address of a guest VM (with static IP address) for its failover (DR) to an IPAM network. After the failover, the other IP addresses of the guest VM have

to be reconfigured manually. To preserve an IP address of a guest VM (with static IP address), ensure that:

> Caution: By default, you cannot preserve statically assigned DNS IP addresses after failover (DR) of guest VMs. However, you can create custom in-guest scripts to preserve the statically assigned DNS IP addresses. For more information, see Creating a Recovery Plan (Leap) on page 60.

- Both the primary and the recovery Nutanix clusters run AOS 5.11 or newer.
- The protected guest VMs have Nutanix Guest Tools (NGT) version 1.5 or newer installed.

  For information about installing NGT, see *Nutanix Guest Tools* in *Prism Web Console Guide*.
- The protected guest VMs have at least one empty CD-ROM slot.

  The empty CD-ROM is required for mounting NGT at the recovery site.
- The protected guest VMs can reach the Controller VM from both the sites.
- The protected guest VMs have NetworkManager command-line tool (nmcli) version 0.9.10.0 or newer installed.

  Also, the NetworkManager must manage the networks on Linux VMs. To enable NetworkManager on a Linux VM, in the interface configuration file, set the value of the `NM_CONTROLLED` field to `yes`. After setting the field, restart the network service on the VM.

  > Tip: In CentOS, the interface configuration file is `/etc/sysconfig/network-scripts/ifcfg-eth0`.

Requirements for static IP address mapping of guest VMs between source and target virtual networks

You can explicitly define IP addresses for guest VMs that have static IP addresses on the primary site. On recovery, such guest VMs retain the explicitly defined IP address. To map static IP addresses of guest VMs between source and target virtual networks, ensure that:

- Both the primary and the recovery Nutanix clusters run AOS 5.17 or newer.
- The protected guest VMs have static IP addresses at the primary site.
- The protected guest VMs have Nutanix Guest Tools (NGT) version 1.5 or newer installed.

  For information about installing NGT, see *Nutanix Guest Tools* in *Prism Web Console Guide*.
- The protected guest VMs have at least one empty CD-ROM slot.

  The empty CD-ROM is required for mounting NGT at the recovery site.
- The protected guest VMs can reach the Controller VM from both the sites.
- The recovery plan selected for failover has VM-level IP address mapping configured.

Virtual network design requirements

You can design the virtual subnets that you plan to use for DR to the recovery site so that they can accommodate the guest VMs running in the source virtual network.

- Maintain a uniform network configuration for all the virtual LANs (VLANs) with the same VLAN ID and network range in all the Nutanix clusters at a site. All such VLANs

must have the same subnet name, IP address range, and IP address prefix length (**(Gateway IP/Prefix Length)**).

For example, if you have VLAN with id 0 and network 10.45.128.0/17, and three clusters PE1, PE2, and PE3 at the site AZ1, all the clusters must maintain the same name, IP address range, and IP address prefix length (**(Gateway IP/Prefix Length)**), for VLAN with id 0.

- To use a virtual network as a recovery virtual network, ensure that the virtual network meets the following requirements.
  - The network prefix is the same as the network prefix of the source virtual network. For example, if the source network address is 192.0.2.0/24, the network prefix of the recovery virtual network must also be 24.
  - The gateway IP address is the same as the gateway IP address in the source network. For example, if the gateway IP address in the source virtual network 192.0.2.0/24 is 192.0.2.10, the last octet of the gateway IP address in the recovery virtual network must also be 10.
- To use a single Nutanix cluster as a target for DR from multiple primary Nutanix clusters, ensure that the number of virtual networks on the recovery cluster is equal to the sum of the number of virtual networks on the individual primary Nutanix clusters. For example, if there are two primary Nutanix clusters, with one cluster having `m` networks and the other cluster having `n` networks, ensure that the recovery cluster has `m + n` networks. Such a design ensures that all recovered VMs attach to a network.

Additional Requirements

- Both the primary and recovery Nutanix clusters must have an external IP address.
- Both the primary and recovery Prism Centrals and Nutanix clusters must have a data services IP address.
- The Nutanix cluster that hosts the Prism Central must meet the following requirements.
  - The Nutanix cluster must be registered to the Prism Central instance.
  - The Nutanix cluster must have an iSCSI data services IP address configured on it.
  - The Nutanix cluster must also have sufficient memory to support a hot add of memory to all Prism Central nodes when you enable Leap. A small Prism Central instance (4 vCPUs, 16 GB memory) requires a hot add of 4 GB, and a large Prism Central instance (8 vCPUs, 32 GB memory) requires a hot add of 8 GB. If you enable Nutanix Flow, each Prism Central instance requires an extra hot-add of 1 GB.
- Each node in a scaled-out Prism Central instance must have a minimum of 4 vCPUs and 16 GB memory.

  For more information about the scaled-out deployments of a Prism Central, see Leap Terminology on page 8.
- The protected guest VMs must have Nutanix VM mobility drivers installed.

  Nutanix VM mobility drivers are required for accessing the guest VMs after failover. Without Nutanix VM mobility drivers, the guest VMs become inaccessible after a failover.
- Maintain a uniform network configuration for all the virtual LANs (VLANs) with the same VLAN ID and network range in all the clusters at an availability zone (site). All such VLANs

must have the same subnet name, IP address range, and IP address prefix length (**(Gateway IP/Prefix Length)**).

For example, if you have VLAN with id 0 and network 10.45.128.0/17, and three clusters PE1, PE2, and PE3 at the site AZ1, all the clusters must maintain the same name, IP address range, and IP address prefix length (**(Gateway IP/Prefix Length)**), for VLAN with id 0.

Leap Limitations

Consider the following general limitations before configuring protection and disaster recovery (DR) with Leap. Along with the general limitations, there are specific protection limitations with the following supported replication schedules.

- For specific limitations of protection with Asynchronous replication schedule (1 hour or greater RPO), see Asynchronous Replication Limitations (Leap) on page 42.
- For specific limitations of protection with NearSync replication schedule (1–15 minutes RPO), see NearSync Replication Limitations (Leap) on page 104.
- For specific limitations of protection with Synchronous replication schedule (0 RPO), see Synchronous Replication Limitations on page 118.

Virtual Machine Limitations

You cannot do or implement the following.

- Deploy witness VMs.
- Protect multiple guest VMs that use disk sharing (for example, multi-writer sharing, Microsoft Failover Clusters, Oracle RAC).
- Protect VMware fault tolerance enabled guest VMs.
- Recover vGPU console enabled guest VMs efficiently.

  When you perform DR of vGPU console-enabled guest VMs, the VMs recover with the default VGA console (without any alert) instead of vGPU console. The guest VMs fail to recover when you perform cross-hypervisor disaster recovery (CHDR).
- Configure NICs for a guest VM across both the virtual private clouds (VPC).

  You can configure NICs for a guest VM associated with either production or test VPC.

Volume Groups Limitation

You cannot protect volume groups.

Network Segmentation Limitation

You cannot apply network segmentation for DR. If you enable network segmentation for DR, the replications fail.

Virtual Network Limitation

Although there is no limit to the number of VLANs that you can create, only the first 500 VLANs list in the drop-down of **Network Settings** while creating a recovery plan. For more information about VLANs in the recovery plan, see Nutanix Virtual Networks on page 191.

Nutanix to vSphere Cluster Mapping Limitation

Due to the way the Nutanix architecture distributes data, there is limited support for mapping a Nutanix cluster to multiple vSphere clusters. If a Nutanix cluster is split into multiple vSphere clusters, migrate and recovery operations fail.

Leap Recommendations

Nutanix recommends the following best practices for configuring protection and disaster recovery (DR) with Leap.

General Recommendations

- Create all entities (protection policies, recovery plans, and VM categories) at the primary availability zone (site).
- Upgrade Prism Central before upgrading Prism Element on the Nutanix clusters registered to it.
- Do not add the guest VMs protected with Asynchronous or NearSync replication schedule, and Synchronous replication schedule in the same recovery plan.
- Disable Synchronous replication before unpairing the sites.

  If you unpair the sites while the guest VMs in the Nutanix clusters are still in synchronization, the Nutanix cluster becomes unstable. For more information about disabling Synchronous replication, see Synchronous Replication Management on page 125.

Recommendation for Migrating Protection Domains to Protection Policies

You can protect a guest VM either with legacy DR solution (protection domain-based) or with Leap. To protect a legacy DR-protected guest VM with Leap, you must migrate the guest VM from protection domain to a protection policy. During the migration, do not delete the guest VM snapshots in the protection domain. Nutanix recommends keeping the guest VM snapshots in the protection domain until the first recovery point for the guest VM is available on Prism Central. For more information, see Migrating Guest VMs from a Protection Domain to a Protection Policy on page 248.

Recommendation for DR to Nutanix Clusters at the Same On-Prem Availability Zone

If the single Prism Central that you use for protection and DR to Nutanix clusters at the same availability zone (site) becomes inactive, you cannot perform a failover when required. To avoid the single point of failure in such deployments, Nutanix recommends installing the single Prism Central at a different site (different fault domain).

Recommendation for Virtual Networks

- Map the networks while creating a recovery plan in Prism Central.
- Recovery plans do not support overlapping subnets in a network-mapping configuration. Do not create virtual networks that have the same name or overlapping IP address ranges.

Recommendation for Container Mapping

Create storage containers with the same name on both the primary and recovery Nutanix clusters.

Leap automatically maps the storage containers during the first replication (seeding) of a guest VM. If a storage container with the same name exists on both the primary and recovery Nutanix clusters, the recovery points replicate to the same name storage container only. For example, if your protected guest VMs are in the *SelfServiceContainer* on the primary Nutanix cluster,

and the recovery Nutanix cluster also has *SelfServiceContainer*, the recovery points replicate to *SelfServiceContainer* only. If a storage container with the same name does not exist at the recovery site, the recovery points replicate to a random storage container at the recovery site. For more information about creating storage containers on the Nutanix clusters, see *Creating a Storage Container* in *Prism Web Console Guide*.

Leap Service-Level Agreements (SLAs)

Leap enables protection of your guest VMs and disaster recovery (DR) to one or more Nutanix clusters at the same or different on-prem sites. A Nutanix cluster is essentially an AHV or an ESXi cluster running AOS. In addition to performing DR to Nutanix clusters running the same hypervisor type, you can also perform cross-hypervisor disaster recovery (CHDR)—DR from AHV clusters to ESXi clusters, or from ESXi clusters to AHV clusters.

Leap supports DR (and CHDR) to maximum two different Nutanix clusters at the same or different availability zones (sites). You can protect your guest VMs with the following replication schedules.

- Asynchronous replication schedule (1 hour or greater RPO). For information about protection with Asynchronous replication schedule, see Protection with Asynchronous Replication Schedule and DR (Leap) on page 40.
- NearSync replication schedule (1–15 minute RPO). For information about protection with NearSync replication schedule, see Protection with NearSync Replication Schedule and DR (Leap) on page 100.
- Synchronous replication schedule (0 RPO). For information about protection with Synchronous replication schedule, see Protection with Synchronous Replication Schedule (0 RPO) and DR on page 116.

  To maintain the efficiency in protection and DR, Leap allows to protect a guest VM with Synchronous replication schedule to only one AHV cluster and at the different on-prem availability zone.

Leap Views

The disaster recovery (DR) views enable you to perform CRUD operations on the following types of Leap entities.

- Configured entities (for example, availability zones, protection policies, and recovery plans)
- Created entities (for example, guest VMs, and recovery points)

This chapter describes the views of Prism Central (on-prem site).

Availability Zones View

The **Availability Zones** view under the **hamburger icon** > **Administration** lists all of your paired availability zones.

The following figure is a sample view, and the tables describe the fields and the actions that you can perform in this view.

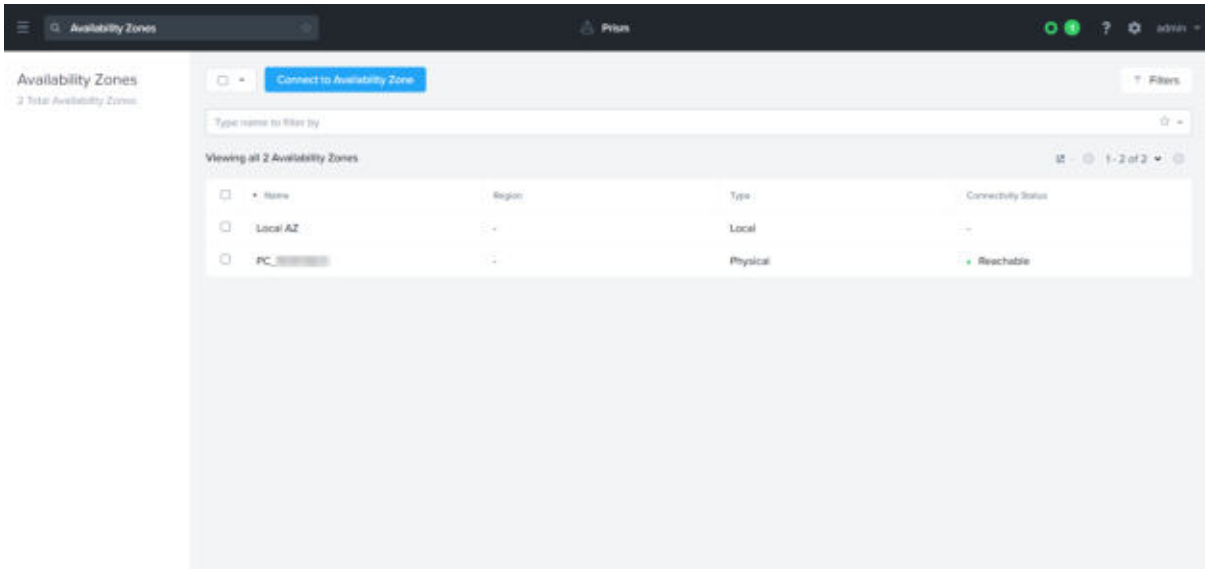


Figure 10: Availability Zones View

Table 2: Availability Zones View Fields

Field	Description
Name	Name of the availability zone.
Region	Region to which the availability zone belongs.
Type	Type of availability zone. Availability zones that are backed by on-prem Prism Central instances are shown to be of type physical. The availability zone that you are logged in to is shown as a local availability zone.
Connectivity Status	Status of connectivity between the local availability zone and the paired availability zone.

Table 3: Workflows Available in the Availability Zones View

Workflow	Description
Connect to Availability Zone (on-prem Prism Central only)	Connect to an on-prem Prism Central or to a Xi Cloud Services for data replication.

Table 4: Actions Available in the Actions Menu

Action	Description
Disconnect	Disconnect the remote availability zone. When you disconnect an availability zone, the pairing is removed.

Protection Summary View

The **Protection Summary** view under the **hamburger icon** > **Data Protection** shows detailed information about the Leap entities in an availability zone (site) and helps you generate DR reports for the specified time. The information in the **Protection Summary** view enables you to monitor the health of your DR deployments (Leap) and the activities performed on Leap entities. Select a topology in the left-hand side pane and protection and recovery information of the selected topology shows up in DR widgets on the right-hand side pane. The following figures are a sample view, and the tables describe the fields and the actions that you can perform in the DR widgets.

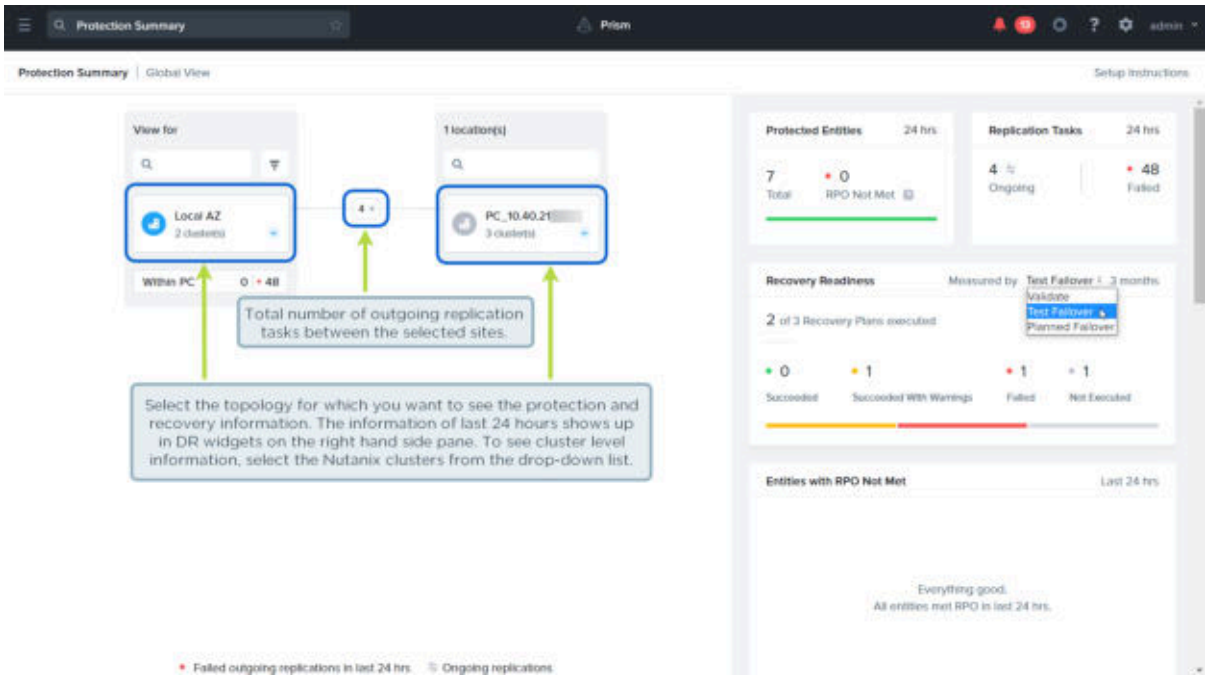


Figure 11: Protection Summary: Protected Entities, Replication Tasks, and Recovery Readiness

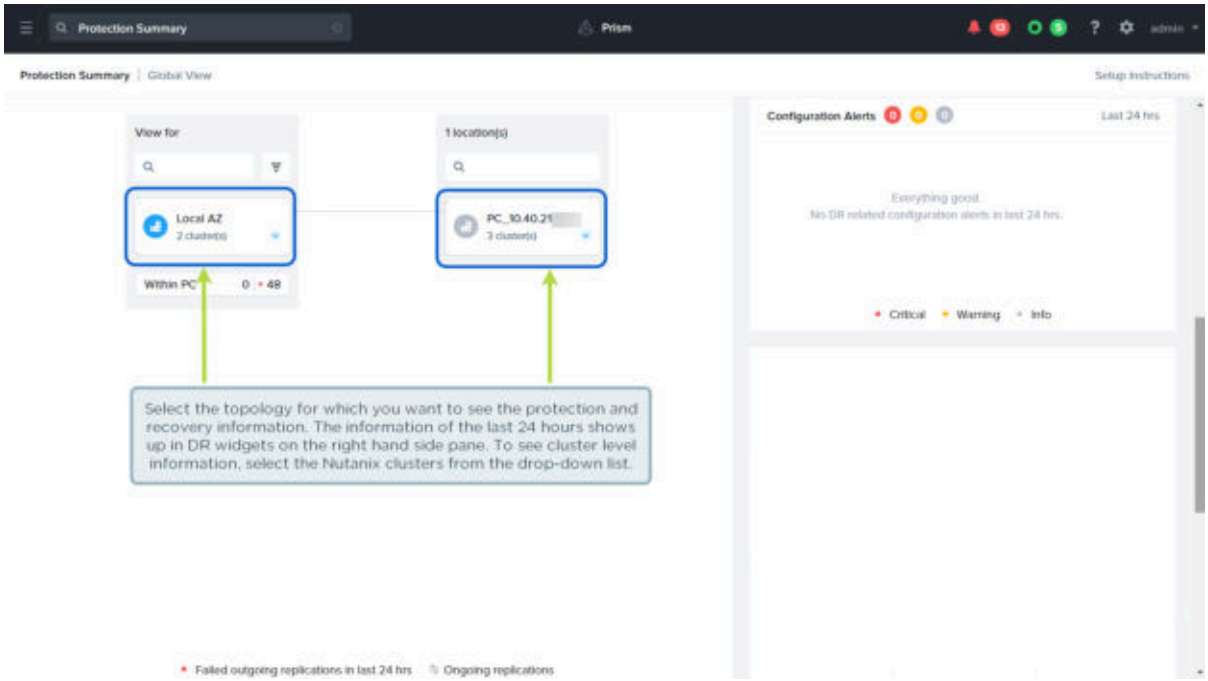


Figure 12: Protection Summary: Configuration Alerts

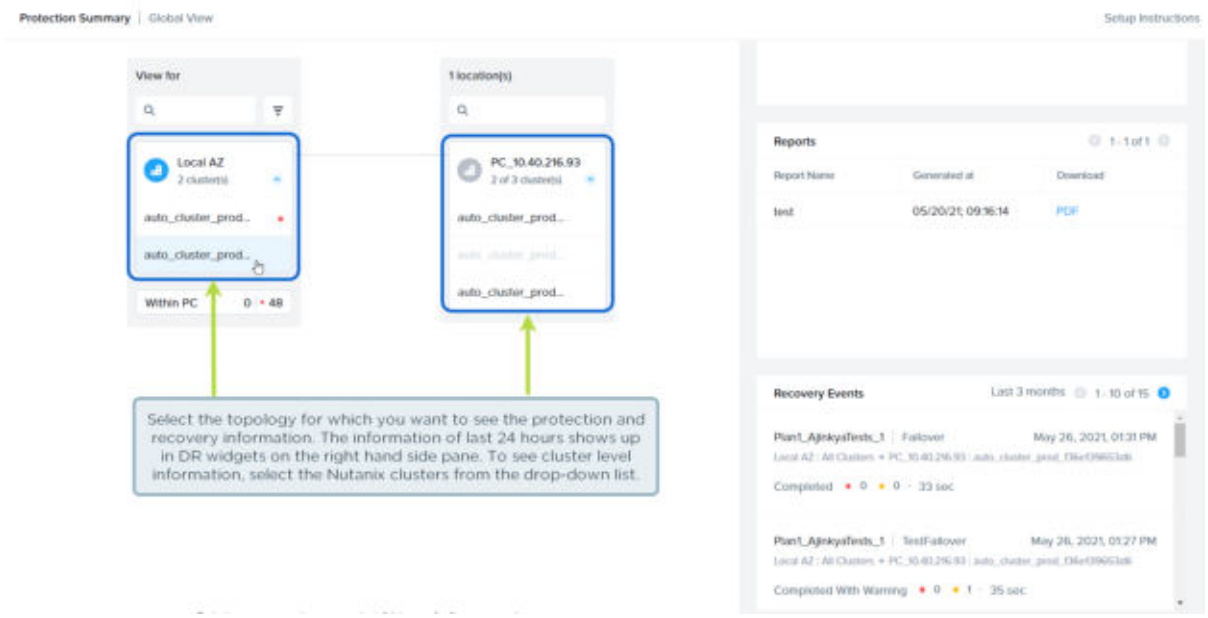


Figure 13: Protection Summary: Reports and Recovery Events

Table 5: Protected Entities

Field	Description
Total	Number of guest VMs protected. Clicking the number shows the guest VMs protected in protection policies.

Field	Description
RPO Not Met	Number of guest VMs that are protected but do not meet the specified RPO. Clicking the number shows the guest VMs that do not meet the specified RPO.

Table 6: Replication Tasks

Field	Description
Ongoing	Number of ongoing replication tasks.
Stuck	Number of replication tasks that are stuck. Clicking the number shows the stuck alerts generated in **Alerts**.
Failed	Number of replication tasks that failed. Clicking the number shows the alerts generated in **Alerts**.

Table 7: Recovery Readiness

Field	Description
Measured by	Failover operations to check the readiness of the recovery plans. You can use Validate, Test Failover, or Planned Failover from the drop-down list to check recovery readiness.
Succeeded	Number of recovery plans on which the selected failover operation ran successfully. Clicking the number shows the recovery plans on which the selected failover operation ran successfully.
Succeeded With Warnings	Number of recovery plans on which the selected failover operation ran successfully but with warnings. Clicking the number shows the recovery plans on which the selected failover operation ran successfully with warnings.
Failed	Number of recovery plans on which the selected failover operation failed to run successfully. Clicking the number shows the recovery plans on which the selected failover operation failed to run successfully.
Not Executed	Number of recovery plans on which no failover operation ran. Clicking the number shows the recovery plans on which no failover operation ran.

Table 8: Entities with RPO Not Met

Field	Description
Name	Names of guest VMs that do not meet the specified RPO. You can use the filters on the guest VMs to investigate the reason for RPO not meeting.

Table 9: Configuration Alerts

Field	Description
Alert Description	Description of configuration alerts raised on protection policies and recovery plans.
Impacted Entity	The entities impacted by the configuration alerts.

Table 10: Reports

Field	Description
Report Name	Name of the report.
Generated at	Date and time when the report was generated.
Download	Option to download the report as a PDF or a CSV document.

Recovery Events

This widget shows you a detailed view of the **Recovery Readiness**. You can view information about all the recovery plans that ran on the selected sites in the last 3 months.

Protection Policies View

The **Protection Policies** view under the **hamburger icon** > **Data Protection** lists all of configured protection policies from all the paired availability zones.

The following figure is a sample view, and the tables describe the fields and the actions that you can perform in this view.

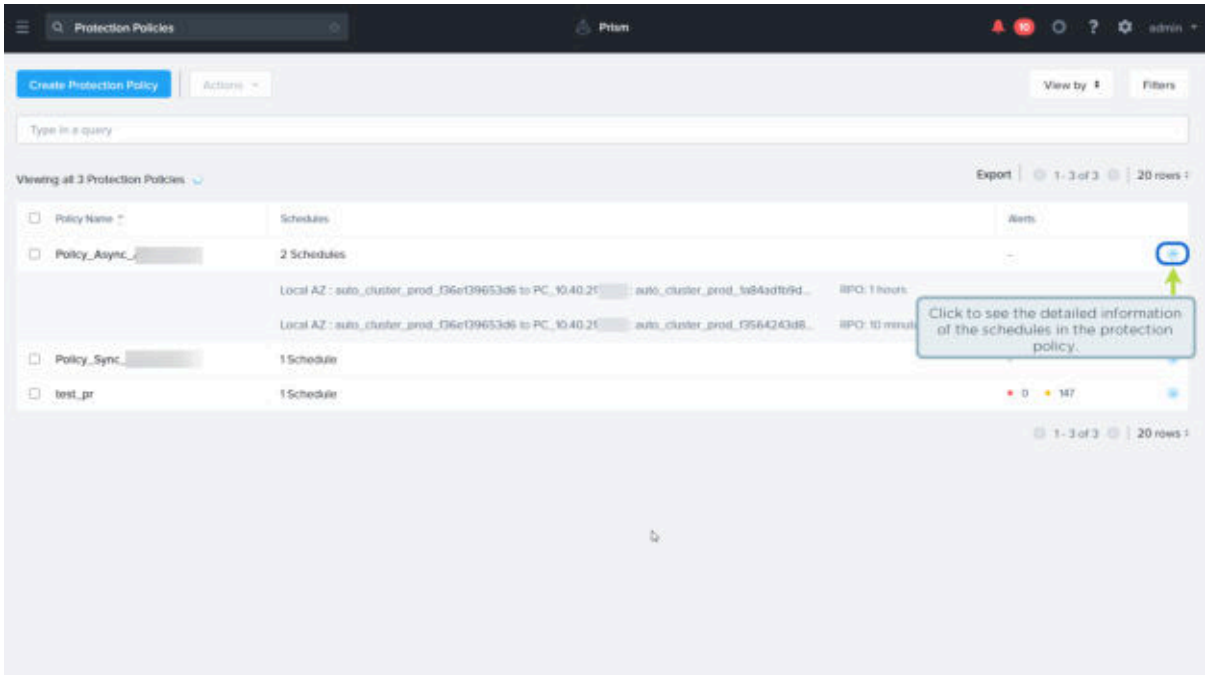


Figure 14: Protection Policies View

Table 11: Protection Policies View Fields

Field	Description
Policy Name	Name of the protection policy.
Schedules	Number of schedules configured in the protection policy. If the protection policy has multiple schedules, a drop-down icon is displayed. Click the drop-down icon to see the primary location:primary Nutanix cluster, recovery location:recovery Nutanix cluster, and RPO of the schedules in the protection policy.
Alerts	Number of alerts issued for the protection policy.

Table 12: Workflows Available in the Protection Policies View

Workflow	Description
Create protection policy	Create a protection policy.

Table 13: Actions Available in the Actions Menu

Action	Description
Update	Update the protection policy.

Action	Description
Clone	Clone the protection policy.
Delete	Delete the protection policy.

Recovery Plans View

The **Recovery Plans** view under the **hamburger icon** > **Data Protection** lists all of configured recovery plans from all the paired availability zones.

The following figure is a sample view, and the tables describe the fields and the actions that you can perform in this view.

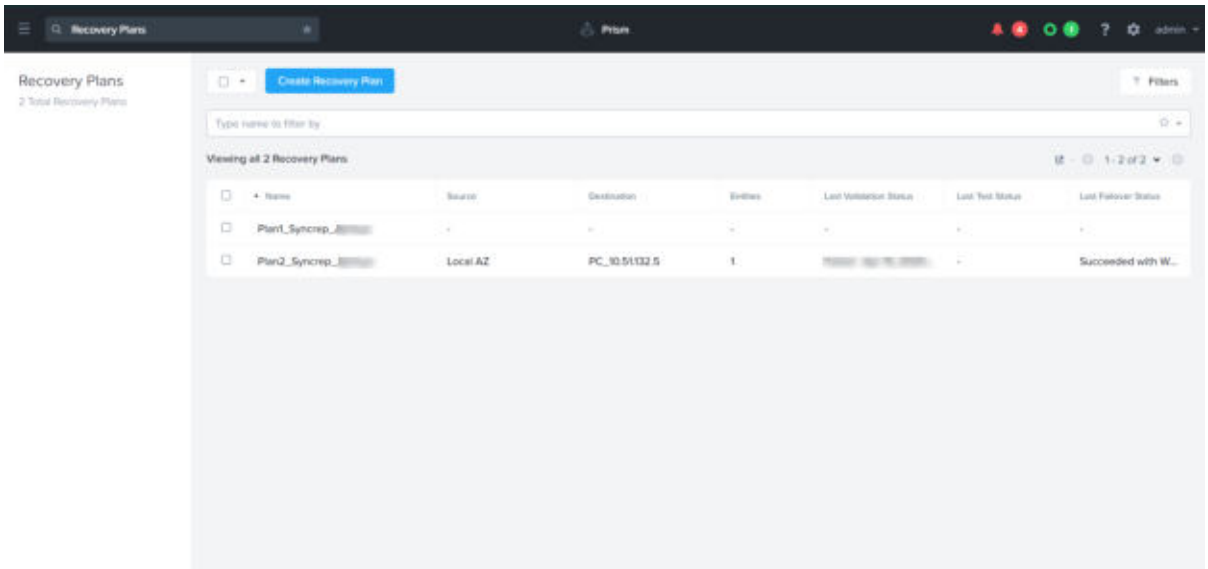


Figure 15: Recovery Plans View

Table 14: Recovery Plans View Fields

Field	Description
Name	Name of the recovery plan.
Primary Location	Replication source site for the recovery plan.
Recovery Location	Replication target site for the recovery plan.
Entities	Sum of the following VMs: • Number of local, live VMs that are specified in the recovery plan. • Number of remote VMs that the recovery plan can recover at this site.
Last Validation Status	Status of the most recent validation of the recovery plan.
Last Test Status	Status of the most recent test performed on the recovery plan.

Field	Description
Last Failover Status	Status of the most recent failover performed on the recovery plan.

Table 15: Workflows Available in the Recovery Plans View

Workflow	Description
Create Recovery Plan	Create a recovery plan.

Table 16: Actions Available in the Actions Menu

Action	Description
Validate	Validates the recovery plan to ensure that the VMs in the recovery plan have a valid configuration and can be recovered.
Test	Tests the recovery plan.
Clean-up test VMs	Cleans up the VMs failed over as a result of testing recovery plan.
Update	Updates the recovery plan.
Failover	Performs a failover.
Delete	Deletes the recovery plan.

VM Recovery Points

The **VM Recovery Points** view under the **hamburger icon** > **Data Protection** lists all the recovery points of all the protected guest VMs (generated over time).

The following figure is a sample view, and the tables describe the fields and the actions that you can perform in this view.

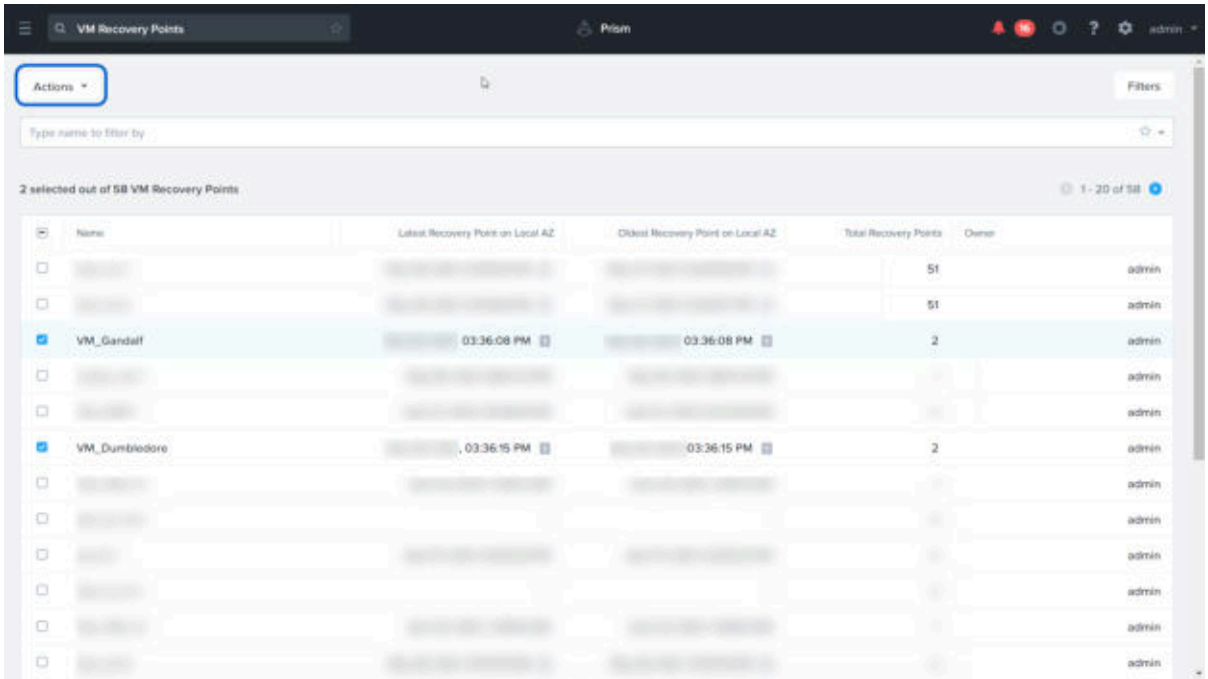


Figure 16: VM Recovery Points

Table 17: VM Recovery Points View Fields

Field	Description
Name	Name of the recovery point.
Latest Recovery Point on Local AZ	Replication source site for the recovery plan.
Oldest Recovery Point on Local AZ	Replication target site for the recovery plan.
Total Recovery Points	Number of recovery points generated for the guest VM.
Owner	Owner account of the recovery point.

Table 18: Actions Available in the Actions Menu

Action	Description
Restore	Restores the guest VMs from the selected recovery points. This option manually recovers guest VMs from recovery points. For more information, see Manual Recovery of Guest VMs on page 144.
Replicate	Manually replicates the selected recovery points to a different Nutanix cluster in the same or different availability zones. For more information, see Replicating Recovery Points Manually on page 144.

Dashboard Widgets

The dashboard includes widgets that display the statuses of configured protection policies and recovery plans. If you have not configured these VMs, the widgets display a summary of the steps required to get started with Leap.

To view these widgets, click the **Dashboard** tab.

The following figure is a sample view of the dashboard widgets.

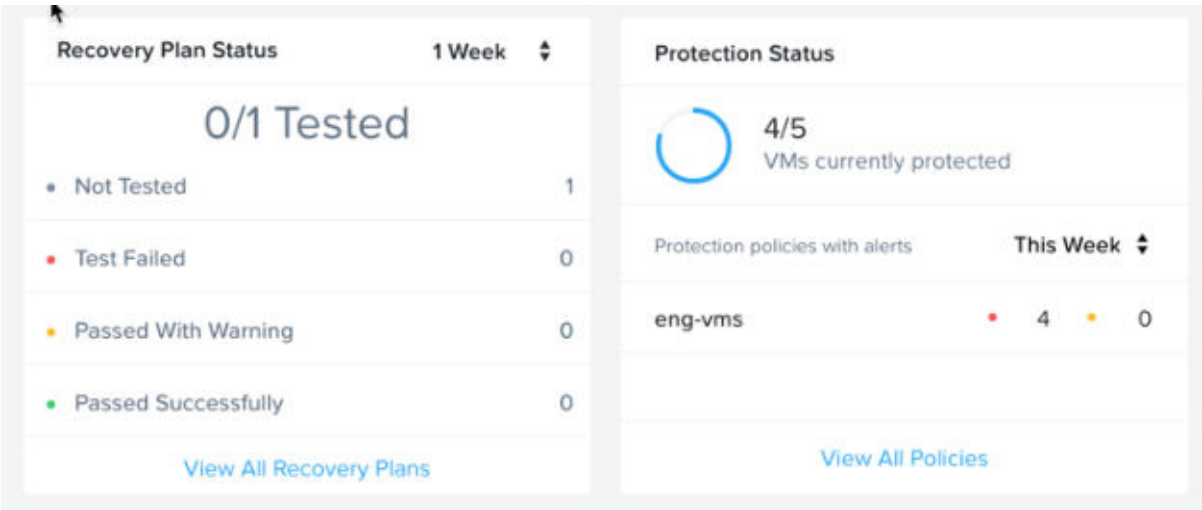


Figure 17: Dashboard Widgets for Leap

Enabling Leap for On-Prem Site

To perform disaster recovery (DR) to Nutanix clusters at different on-prem available zones (sites), enable Leap at both the primary and recovery sites (Prism Central). Without enabling Leap, you can configure protection policies and recovery plans that synchronize to the paired sites but you cannot perform failover and failback operations. To perform DR to different Nutanix clusters at the same site, enable Leap in the single Prism Central.

About this task

To enable Leap, perform the following procedure.

> Note: You cannot disable Leap once you have enabled it.

Procedure

1. Log on to the Prism Central web console.
2. Click the settings button (gear icon) at the top-right corner of the window.

3. Click **Enable Leap** in the **Setup** section on the left pane.

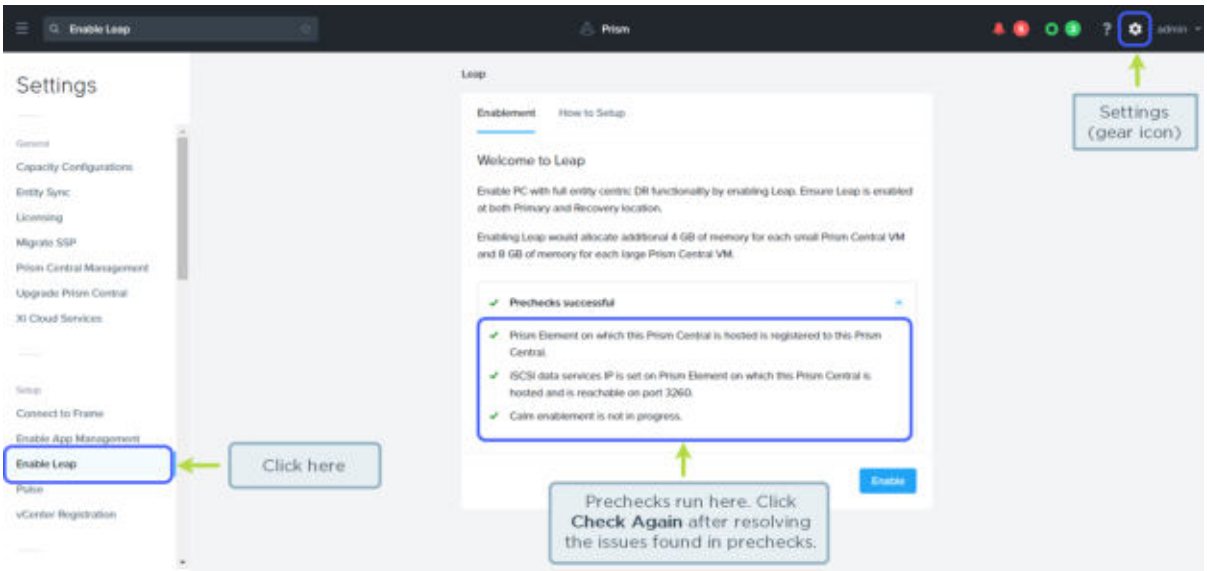


Figure 18: Enabling Leap

The **Leap** dialog box run prechecks. If any precheck fails, resolve the issue that is causing the failure and click **check again**.

4. Click **Enable** after all the prechecks pass.
   Leap is enabled after at least 10 seconds.

Pairing Availability Zones (Leap)

To replicate entities (protection policies, recovery plans, and recovery points) to different on-prem availability zones (sites) bidirectionally, pair the sites with each other. To replicate entities to different Nutanix clusters at the same site bidirectionally, you need not pair the sites because the primary and the recovery Nutanix clusters are registered to the same site (Prism Central). Without pairing the sites, you cannot perform DR to a different site.

About this task

To pair an on-prem site with another on-prem site, perform the following procedure at both the sites.

Procedure

1. Log on to the Prism Central web console.

2. Click the hamburger icon at the top-left corner of the window. Go to **Administration** > **Availability Zones** in the left pane.

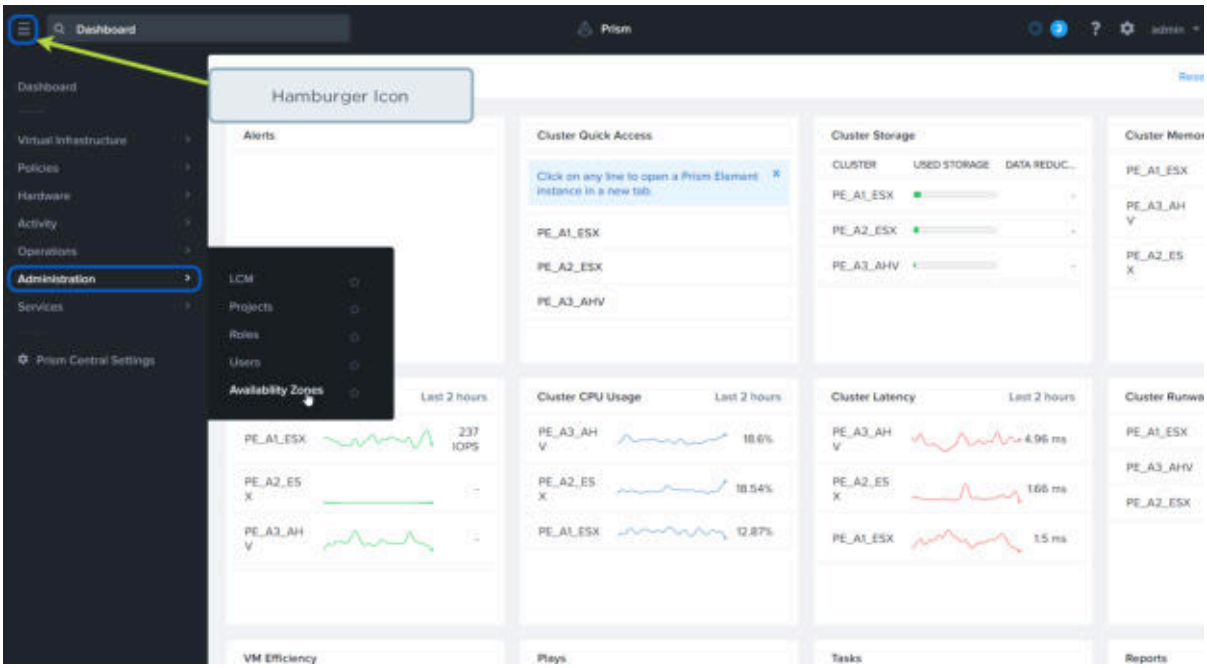


Figure 19: Pairing Availability Zone

3. Click **Connect to Availability Zone**.
   Specify the following information in the Connect to Availability Zone window.

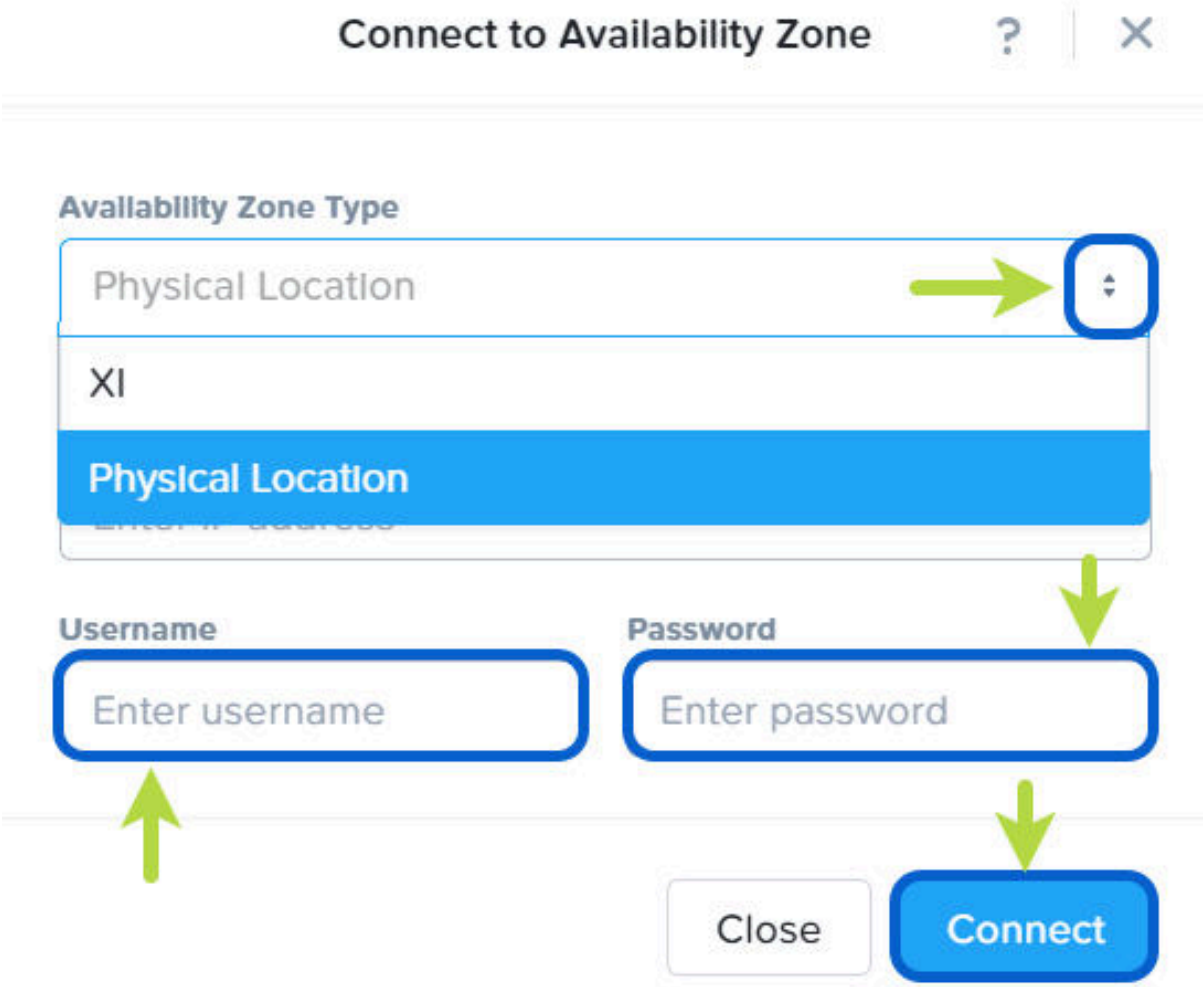


Figure 20: Connect to Availability Zone

   a. **Availability Zone Type**: Select **Physical Location** from the drop-down list.
      A physical location is an on-prem availability zone (site). To pair the on-prem site with Xi Cloud Services, select **XI** from the drop-down list, and enter the credentials of your Xi Cloud Services account in step c and set d.
   b. **IP Address for Remote PC**: Enter the IP address of the recovery site Prism Central.
   c. **Username**: Enter the username of your recovery site Prism Central.
   d. **Password**: Enter the password of your recovery site Prism Central.

4. Click **Connect**.

Protection and Automated DR (Leap)

Automated disaster recovery (DR) configurations use protection policies to protect your guest VMs, and recovery plans to orchestrate the recovery of those guest VMs to different Nutanix clusters at the same or different availability zones (sites). You can automate protection of your guest VMs with the following supported replication schedules in Leap.

- Asynchronous replication schedule (1 hour or greater RPO). For information about protection with Asynchronous replication schedule, see Protection with Asynchronous Replication Schedule and DR (Leap) on page 40.
- NearSync replication schedule (1–15 minute RPO). For information about protection with NearSync replication schedule, see Protection with NearSync Replication Schedule and DR (Leap) on page 100.
- Synchronous replication schedule (0 RPO). For information about protection with Synchronous replication schedule, see Protection with Synchronous Replication Schedule (0 RPO) and DR on page 116.

  To maintain the efficiency in protection and DR, Leap allows to protect a guest VM with Synchronous replication schedule to only one AHV cluster and at the different on-prem availability zone.

Protection with Asynchronous Replication Schedule and DR (Leap)

Asynchronous replication schedules enable you to protect your guest VMs with an RPO of 1 hour or beyond. A protection policy with an Asynchronous replication schedule creates a recovery point in an hourly time interval, and replicates it to the recovery availability zones (sites) for High Availability. For guest VMs protected with Asynchronous replication schedule, you can perform disaster recovery (DR) to different Nutanix clusters at same or different sites. In addition to performing DR to Nutanix clusters running the same hypervisor type, you can also perform cross-hypervisor disaster recovery (CHDR)—DR from AHV clusters to ESXi clusters, or from ESXi clusters to AHV clusters.

> Note: Nutanix provides multiple DR solutions to protect your environment. See Nutanix Disaster Recovery Solutions on page 11 for the detailed representation of the DR offerings of Nutanix.

Asynchronous Replication Requirements (Leap)

The following are the specific requirements for protecting your guest VMs with Asynchronous replication schedule. Ensure that you meet the following requirements in addition to the general requirements of Leap.

For information about the general requirements of Leap, see Leap Requirements on page 20.

For information about node, disk and Foundation configurations required to support Asynchronous replication schedules, see On-Prem Hardware Resource Requirements on page 15.

Hypervisor Requirements

AHV or ESXi

- The AHV clusters must be running on AHV versions that come bundled with the supported version of AOS.
- The ESXi clusters must be running on version ESXi 6.5 GA or newer.

Nutanix Software Requirements

Each on-prem site must have a Leap enabled Prism Central instance.

The primary and recovery Prism Central and Prism Element on the Nutanix clusters must be running the following versions of AOS.

- AHV clusters
  - AOS 5.17 or newer for DR to different Nutanix clusters at the same site.
  - AOS 5.10 or newer for DR to Nutanix clusters at the different sites.
- ESXi clusters
  - AOS 5.17 or newer for DR to different Nutanix clusters at the same site.
  - AOS 5.11 or newer for DR to Nutanix clusters at the different sites.

Cross Hypervisor Disaster Recovery (CHDR) Requirements

Guest VMs protected with Asynchronous replication schedule support cross-hypervisor disaster recovery. You can perform failover (DR) to recover guest VMs from AHV clusters to ESXi clusters or guest VMs from ESXi clusters to AHV clusters by considering the following requirements.

- Both the primary and the recovery Nutanix clusters must be running AOS 5.17 or newer for CHDR to Nutanix clusters at the same availability zone (site).
- Both the primary and the recovery Nutanix clusters must be running AOS 5.11.2 or newer for CHDR to Nutanix clusters at different availability zones (sites).
- Install and configure Nutanix Guest Tools (NGT) on all the guest VMs. For more information, see *Enabling and Mounting Nutanix Guest Tools* in *Prism Web Console Guide*.

  NGT configures the guest VMs with all the required drivers for VM portability. For more information about general NGT requirements, see *Nutanix Guest Tools Requirements and Limitations* in *Prism Web Console Guide*.
- CHDR supports guest VMs with flat files only.
- CHDR supports IDE/SCSI and SATA disks only.
- For all the non-boot SCSI disks of Windows guest VMs, set the SAN policy to OnlineAll so that they come online automatically.
- In vSphere 6.7, guest VMs are configured with UEFI secure boot by default. Upon CHDR to an AHV cluster, these guest VMs do not start if the host does not support the UEFI secure boot feature. For more information about supportability of UEFI secure boot on Nutanix clusters, see https://portal.nutanix.com/page/documents/compatibility-interoperability-matrix/guestos.
- For information about operating systems that support UEFI and Secure Boot, see UEFI and Secure Boot Support for CHDR on page 227.
- Nutanix does not support vSphere inventory mapping (for example, VM folder and resource pools) when protecting workloads between VMware clusters.

- Nutanix does not support vSphere snapshots or delta disk files.

  If you have delta disks attached to a guest VM and you proceed with failover, you get a validation warning and the guest VM does not recover. Contact Nutanix Support for assistance.

Table 19: Operating Systems Supported for CHDR (Asynchronous Replication)

Operating System	Version	Requirements and limitations
Windows	• Windows 2008 R2 or newer versions • Windows 7 or newer versions	• Only 64-bit operating systems are supported.
Linux	• CentOS 6.5 and 7.0 • RHEL 6.5 or newer and RHEL 7.0 or newer. • Oracle Linux 6.5 and 7.0 • Ubuntu 14.04	• SLES operating system is not supported.

Additional Requirement

The storage container name of the protected guest VMs must be the same on both the primary and recovery clusters. Therefore, a storage container must exist on the recovery cluster with the same name as the one on the primary cluster. For example, if the protected VMs are in the *SelfServiceContainer* storage container on the primary cluster, there must also be a *SelfServiceContainer* storage container on the recovery cluster.

Asynchronous Replication Limitations (Leap)

Consider the following specific limitations before protecting your guest VMs with Asynchronous replication schedule. These limitations are in addition to the general limitations of Leap.

For information about the general limitations of Leap, see Leap Limitations on page 24.

- You cannot restore guest VMs with incompatible GPUs at the recovery Nutanix cluster.
- You cannot protect guest VMs configured as part of a network function chain.
- You cannot retain hypervisor-specific properties after cross hypervisor disaster recovery (CHDR).

  CHDR does not preserve hypervisor-specific properties (for example, multi-writer flags, independent persistent and non-persistent disks, changed block tracking (CBT), PVSCSI disk configurations).

Creating a Protection Policy with an Asynchronous Replication Schedule (Leap)

To protect the guest VMs in an hourly replication schedule, configure an Asynchronous replication schedule while creating the protection policy. The policy takes recovery points of those guest VMs in the specified time intervals (hourly) and replicates them to the recovery availability zones (sites) for High Availability. To protect the guest VMs at the same or different recovery sites, the protection policy allows you to configure Asynchronous replication

schedules to at most two recovery sites—a unique replication schedule to each recovery site. The policy synchronizes continuously to the recovery sites in a bidirectional way.

Before you begin

See Asynchronous Replication Requirements (Leap) on page 40 and Asynchronous Replication Limitations (Leap) on page 42 before you start.

About this task

To create a protection policy with an Asynchronous replication schedule, do the following at the primary site. You can also create a protection policy at the recovery site. Protection policies you create or update at a recovery site synchronize back to the primary site.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Protection Policies** in the left pane.

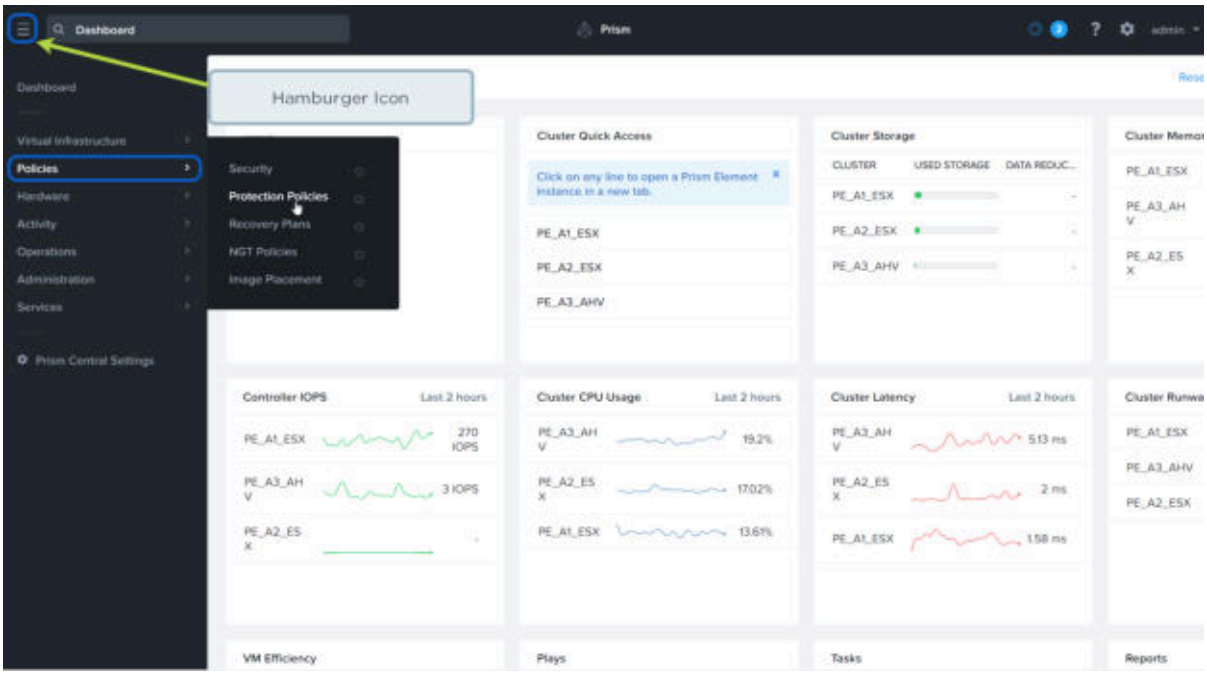


Figure 21: Protection Policy Configuration: Protection Policies

3. Click **Create Protection Policy**.
   Specify the following information in the Create Protection Policy window.

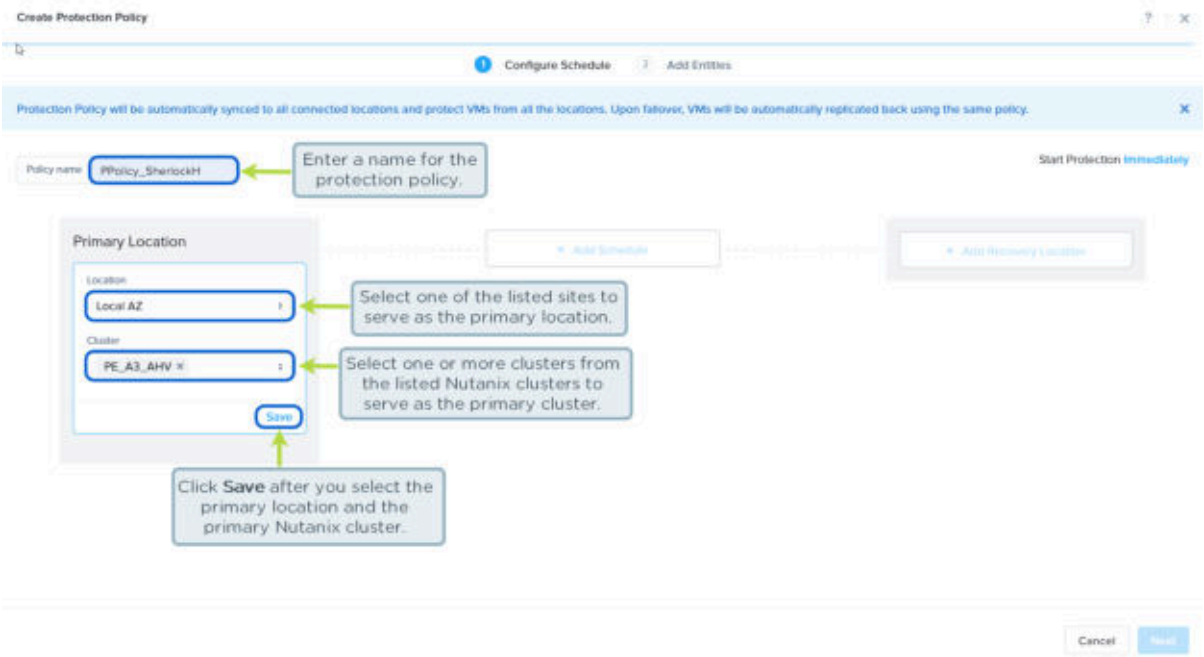


Figure 22: Protection Policy Configuration: Select Primary Location

a. **Policy name**: Enter a name for the protection policy.

> Caution: The name can be of only alphanumeric, dot, dash, and underscore characters.

b. In the **Primary Location** pane, specify the following information.

- 1. **Location**: From the drop-down list, check an availability zone (site) that hosts the guests VMs to protect.

     The drop-down lists all the sites paired with the local site. Local AZ represents the local site (Prism Central). For your primary site, you can check either the local site or a non-local site.
  2. **Cluster**: From the drop-down list, check the Nutanix cluster that hosts the guest VMs to protect.

     The drop-down lists all the Nutanix clusters registered to Prism Central representing the selected site. If you want to protect the guest VMs from multiple Nutanix clusters in the same protection policy, check the clusters that host those guest VMs. All Clusters protects the guest VMs of all Nutanix clusters registered to Prism Central.
  3. Click **Save**.

     Clicking **Save** activates the **Recovery Location** pane. After saving the primary site configuration, you can optionally add a local schedule (step iv) to retain the recovery points at the primary site.
  4. Click **+ Add Local Schedule** if you want to retain recovery points locally in addition to retaining recovery points in a replication schedule (step d.iv). For example, you can create a local schedule to retain 15 minute recovery points locally and also

an hourly replication schedule to retain recovery points and replicate them to a recovery site every 2 hours. The two schedules apply differently on the guest VMs.

Specify the following information in the Add Schedule window.

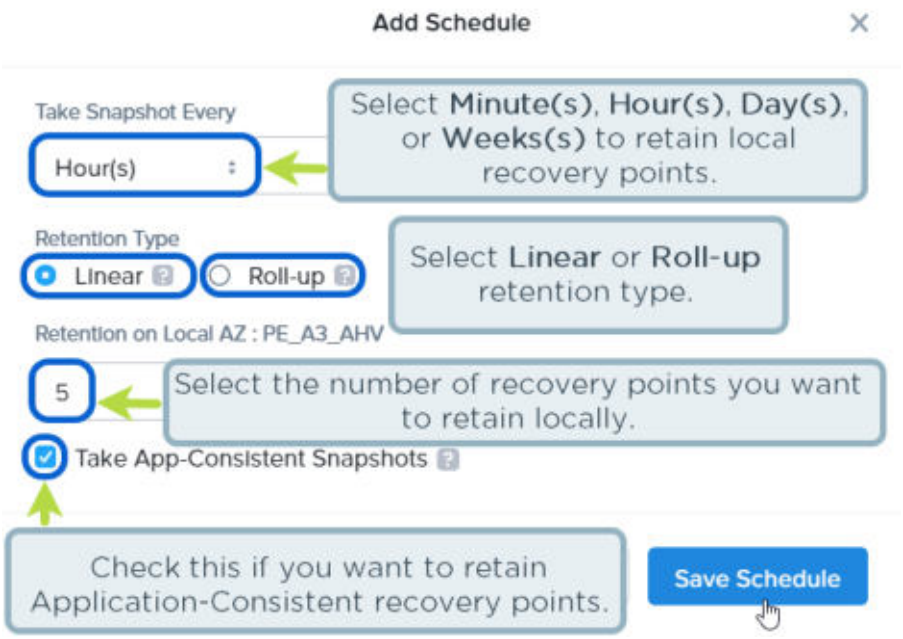


Figure 23: Protection Policy Configuration: Add Local Schedule

1. **Take Snapshot Every**: Specify the frequency in minutes, hours, days, or weeks at which you want the recovery points to be taken locally.
2. **Retention Type**: Specify one of the following two types of retention policy.

   - **Linear**: Implements a simple retention scheme at the local site. If you set the retention number to n, the local site retains the n recent recovery points.

     When you enter the frequency in minutes, the system selects the **Roll-up** retention type by default because minutely recovery points do not support **Linear** retention types.

   - **Roll-up**: Rolls up the recovery points into a single recovery point at the local site.

     For more information about the roll-up recovery points, see step d.iii.
3. **Retention on Local AZ:PE_A3_AHV**: Specify the retention number for the local site.
4. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Irrespective of the local or replication schedules, the recovery points are of the specified type. If you check **Take App-Consistent Recovery Point**, the recovery points generated are application-consistent and if you do not check **Take App-Consistent Recovery Point**, the recovery points generated are crash-consistent.

If the time in the local schedule and the replication schedule match, the single recovery point generated is application-consistent.

> Note: See Application-consistent Recovery Point Conditions and Limitations on page 55 before you take application-consistent snapshot.

5. Click **Save Schedule**.

c. In the **Recovery Location** pane, specify the following information.

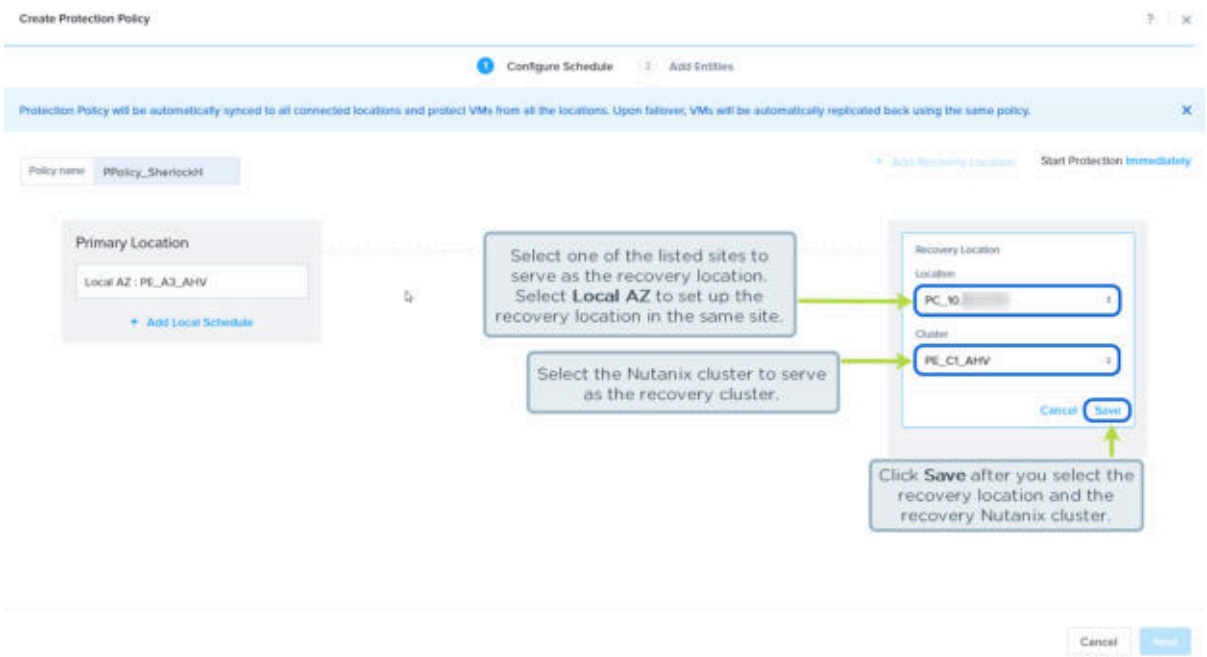


Figure 24: Protection Policy Configuration: Select Recovery Location

- 1. **Location**: From the drop-down list, select the availability zone (site) where you want to replicate the recovery points.

  The drop-down lists all the sites paired with the local site. Local AZ represents the local site (Prism Central). Select Local AZ if you want to configure DR to a different Nutanix cluster at the same site.

  If you do not select a site, local recovery points that are created by the protection policy do not replicate automatically. You can, however, replicate the recovery points manually and use recovery plans to recover the guest VMs. For more information, see Protection and Manual DR (Leap) on page 143.

  2. **Cluster**: From the drop-down list, select the Nutanix cluster where you want to replicate the recovery points.

  The drop-down lists all the Nutanix clusters registered to Prism Central representing the selected site. You can select one cluster at the recovery site. If you want to replicate the recovery points to more clusters at the same or different sites, add

another recovery site with a replication schedule. For more information to add another recovery site with a replication schedule, see step e.

> Note: Selecting auto-select from the drop-down list replicates the recovery points to any available cluster at the recovery site. Select auto-select from the drop-down list only if all the clusters at the recovery site are up and running.

> Caution: If the primary Nutanix cluster contains an IBM POWER Systems server, you can replicate recovery points to an on-prem site only if that on-prem site contains an IBM Power Systems server.

3. Click **Save**.

   Clicking **Save** activates the **+ Add Schedule** button between the primary and the recovery site. After saving the recovery site configuration, you can optionally add a local schedule to retain the recovery points at the recovery site.

4. Click **+ Add Local Schedule** if you want to retain recovery points locally in addition to retaining recovery points in a replication schedule (step d.iv). For example, you can create a local schedule to retain one hourly recovery points locally to supplement the hourly replication schedule. The two schedules apply differently on the guest VMs after failover, when the recovery points replicate back to the primary site.

   Specify the following information in the Add Schedule window.

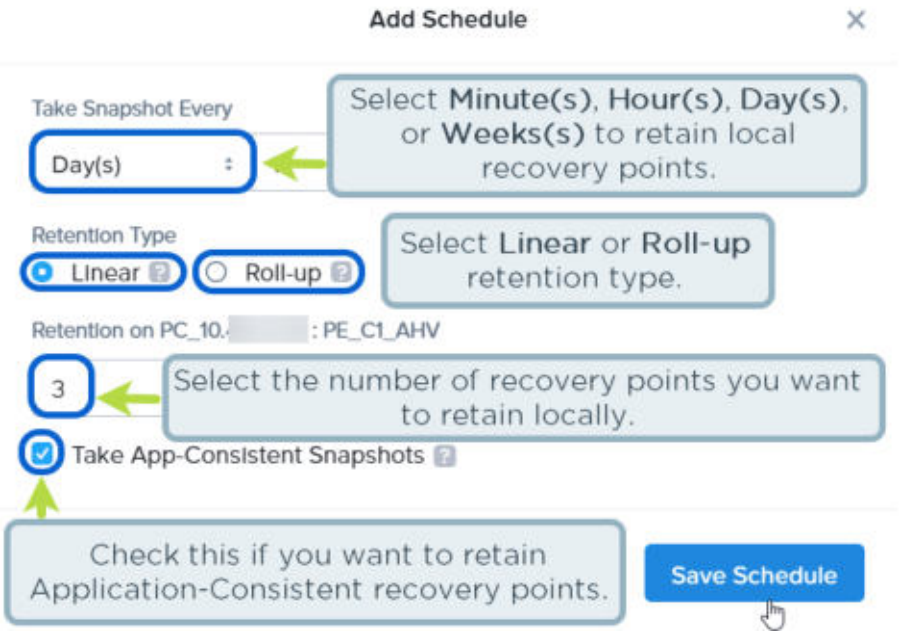


Figure 25: Protection Policy Configuration: Add Local Schedule

1. **Take Snapshot Every**: Specify the frequency in minutes, hours, days, or weeks at which you want the recovery points to be taken locally.

2. **Retention Type**: Specify one of the following two types of retention policy.

   - **Linear**: Implements a simple retention scheme at the local site. If you set the retention number to n, the local site retains the n recent recovery points.

     When you enter the frequency in minutes, the system selects the **Roll-up** retention type by default because minutely recovery points do not support **Linear** retention types.

   - **Roll-up**: Rolls up the recovery points into a single recovery point at the local site.

     For more information about the roll-up recovery points, see step d.iii.

3. **Retention on 10.xx.xx.xxx:PE_C1_AHV**: Specify the retention number for the local site.
4. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Irrespective of the local or replication schedules, the recovery points are of the specified type. If you check **Take App-Consistent Recovery Point**, the recovery points generated are application-consistent and if you do not check **Take App-Consistent Recovery Point**, the recovery points generated are crash-consistent.

If the time in the local schedule and the replication schedule match, the single recovery point generated is application-consistent.

> Note: See Application-consistent Recovery Point Conditions and Limitations on page 55 before you take application-consistent snapshot.

5. Click **Save Schedule**.

d. Click **+ Add Schedule** to add a replication schedule between the primary and the recovery site.

   Specify the following information in the Add Schedule window. The window auto-populates the Primary Location and Recovery Location that you have selected in step b and step c.

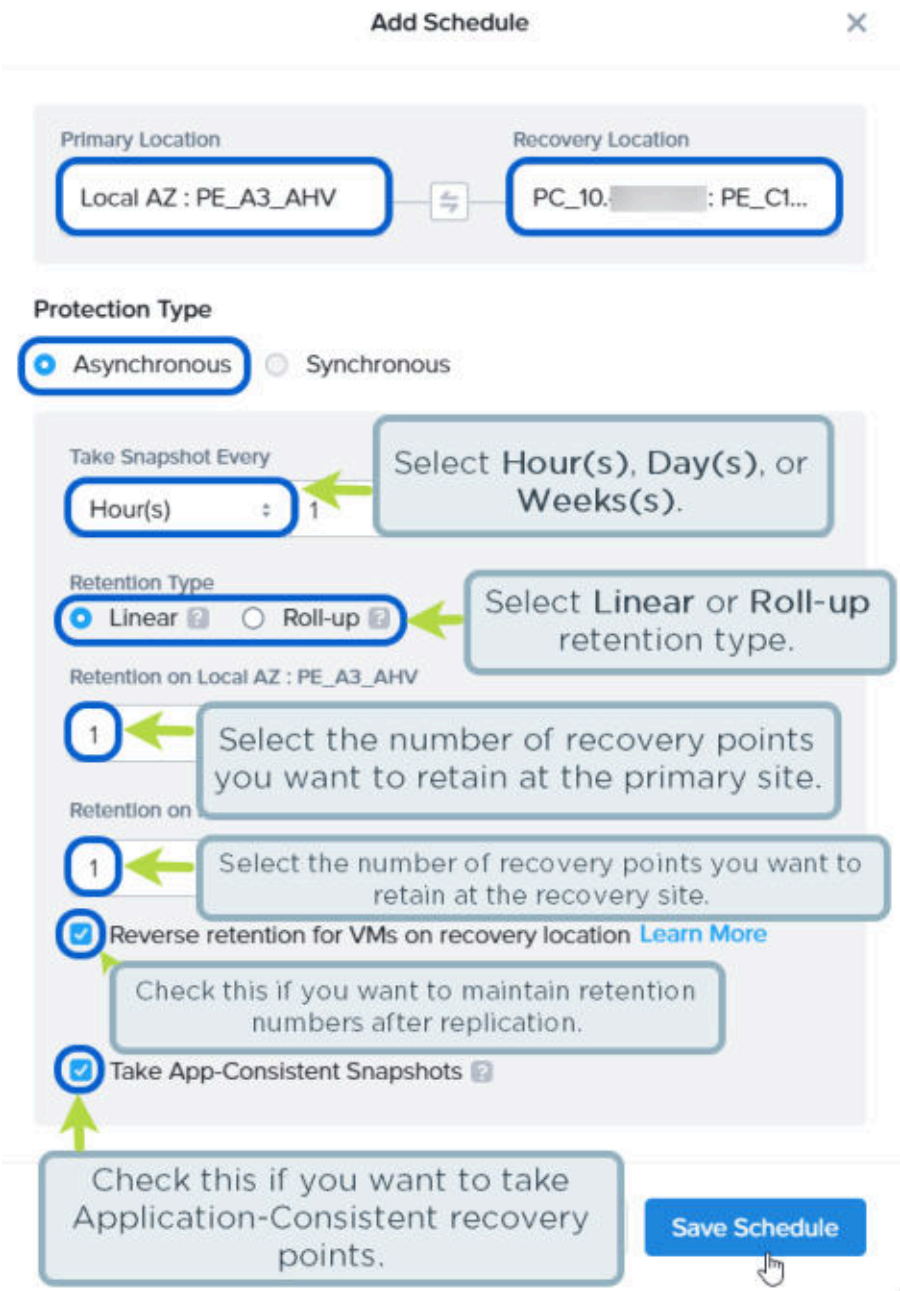


Figure 26: Protection Policy Configuration: Add Schedule (Asynchronous)

- 1. **Protection Type**: Click **Asynchronous**.
  2. **Take Snapshot Every**: Specify the frequency in hours, days, or weeks at which you want the recovery points to be taken.

     The specified frequency is the RPO. For more information about RPO, see Leap Terminology on page 8.
  3. **Retention Type**: Specify one of the following two types of retention policy.

     - **Linear**: Implements a simple retention scheme at both the primary (local) and the recovery (remote) site. If you set the retention number for a given site to n, that

site retains the n recent recovery points. For example, if the RPO is 1 hour, and the retention number for the local site is 48, the local site retains 48 hours (48 X 1 hour) of recovery points at any given time.

> Tip: Use linear retention policies for small RPO windows with shorter retention periods or in cases where you always want to recover to a specific RPO window.

- **Roll-up**: Rolls up the recovery points as per the RPO and retention period into a single recovery point at a site. For example, if you set the RPO to 1 hour, and the retention time to 5 days, the 24 oldest hourly recovery points roll up into a single daily recovery point (one recovery point = 24 hourly recovery points) after every 24 hours. The system keeps one day (of rolled-up hourly recovery points) and 4 days of daily recovery points.

> Note:
>
> - If the retention period is n days, the system keeps 1 day of RPO (rolled-up hourly recovery points) and n-1 days of daily recovery points.
> - If the retention period is n weeks, the system keeps 1 day of RPO, 1 week of daily and n-1 weeks of weekly recovery points.
> - If the retention period is n months, the system keeps 1 day of RPO, 1 week of daily, 1 month of weekly, and n-1 months of monthly recovery points.
> - If the retention period is n years, the system keeps 1 day of RPO, 1 week of daily, 1 month of weekly, and n-1 months of monthly recovery points.

> Note: The recovery points that are used to create a rolled-up recovery point are discarded.

> Tip: Use roll-up retention policies for anything with a longer retention period. Roll-up policies are more flexible and automatically handle recovery point aging/pruning while still providing granular RPOs for the first day.

4. To specify the retention number for the primary and recovery sites, do the following.

   - **Retention on Local AZ: PE_A3_AHV**: Specify the retention number for the primary site.

     This field is unavailable if you do not specify a recovery location.

   - **Retention on 10.xx.xx.xxx:PE_C1_AHV**: Specify the retention number for the recovery site.

     If you select linear retention, the remote and local retention count represents the number of recovery points to retain at any given time. If you select roll-up retention, these numbers specify the retention period.

5. If you want to enable reverse retention of the recovery points, check **Reverse retention for VMs on recovery location**.

   > Note: **Reverse retention for VMs on recovery location** is available only when the retention numbers on the primary and recovery sites are different.

   Reverse retention maintains the retention numbers of recovery points even after failover to a recovery site in the same or different availability zones. For example, if

you retain two recovery points at the primary site and three recovery points at the recovery site, and you enable reverse retention, a failover event does not change the initial retention numbers when the recovery points replicate back to the primary site. The recovery site still retains two recovery points while the primary site retains three recovery points. If you do not enable reverse retention, a failover event changes the initial retention numbers when the recovery points replicate back to the primary site. The recovery site retains three recovery points while the primary site retains two recovery points.

Maintaining the same retention numbers at a recovery site is required if you want to retain a particular number of recovery points, irrespective of where the guest VM is after its failover.

6. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Application-consistent recovery points ensure that application consistency is maintained in the replicated recovery points. For application-consistent recovery points, install NGT on the guest VMs running on AHV clusters. For guest VMs running on ESXi clusters, you can take application-consistent recovery points without installing NGT, but the recovery points are hypervisor-based, and leads to VM stuns (temporary unresponsive VMs) after failover to the recovery sites.

   > Note: See Application-consistent Recovery Point Conditions and Limitations on page 55 before you take application-consistent snapshot.

   > Caution: Application-consistent recovery points fail for EFI-boot enabled Windows 2019 VM running on ESXi when NGT is not installed. Nutanix recommends installing NGT on guest VMs running on ESXi also.

7. Click **Save Schedule**.

e. Click **+ Add Recovery Location** at the top-right if you want to add an additional recovery site for the guest VMs in the protection policy.

   » To add an on-prem site for recovery, see Protection and DR between On-Prem Sites (Leap) on page 18

   » To add Xi Cloud Services for recovery, see Protection and DR between On-Prem Site and Xi Cloud Service (Xi Leap) on page 154.

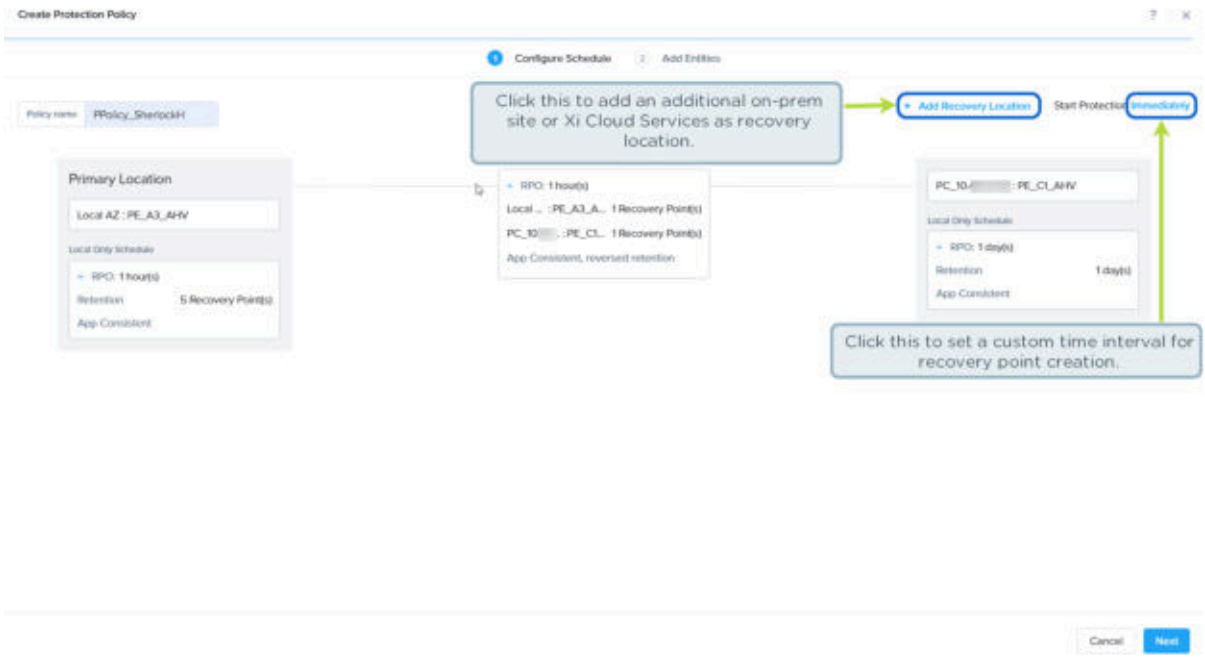


Figure 27: Protection Policy Configuration: Additional Recovery Location

f. Click **+ Add Schedule** to add a replication schedule between the primary site and the additional recovery site you specified in step e.

   Perform step d again in the Add Schedule window to add the replication schedule. The window auto-populates the Primary Location and the additional Recovery Location that you have selected in step b and step c.

   By default, recovery point creation begins immediately after you create the protection policy. If you want to specify when recovery point creation must begin, click **Immediately** at the top-right corner, and then, in the **Start Time** dialog box, do the following.

   - 1. Click **Start protection at specific point in time**.
     2. Specify the time at which you want to start taking recovery points.
     3. Click **Save**.

g. Click **Next**.

   Clicking **Next** shows a list of VM categories where you can optionally check one or more VM categories to protect in the protection policy. DR configurations using Leap allows you to protect a guest VM by using only one protection policy. Therefore, VM categories specified in another protection policy are not in the list. If you protect a guest VM in another protection policy by specifying the VM category of the guest VM (category-based inclusion), and if you protect the guest VM from the **VMs** page in this policy (individual inclusion), the individual inclusion supersedes the category-based inclusion.

Effectively, only the protection policy that protected the individual guest VM protects the guest VM.

For example, the guest VM *VM_SherlockH* is in the category *Department:Admin*, and you add this category to the protection policy named *PP_AdminVMs*. Now, if you add *VM_SherlockH* from the **VMs** page to another protection policy named *PP_VMs_UK*, *VM_SherlockH* is protected in *PP_VMs_UK* and unprotected from *PP_AdminVMs*.

h. If you want to protect the guest VMs category wise, check the VM categories that you want to protect from the list and click **Add**.

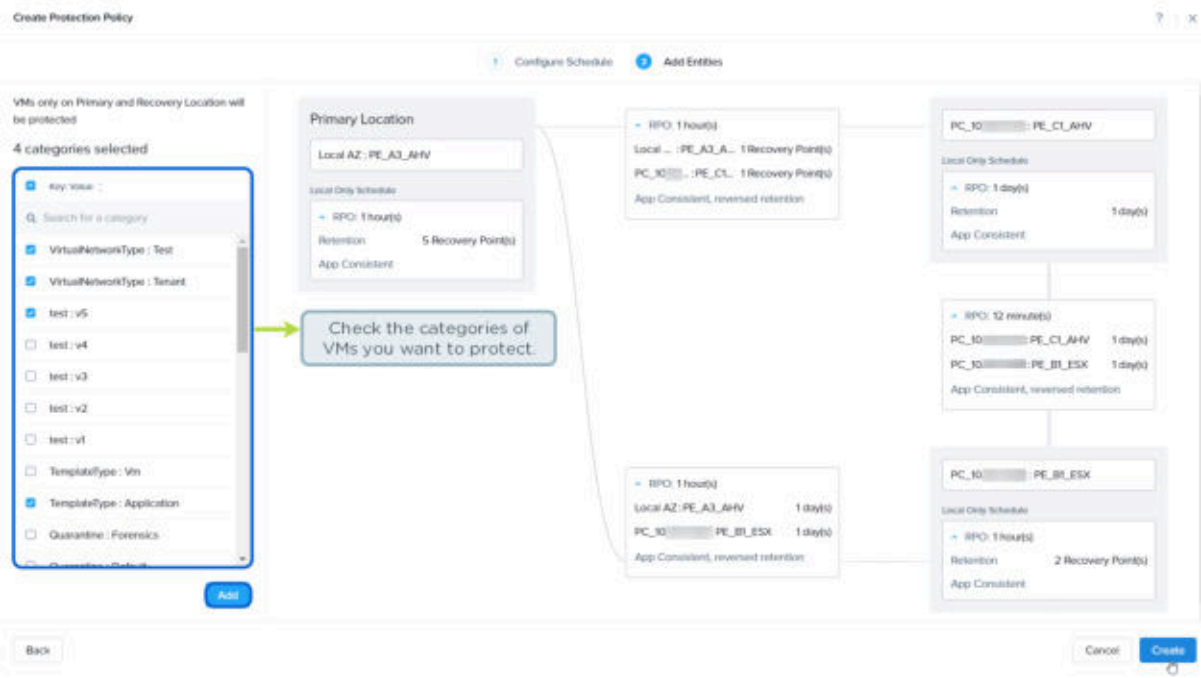


Figure 28: Protection Policy Configuration: Add VM Categories

Prism Central includes built-in VM categories for frequently encountered applications (for example, MS Exchange and Oracle). If the VM category or value you want is not available, first create the category with the required values, or update an existing category so that it has the values you require. Doing so ensures that the VM categories and values are available for selection. You can add VMs to the category either before or after you configure the protection policy. If the guest VMs have a common characteristic, such as belonging to a specific application or location, create a VM category and add the guest VMs into the category.

If you do not want to protect the guest VMs category wise, proceed to the next step without checking VM categories. You can add the guest VMs individually to the protection policy later from the **VMs** page (see Adding Guest VMs Individually to a Protection Policy on page 136).

i. Click **Create**.
The protection policy with an Asynchronous replication schedule is created. To verify the protection policy, see the **Protection Policies** page. If you check VM categories in step h, the protection policy starts generating recovery points of the guest VMs in those VM categories. To see the generated recovery points, click the hamburger icon at the top-left corner of the window and go to **VM Recovery Points**. Click the recovery points for its

information. You can see the time estimated for the very first replication (seeding) to the recovery sites.

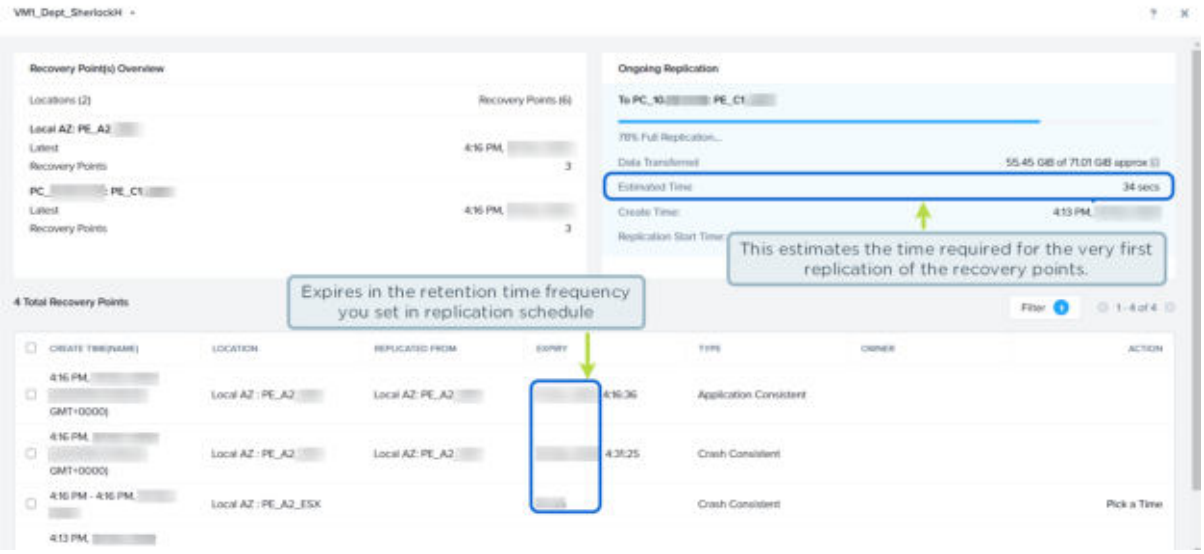


Figure 29: Recovery Points Overview

Application-consistent Recovery Point Conditions and Limitations

This topic describes the conditions and limitations for application-consistent recovery points that you can generate through a protection policy. For information about the operating systems that support the AOS version you have deployed, see the Compatibility Matrix.

- Before taking an application-consistent recovery point, consider the workload type of your guest VM.

  Applications running in your guest VM must be able to quiesce I/O operations. For example, For example, you can quiesce I/O operations for database applications and similar workload types.

- Before taking an application-consistent recovery point, install and enable Nutanix Guest Tools (NGT) on your guest VM.

  For installing and enabling NGT, see *Nutanix Guest Tools* in the *Prism Web Console Guide*.

  For guest VMs running on ESXi, consider these points.

- Install and enable NGT on guest VMs running on ESXi also. Application-consistent recovery points fail for EFI boot-enabled Windows 2019 VMs running on ESXi without installing NGT.

- (vSphere) If you do not enable NGT and then try to take an application-consistent recovery point, the system creates a Nutanix native recovery point with a single vSphere host-based recovery point. The system deletes the vSphere host-based recovery point. If you enable NGT and then try to take application-consistent recovery point, the system directly captures a Nutanix native recovery point.

- Do not delete the `.snapshot` folder in the vCenter.

- The following table lists the operating systems that support application-consistent recovery points with NGT installed.

Table 20: Supported Operating Systems (NGT Installed)

Operating system	Version
Windows	• Windows 2008 R2 through Windows 2019
Linux	• CentOS 6.5 through 6.9 and 7.0 through 7.3 • Red Hat Enterprise Linux (RHEL) 6.5 through 6.9 and 7.0 through 7.3. • Oracle Linux 6.5 and 7.0 • SUSE Linux Enterprise Server (SLES) 11 SP1 through 11 SP4 and 12 SP1 through 12 SP3 • Ubuntu 14.04

Application-consistent Recovery Points with Microsoft Volume Shadow Copy Service (VSS)

- To take application-consistent recovery points on Windows guest VMs, enable Microsoft VSS services.

  When you configure a protection policy and select **Take App-Consistent Recovery Point**, the Nutanix cluster transparently invokes the VSS (also known as Shadow copy or volume snapshot service).

  > Note: This option is available for ESXi and AHV only. However, you can use third-party backup products to invoke VSS for Hyper-V.

- To take application-consistent recovery points on guest VMs that use VSS, systems invoke Nutanix native in-guest VmQuiesced Snapshot Service (VSS) agent. VSS framework takes application-consistent recovery points without causing VM stuns (temporary unresponsive VMs).
- VSS framework enables third-party backup providers like Commvault and Rubrik to take application-consistent snapshots on Nutanix platform in a hypervisor-agnostic manner.
- The default and only backup type for VSS snapshots is VSS_BT_COPY (copy backup).

  Third party Backup products can choose between VSS_BT_FULL (full backup)and VSS_BT_COPY (copy backup) backup types.
- Guest VMs with any iSCSI attachments (LUNs) do not support Nutanix VSS recovery points.

  Nutanix VSS recovery points fail for such guest VMs.
- Do not take Nutanix enabled application-consistent recovery points while using any third-party backup provider enabled VSS snapshots (for example, Veeam).

Pre-freeze and Post-thaw scripts

- You can take application-consistent recovery points on NGT and Volume Shadow Copy Service (VSS) enabled guest VMs. However, some applications require more steps before or after the VSS operations to fully quiesce the guest VMs to an appropriate restore point or state in which the system can capture a recovery point. Such applications need pre-freeze and post-thaw scripts to run the necessary extra steps.

- Any operation that the system must perform on a guest VM before replication or a recovery point capture is a pre-freeze operation. For example, if a guest VM hosts a database, you can enable hot backup of the database before replication using a pre-freeze script. Similarly, any operation that the system must perform on guest VM after replication or a recovery point capture is a post-thaw operation.

  > Tip: Vendors such as *CommVault* provide pre-freeze and post-thaw scripts. You can also write your own pre-freeze and post-thaw scripts.

  Script Requirements

  - For Windows VMs, you must administrator and have read, write, and execute permissions on the scripts.
  - For Linux VMs, you must have root ownership and root access with 700 permissions on the scripts.
  - For completion of any operation before or after replication or recovery point capture, you must have both the pre_freeze and post_thaw scripts for the operation.
  - Timeout for both the scripts is 60 seconds.
  - A script must return 0 to indicate a successful run. A non-zero return value implies that the script execution failed. The necessary log entries are available in the NGT logs.

    > Tip: (AHV) For a non-zero return value from the pre-freeze script, the system captures a non application-consistent snapshot and raises an alert on the Prism web console. Similarly, for a non-zero return value from the post-thaw script, the system attempts to capture an application-consistent snapshot once again. If the attempt fails, the system captures a non application-consistent snapshot, and raises an alert on the Prism web console.

  - Irrespective of whether the pre-freeze script execution is successful, the corresponding post-thaw script runs.

  Script Location

  You can define Python or shell scripts or any executable or batch files at the following locations in Linux or Windows VMs. The scripts can contain commands and routines necessary to run specific operations on one or more applications.

  - In Windows VMs,
    - Batch script file path for pre_freeze scripts:

      ```
      C:\Program Files\Nutanix\Scripts\pre_freeze.bat
      ```

    - Batch script file path for post_thaw scripts:

      ```
      C:\Program Files\Nutanix\Scripts\post_thaw.bat
      ```

- In Linux VMs,
  - Shell script file path for production failover:

    `/usr/local/sbin/pre_freeze`

    Replace `pre_freeze` with the script name (without extension).
  - Shell script file path for test failover:

    `/usr/local/sbin/post_thaw`

    Replace `post_thaw` with the script name (without extension).

    > Note: The scripts must have root ownership and root access with 700 permissions.

Script Sample

> Note: The following are only sample scripts and therefore must be modified to fit your deployment.

- For Linux VMs

```
#!/bin/sh
#pre_freeze-script
date >> '/scripts/pre_root.log'
echo -e "\n attempting to run pre_freeze script for MySQL as root user\n" >> /scripts/
pre_root.log
if [ "$(id -u)" -eq "0" ]; then
python '/scripts/quiesce.py' &
echo -e "\n executing query flush tables with read lock to quiesce the database\n"
 >> /scripts/pre_freeze.log
echo -e "\n Database is in quiesce mode now\n" >> /scripts/pre_freeze.log
else
date >> '/scripts/pre_root.log'
echo -e "not root useri\n" >> '/scripts/pre_root.log'
fi
```

```
#!/bin/sh
#post_thaw-script
date >> '/scripts/post_root.log'
echo -e "\n attempting to run post_thaw script for MySQL as root user\n" >> /scripts/
post_root.log
if [ "$(id -u)" -eq "0" ]; then
python '/scripts/unquiesce.py'
else
date >> '/scripts/post_root.log'
echo -e "not root useri\n" >> '/scripts/post_root.log'
fi
```

- For Windows VMs

```
@echo off
echo Running pre_freeze script >C:\Progra~1\Nutanix\script\pre_freeze_log.txt
```

```
@echo off
echo Running post_thaw script >C:\Progra~1\Nutanix\script\post_thaw_log.txt
```

> Note: If any of these scripts prints excessive output to the console session, the script freezes. To avoid script freeze, perform the following.
>
> - Add `@echo off` to your scripts.

- Redirect the script output to a log file.

If you receive a non-zero return code from the pre-freeze script, the system captures a non application-consistent recovery point and raises an alert on the Prism web console. If you receive a non-zero return code from the post-thaw script, the system attempts to capture an application-consistent snapshot once again. If that attempt fails, the system captures a non application-consistent snapshot, and raises an alert on the Prism web console.

Applications supporting application-consistent recovery points without scripts

Only the following applications support application-consistent recovery points without pre-freeze and post-thaw scripts.

- Microsoft SQL Server 2008, 2012, 2016, and 2019
- Microsoft Exchange 2010
- Microsoft Exchange 2013
- Microsoft Exchange 2016

- Nutanix does not support application-consistent recovery points on Windows VMs that have mounted VHDX disks.
- The system captures hypervisor-based recovery points only when you have VMware Tools running on the guest VM and the guest VM does not have any independent disks attached to it.

  If these requirements are not met, the system captures crash-consistent snapshots.

- The following table provides detailed information on whether a recovery point is application-consistent or not depending on the operating systems and hypervisors running in your environment.

  Note:

  - Installed and active means that the guest VM has the following.
    - NGT installed.
    - VSS capability enabled.
    - Powered on.
    - Actively communicating with the CVM.

Table 21: Application-consistent Recovery Points

Server	ESXi		AHV	
	NGT status	Result	NGT status	Result
Microsoft Windows Server edition	Installed and active. Also pre-freeze and post-thaw scripts are present.	Nutanix script-based VSS snapshots	Installed and active. Also pre-freeze and post-thaw scripts are present.	Nutanix script-based VSS snapshots
	Installed and active	Nutanix VSS-enabled snapshots.	Installed and active	Nutanix VSS-enabled snapshots
	Not enabled	Hypervisor-based application-consistent or crash-consistent snapshots.	Not enabled	Crash-consistent snapshots
Microsoft Windows Client edition	Installed and active. Also pre-freeze and post-thaw scripts are present.	Nutanix script-based VSS snapshots	Installed and active. Also pre-freeze and post-thaw scripts are present.	Nutanix script-based VSS snapshots
	Not enabled	Hypervisor-based snapshots or crash-consistent snapshots.	Not enabled	Crash-consistent snapshots
Linux VMs	Installed and active. Also pre-freeze and post-thaw scripts are present.	Nutanix script-based VSS snapshots	Installed and active. Also pre-freeze and post-thaw scripts are present.	Nutanix script-based VSS snapshots
	Not enabled	Hypervisor-based snapshots or crash-consistent snapshots.	Not enabled	Crash-consistent snapshots

Creating a Recovery Plan (Leap)

To orchestrate the failover (disaster recovery) of the protected guest VMs to the recovery site, create a recovery plan. After a failover, a recovery plan recovers the protected guest VMs to the recovery site. If you have configured two on-prem recovery sites in a protection policy, create two recovery plans for DR—one for recovery to each recovery site. The recovery plan synchronizes continuously to the recovery site in a bidirectional way.

About this task

To create a recovery plan, do the following at the primary site. You can also create a recovery plan at a recovery site. The recovery plan you create or update at a recovery site synchronizes back to the primary site.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.

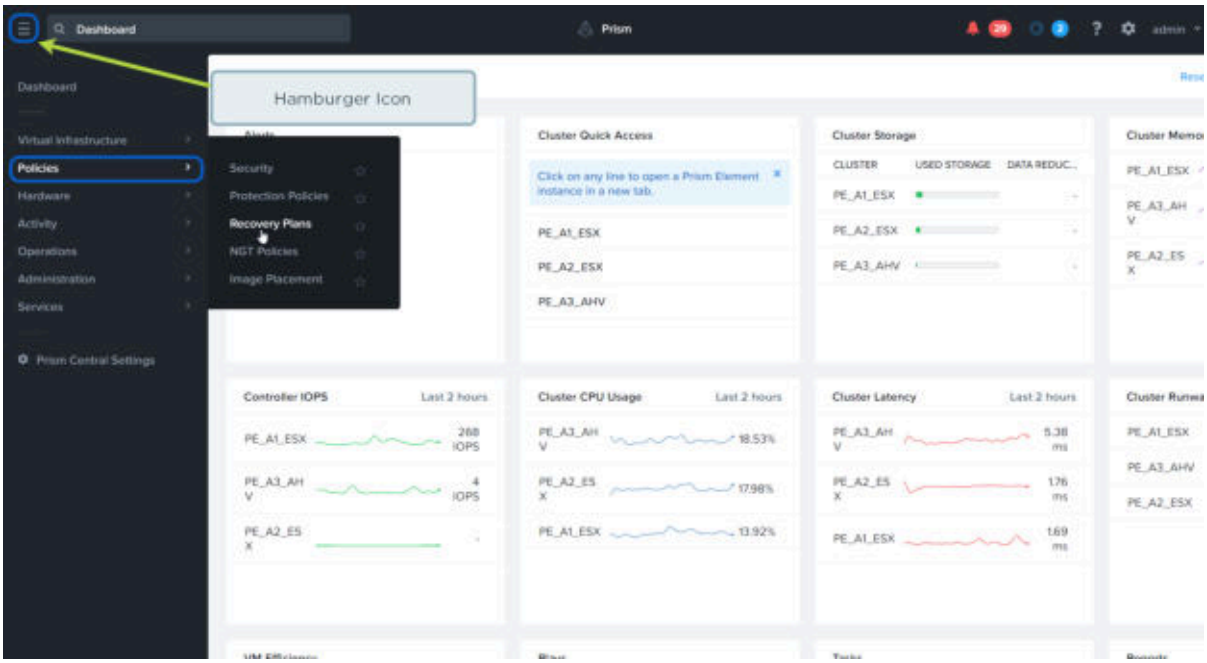


Figure 30: Recovery Plan Configuration: Recovery Plans

3. Click **Create Recovery Plan**.
   Specify the following information in the Create Recovery Plan window.

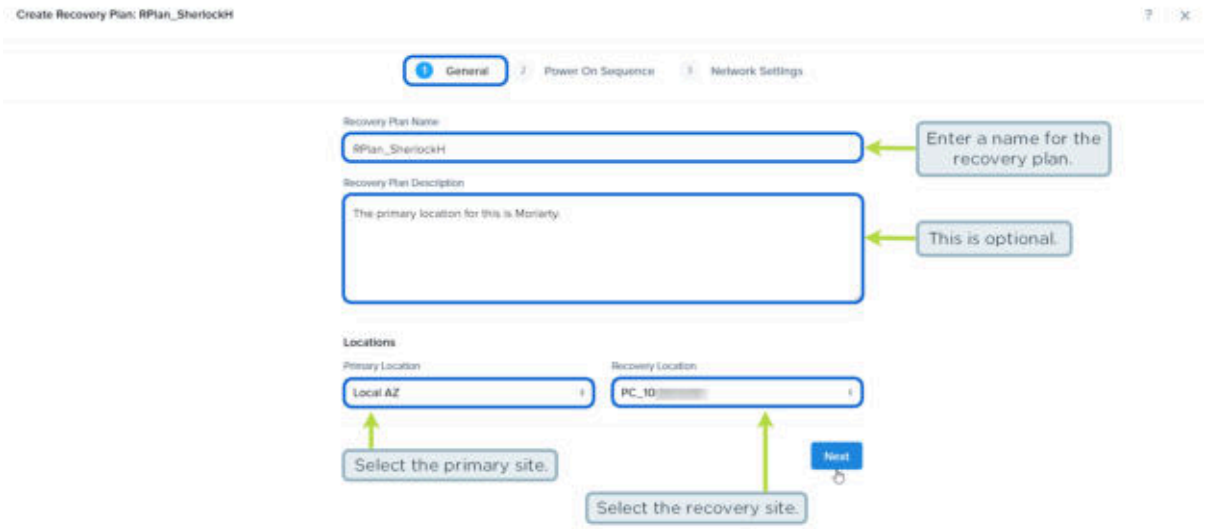


Figure 31: Recovery Plan Configuration: General

4. In the **General** tab, enter **Recovery Plan Name**, **Recovery Plan Description**, **Primary Location**, **Recovery Location**, and click **Next**.
   From **Primary Location** and **Recovery Location** drop-down lists, you can select either the local availability zone (site) or a non-local site to serve as your primary and recovery sites respectively. Local AZ represents the local site (Prism Central). If you are configuring recovery plan to recover the protected guest VMs to another Nutanix cluster at the same site, select Local AZ from both **Primary Location** and **Recovery Location** drop-down lists.

5. In the **Power On Sequence** tab, click **+ Add Entities** to add the guest VMs to the start sequence.

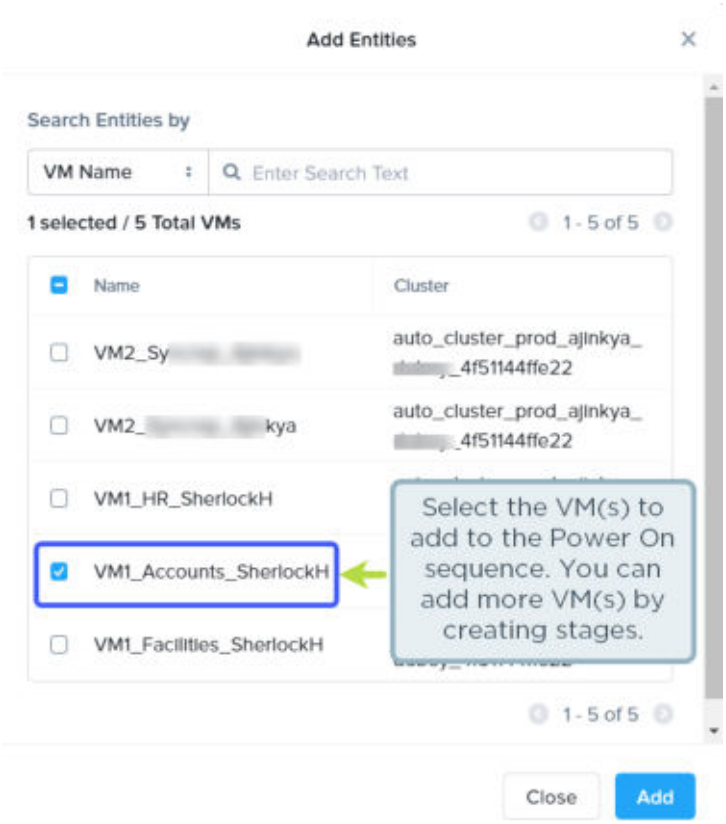


Figure 32: Recovery Plan Configuration: Add Entities

a. In the **Search Entities by**, select **VM Name** from the drop-down list to specify guest VMs by name.

b. In the **Search Entities by**, select **Category** from the drop-down list to specify guest VMs by category.

c. To add the guest VMs or VM categories to the stage, select the VMs or VM categories from the list.

> Note: The VMs listed in the search result are in the active state of replication.

d. Click **Add**.
The selected guest VMs are added to the start sequence in a single stage by default. You can also create multiple stages to add guest VMs and define the order of their power-on sequence. For more information about stages, see Stage Management on page 68.

> Caution: Do not include the guest VMs protected with Asynchronous, NearSync, and Synchronous replication schedules in the same recovery plan. You can include guest VMs protected with Asynchronous or NearSync replication schedules in the same recovery plan.

However, if you combine these guest VMs with the guest VMs protected by Synchronous replication schedules in a recovery plan, the recovery fails.

e. To automate in-guest script execution on the guest VMs during recovery, select the individual guest VMs or VM categories in the stage and click **Manage Scripts**.

Note: In-guest scripts allow you to automate various task executions upon recovery of the guest VMs. For example, in-guest scripts can help automate the tasks in the following scenarios.

- After recovery, the guest VMs must use new DNS IP addresses and also connect to a new database server that is already running at the recovery site.

  Traditionally, to achieve this new configuration, you would manually log on to the recovered VM and modify the relevant files. With in-guest scripts, you have to write a script to automate the required steps and enable the script when you configure a recovery plan. The recovery plan execution automatically invokes the script and performs the reassigning of DNS IP address and reconnection to the database server at the recovery site.

- If guest VMs are part of domain controller *siteA.com* at the primary site *AZ1*, and after the guest VMs recover at the site *AZ2*, you want to add the recovered guest VMs to the domain controller *siteB.com*.

  Traditionally, to reconfigure, you would manually log on to the VM, remove the VM from an existing domain controller, and then add the VM to a new domain

controller. With in-guest scripts, you can automate the task of changing the domain controller.

Note: In-guest script execution requires NGT version 1.9 or newer installed on the VM. The in-guest scripts run as a part of the recovery plan only if they have executable permissions for the following.

- Administrator user (Windows)
- Root user (Linux)

Note: You can define a batch or shell script that executes automatically in the guest VMs after their disaster recovery. Place two scripts—one for production failover and the other for test failover—at the following locations in the guest VMs with the specified name.

- In Windows VMs,
  - Batch script file path for production failover:

    `C:\Program Files\Nutanix\Scripts\production\vm_recovery`

    Replace *vm_recovery* with the script name (without extension).
  - Batch script file path for test failover:

    `C:\Program Files\Nutanix\Scripts\test\vm_recovery`

    Replace *vm_recovery* with the script name (without extension).
- In Linux VMs,
  - Shell script file path for production failover:

    `/usr/local/sbin/production_vm_recovery`

    Replace *production_vm_recovery* with the script name (without extension).
  - Shell script file path for test failover:

    `/usr/local/sbin/test_vm_recovery`

    Replace *test_vm_recovery* with the script name (without extension).

Note: When an in-guest script runs successfully, it returns code **0**. Any non-zero error code signifies that the execution of the in-guest script was unsuccessful.

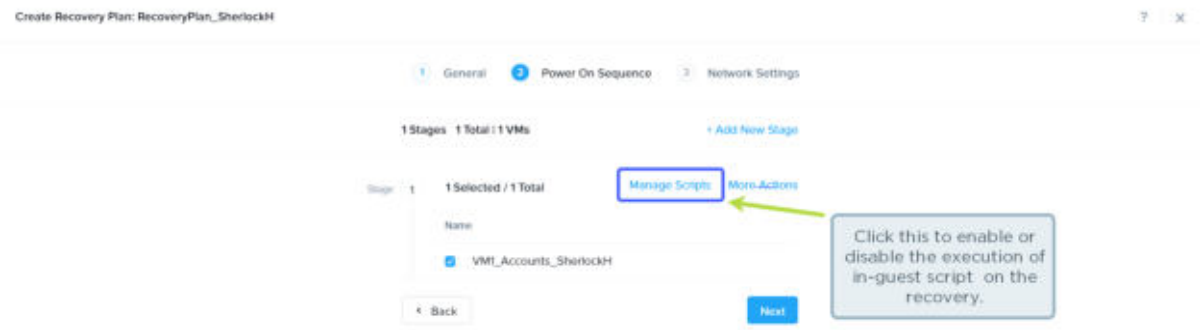


Figure 33: Recovery Plan Configuration: In-guest Script Execution

- 1. To enable script execution, click **Enable**.

  A command prompt icon appears against the guest VMs or VM categories to indicate that in-guest script execution is enabled on those guest VMs or VM categories.
  2. To disable script execution, click **Disable**.

6. In the **Network Settings** tab, map networks in the primary cluster to networks at the recovery cluster.

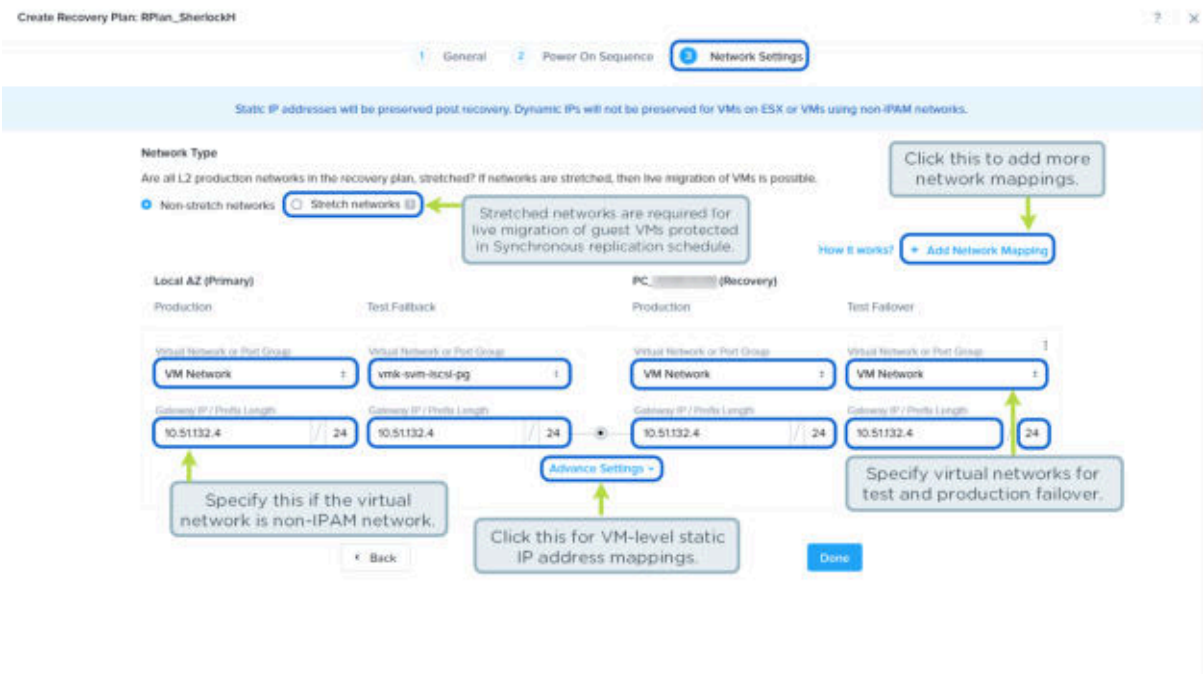


Figure 34: Recovery Plan Configuration: Network Settings

Network mapping enables replicating the network configurations of the primary Nutanix clusters to the recovery Nutanix clusters, and recover guest VMs into the same subnet at the recovery Nutanix cluster. For example, if a guest VM is in the *vlan0* subnet at the primary Nutanix cluster, you can configure the network mapping to recover that guest VM in the same *vlan0* subnet at the recovery Nutanix cluster. To specify the source (primary Nutanix cluster) and destination (recovery Nutanix cluster) network information for network mapping, do the following in **Local AZ (Primary)** and **PC 10.xx.xx.xxx (Recovery)** panes.

a. Under **Production** in **Virtual Network or Port Group** drop-down list, select the production subnet that contains the protected guest VMs. (optional) If the virtual network is a non-IPAM network, specify the gateway IP address and prefix length in the **Gateway IP/Prefix Length** field.

b. Under **Test Failback** in **Virtual Network or Port Group** drop-down list, select the test subnet that you want to use for testing failback from the recovery Nutanix cluster. (optional) If the virtual network is a non-IPAM network, specify the gateway IP address and prefix length in the **Gateway IP/Prefix Length** field.

c. To add more network mappings, click **Add Networks** at the top-right corner of the page, and then repeat the steps 6.a-6.b.

> Note: The primary and recovery Nutanix clusters must have identical gateway IP addresses and prefix length. Therefore you cannot use a test failover network for two or more network mappings in the same recovery plan.

d. Click **Done**.

> Note: For ESXi, you can configure network mapping for both standard and distributed (DVS) port groups. For more information about DVS, see VMware documentation.

> Caution: Leap does not support VMware NSX-T datacenters. For more information about NSX-T datacenters, see VMware documentation.

7. To perform VM-level static IP address mapping between the primary and the recovery sites, click **Advanced Settings**, click **Custom IP Mapping**, and then do the following.

> Note: The **Custom IP Mapping** shows all the guest VMs with static IP address configured, NGT installed, and VNIC in the source subnet specified in the network mapping.

   a. To locate the guest VM, type the name of the guest VM in the filter field.

      A guest VM that has multiple NICs lists in multiple rows, allowing you to specify an IP address mapping for each VNIC. All the fields auto-populate with the IP addresses generated based on the offset IP address-mapping scheme.

   b. In the **Test Failback** field for the local site, **Production** field for the remote (recovery) site, and **Test Failover** for the remote site, edit the IP addresses.

      Perform this step for all the IP addresses that you want to map.

      > Caution: Do not edit the IP address assigned to the VNIC in the local site. If you do not want to map static IP addresses for a particular VNIC, you can proceed with the default entries.

   c. Click **Save**.

   d. If you want to edit one or more VM-level static IP address mappings, click **Edit**, and then change the IP address mapping.

8. If VM-level static IP address mapping is configured between the primary and the recovery Nutanix clusters and you want to use the default, offset-based IP address-mapping scheme, click **Reset to Matching IP Offset**.

9. Click **Done**.
   The recovery plan is created. To verify the recovery plan, see the **Recovery Plans** page. You can modify the recovery plan to change the recovery location, add, or remove the protected guest VMs. For information about various operations that you can perform on a recovery plan, see Recovery Plan Management on page 141.

Stage Management

A stage defines the order in which the protected guest VMs start at the recovery cluster. You can create multiple stages to prioritize the start sequence of the guest VMs. In the **Power On Sequence**, the VMs in the preceding stage start before the VMs in the succeeding stages. On recovery, it is desirable to start some VMs before the others. For example, database VMs must start before the application VMs. Place all the database VMs in the stage before the stage containing the application VMs, in the **Power On Sequence**.

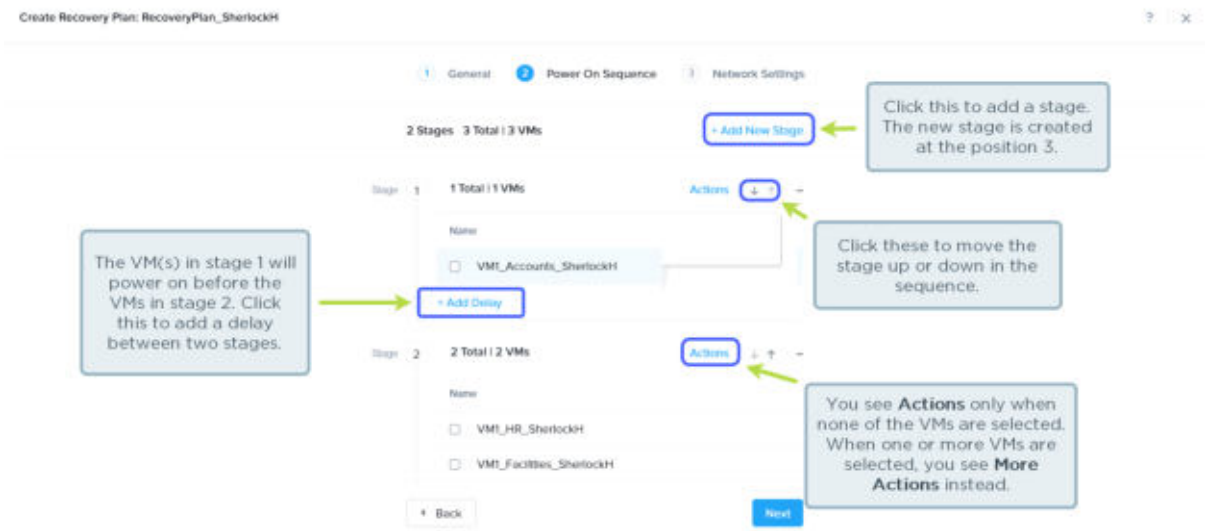


Figure 35: Recovery Plan Configuration: Power On Sequence

To Add a Stage in the Power-On Sequence and Add Guest VMs to It, Do the Following.

1. Click **+Add New Stage**.
2. Click **+Add Entities**.
3. To add guest VMs to the current stage in the power-on sequence, do the following.
   1. In the **Search Entities by**, select **VM Name** from the drop-down list to specify guest VMs by name.
   2. In the **Search Entities by**, select **Category** from the drop-down list to specify guest VMs by category.
   3. To add the guest VMs or VM categories to the stage, select the guest VMs or VM categories from the list.

      Note: The VMs listed in the search result are in the active state of replication.
4. Click **Add**.

To Remove a Stage from the Power-On Sequence, Do the Following.

Click **Actions** > **Remove Stage**

Note: You see **Actions** in a stage only when none of the VMs in the stage are selected. When one or more VMs in the stage are selected, you see **More Actions**.

To Change the Position of a Stage in the Power-On Sequence, Do the Following.

- To move a stage up or down in the power-on sequence, click **#** or **#** respectively, at the top-right corner of the stage.
- To expand or collapse a stage, click **+** or **-** respectively, at the top-right corner of the stage.
- To move VMs to a different stage, select the VMs, do the following.
  1. Click **More Actions** > **Move**.
  2. Select the target stage from the list.

     Note: You see **Move** in the **More Actions** only when you have defined two or more stages.

To Set a Delay Between the Power-On Sequence of Two Stages, Do the Following.

1. Click **+Add Delay**.
2. Enter the time in seconds.
3. Click **Add**.

To Add Guest VMs to an Existing Stage, Do the Following.

1. Click **Actions** > **Add Entities**.

   > Note: You see **Actions** in a stage only when none of the VMs in the stage are selected. When one or more VMs in the stage are selected, you see **More Actions**.

2. To add VMs to the current stage in the power-on sequence, do the following.
   1. In the **Search Entities by**, select **VM Name** from the drop-down list to specify guest VMs by name.
   2. In the **Search Entities by**, select **Category** from the drop-down list to specify guest VMs by category.
   3. To add the guest VMs or VM categories to the stage, select the guest VMs or VM categories from the list.

      > Note: The VMs listed in the search result are in the active state of replication.

3. Click **Add**.

To Remove Guest VMs from an Existing Stage, Do the Following.

1. Select the VMs from the stage.
2. Click **More Actions** > **Remove**.

   > Note: You see **More Actions** in a stage only when one or more VMs in the stage are selected. When none of the VMs in the stage are selected, you see **Actions**.

To Move Guest VMs to a Different Stage, Do the Following.

1. Select the VMs from the stage.
2. Click **More Actions** > **Move**.

   > Note: You see **More Actions** in a stage only when one or more VMs in the stage are selected. When none of the VMs in the stage are selected, you see **Actions**.

3. Select the target stage from the list.

Failover and Failback Management

You perform failover of the protected guest VMs when unplanned failure events (for example, natural disasters) or planned events (for example, scheduled maintenance) happen at the primary availability zone (site) or the primary cluster. The protected guest VMs migrate to the recovery site where you perform the failover operations. On recovery, the protected guest VMs start in the Nutanix cluster you specify in the recovery plan that orchestrates the failover.

The following are the types of failover operations.

Test failover

To ensure that the protected guest VMs failover efficiently to the recovery site, you perform a test failover. When you perform a test failover, the guest VMs recover in the virtual network designated for testing purposes at the recovery site. However, the guest VMs at the primary site are not affected. Test failovers rely on the presence of VM recovery points at the recovery sites.

Planned failover

To ensure VM availability when you foresee service disruption at the primary site, you perform a planned failover to the recovery site. For a planned failover to succeed, the guest VMs must be available at the primary site. When you perform a planned failover, the recovery plan first creates a recovery point of the protected guest VM, replicates the recovery point to the recovery site, and then starts the guest VM at the recovery site. The recovery point used for migration is retained indefinitely. After a planned failover, the guest VMs no longer run at the primary site.

Unplanned failover

To ensure VM availability when a disaster causing service disruption occurs at the primary site, you perform an unplanned failover to the recovery site. In an unplanned failover, you can expect some data loss to occur. The maximum data loss possible is equal to the least RPO you specify in the protection policy, or the data that was generated after the last manual recovery point for a given guest VM. In an unplanned failover, by default, the protected guest VMs recover from the most recent recovery point. However, you can recover from an earlier recovery point by selecting a date and time of the recovery point.

At the recovery site, the guest VMs can recover using the recovery points replicated from the primary site only. The guest VMs cannot recover using the local recovery points. For example, if you perform an unplanned failover from the primary site *AZ1* to the recovery site *AZ2*, the guest VMs recover at *AZ2* using the recovery points replicated from *AZ1* to *AZ2*.

You can perform a planned or an unplanned failover in different scenarios of network failure. For more information about network failure scenarios, see Leap and Xi Leap Failover Scenarios on page 71.

At the recovery site after a failover, the recovery plan creates only the VM category that was used to include the guest VM in the recovery plan. Manually create the remaining VM categories at the recovery site and associate the guest VMs with those categories.

The recovered guest VMs generate recovery points as per the replication schedule that protects it even after recovery. The recovery points replicate back to the primary site when the primary site starts functioning. The approach for reverse replication enables you to perform failover of the guest VMs from the recovery site back to the primary site (failback). The same recovery plan applies to both the failover and the failback operations. Only, for failover, you must perform the failover operations on the recovery plan at the recovery site while for failback, you must perform the failover operations on the recovery plan at the primary site. For example, if a guest VM fails over from *AZ1 (Local)* to *AZ2*, the failback fails over the same VMs from *AZ2 (Local)* back to *AZ1*.

Leap and Xi Leap Failover Scenarios

You have the flexibility to perform a real or simulated failover for the full and partial workloads (with or without networking). The term virtual network is used differently on on-prem clusters and Xi Cloud Services. In Xi Cloud Services, the term virtual network is used to describe the two built-in virtual networks—production and test. Virtual networks on the on-prem clusters are virtual subnets bound to a single VLAN. Manually create these virtual subnets, and create separate virtual subnets for production and test purposes. Create these virtual subnets before you configure recovery plans. When configuring a recovery plan, you map the virtual subnets at the primary site to the virtual subnets at the recovery site.

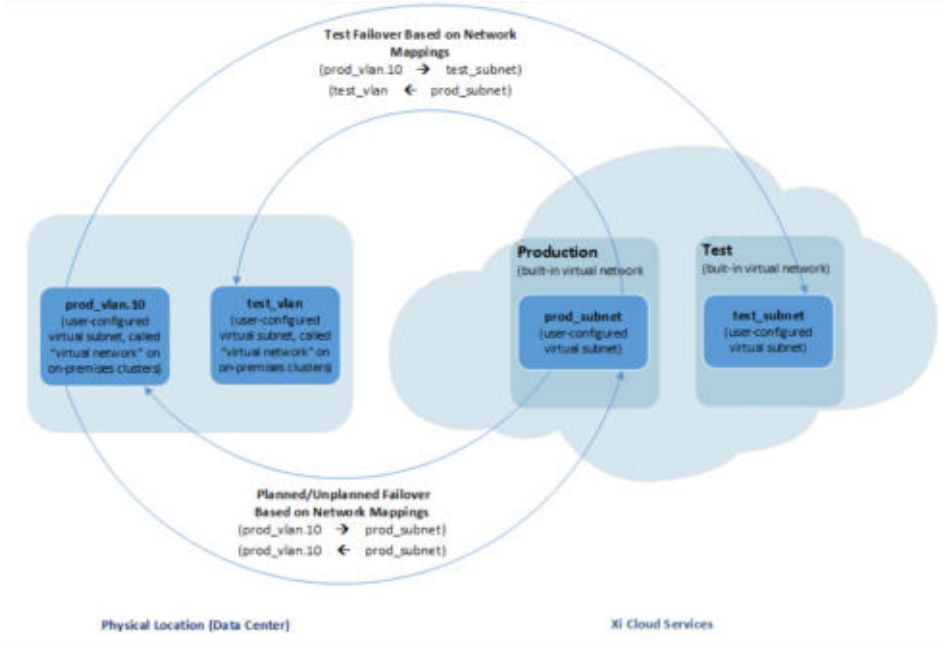


Figure 36: Failover in Network Mapping

The following are the various scenarios that you can encounter in Leap configurations for disaster recovery (DR) to an on-prem availability zone (site) or to Xi Cloud (Xi Leap). Each scenario is explained with the required network-mapping configuration for Xi Leap. However, the configuration remains the same irrespective of disaster recovery (DR) using Leap or Xi Leap. You can either create a recovery plan with the following network mappings (see Creating a Recovery Plan (Leap) on page 60) or update an existing recovery plan with the following network mappings (see Updating a Recovery Plan on page 143).

Scenario 1: Leap Failover (Full Network Failover)

Full network failure is the most common scenario. In this case, it is desirable to bring up the whole primary site in the Xi Cloud. All the subnets must failover, and the WAN IP address must change from the on-prem IP address to the Xi WAN IP address. Floating IP addresses can be assigned to individual guest VMs, otherwise, everything use Xi network address translation (NAT) for external communication.

Perform the failover when the on-prem subnets are down and jump the host available on the public Internet through the floating IP address of Xi production network.

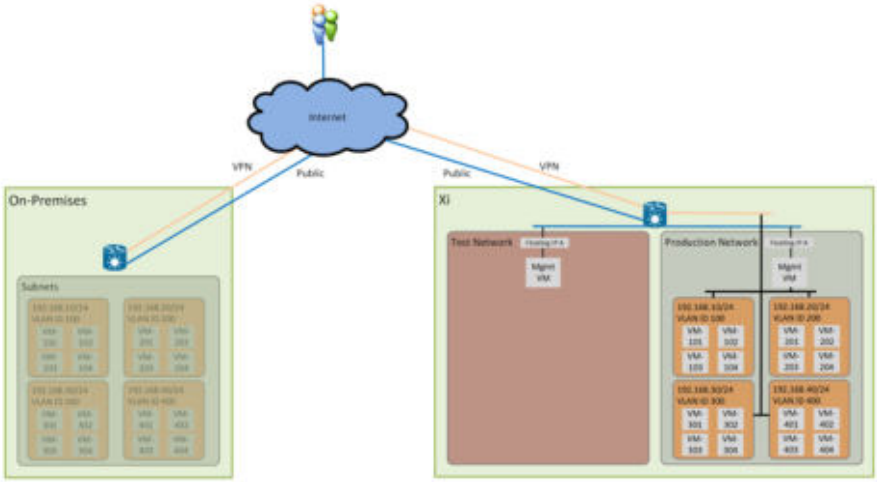


Figure 37: Full Network Failover

To set up the recovery plan that orchestrates the full network failover, perform the following.

1. Open the **Network Settings** page to configure network mappings in a recovery plan.
2. Select the **Local AZ** > **Production** > **Virtual Network or Port Group**.

   The selection auto-populates the Xi production and test failover subnets.
3. Select the **Outbound Internet Access** switch to allow the Xi NAT to use for Internet access.
4. Dynamically assign the floating IP addresses to the guest VMs you select in the recovery plan.

   Perform steps 1–4 for every subnet.

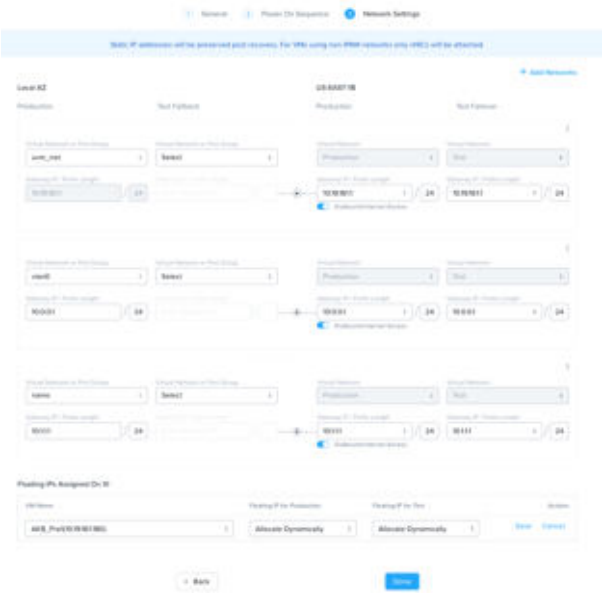


Figure 38: Recovery Plan Configuration: Network Settings

Scenario 2: Xi Network Failover (Partial Network Failover)

You want to failover one or more subnets from the primary site to Xi Cloud. The communications between the sites happen through the VPN or using the external NAT

or floating IP addresses. A use case of this type of scenario is that the primary site needs maintenance, but some of its subnets must see no downtime.

Perform partial failover when some subnets are active in the production networks at both on-prem and Xi Cloud, and jump the host available on the public Internet through the floating IP address of Xi production network.

On-prem guest VMs can connect to the guest VMs on the Xi Cloud Services.

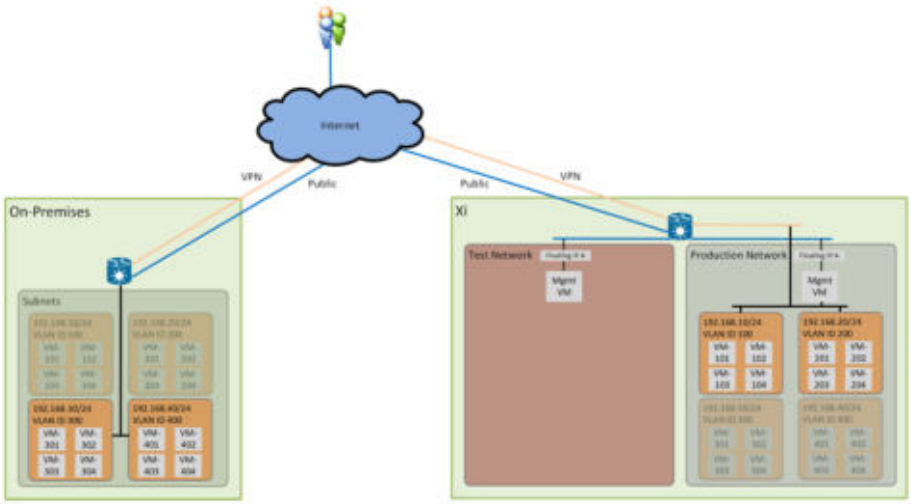


Figure 39: Partial Network Failover

To set up the recovery plan that orchestrates the partial network failover, perform the following.

1. Open the **Network Settings** page to configure network mappings in a recovery plan.
2. Select the **Local AZ** > **Production** > **Virtual Network or Port Group**.

   The selection auto-populates the Xi production and test failover subnets.
3. Select the **Outbound Internet Access** switch to allow the Xi NAT to use for Internet access.
4. Dynamically assign the floating IP addresses to the guest VMs you select in the recovery plan.

   Perform steps 1–4 for one or more subnets based on the maintenance plan.

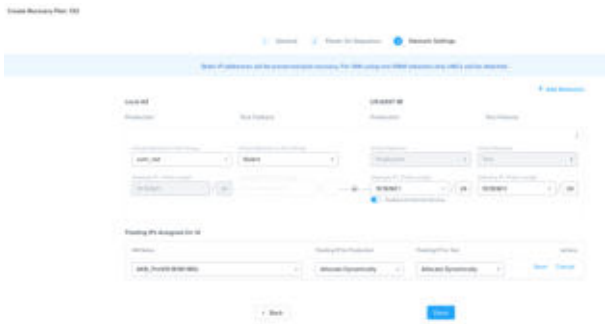


Figure 40: Recovery Plan Configuration: Network Settings

Scenario 3: Xi Network Failover (Partial Subnet Network Failover)

You want to failover some guest VMs to Xi Cloud, while keeping the other guest VMs up and running at the on-prem cluster (primary site). A use case of this type of scenario is that the primary site needs maintenance, but some of its guest VMs must see no downtime.

This scenario requires changing IP addresses for the guest VMs running at Xi Cloud. Since you cannot have the subnet active on both the sites, create a subnet to host the failed over guest VMs. Jump the host available on the public Internet through the floating IP address of Xi production network.

On-prem guest VMs can connect to the guest VMs on the Xi Cloud Services.

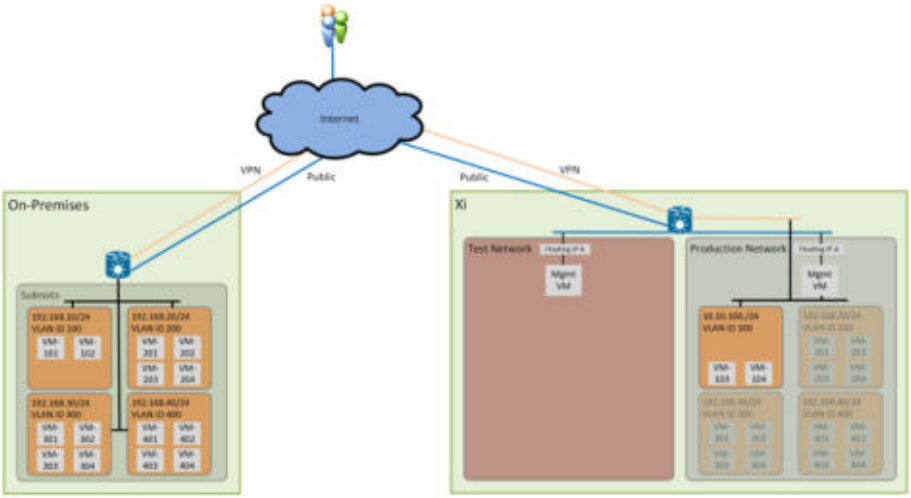


Figure 41: Partial Subnet Network Failover

To set up the recovery plan that orchestrates the partial subnet network failover, perform the following.

1. Open the **Network Settings** page to configure network mappings in a recovery plan.
2. Select the **Local AZ** > **Production** > **Virtual Network or Port Group**.

   The selection auto-populates the Xi production and test failover subnets for a full subnet failover

   > Note: In this case, you have created subnets on the Xi Cloud Services also. Choose that subnets to avoid full subnet failover (scenario 1).

3. Select the **Outbound Internet Access** switch to allow the Xi NAT to use for Internet access.
4. Dynamically assign the floating IP addresses to the guest VMs you select in the recovery plan.

   Perform steps 1–4 for one or more subnets based on the maintenance plan.

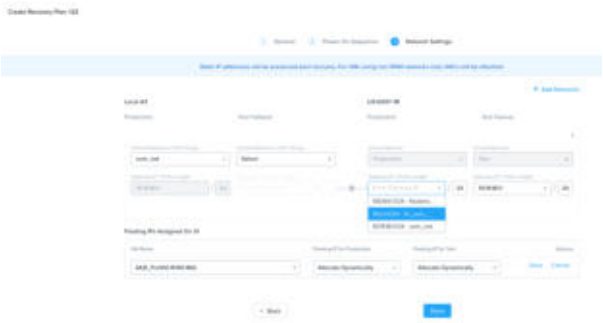


Figure 42: Recovery Plan Configuration: Network Settings

Scenario 4: Xi Network Failover (Test Failover and Failback)

You want to test all the preceding three scenarios by creating an isolated test network so that no routing or IP address conflict happens. Clone all the guest VMs from a local recovery point and bring up to test failover operations. Test failover test when all on-prem subnets are active and on-prem guest VMs can connect to the guest VMs at the Xi Cloud. Jump the host available on the public Internet through the floating IP address of Xi production network.

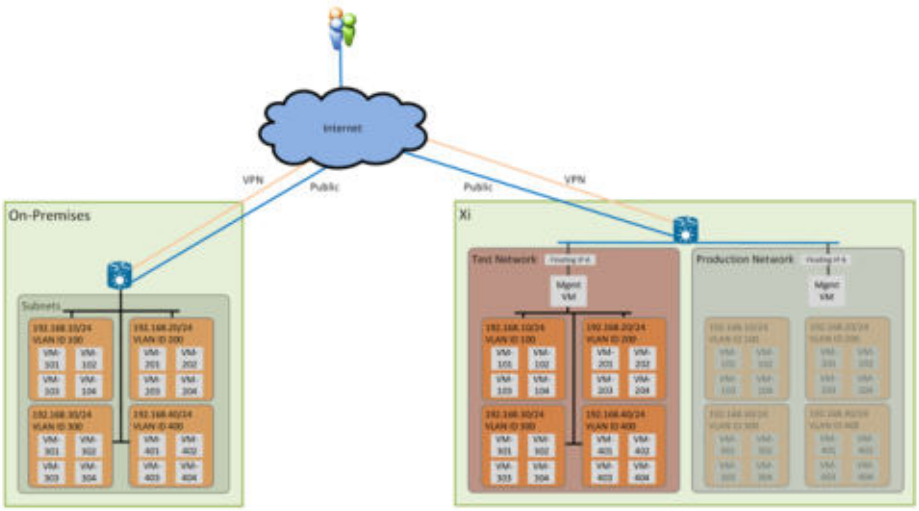


Figure 43: Test Failover & Failback

In this case, focus on the test failover section when creating the recovery plan. When you select a local AZ production subnet, it copies to the test network. You can go one step further and create a test subnet at the Xi Cloud.

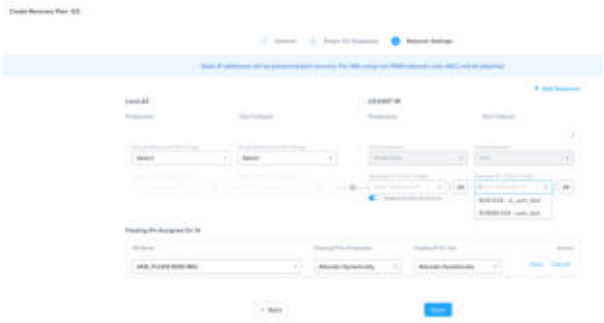


Figure 44: Recovery Plan Configuration: Network Settings

After the guest VMs test failover to Xi Cloud, you can do a test failback to the primary site.

Note: Make a test subnet in advance for the failback to the on-prem cluster.

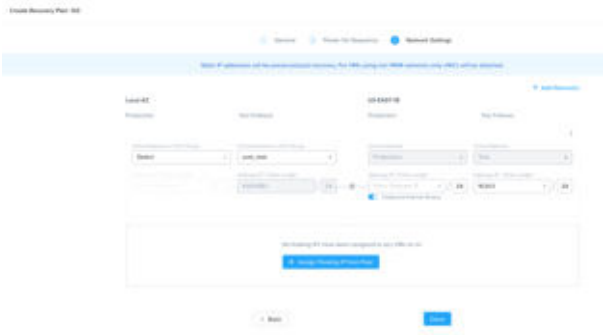


Figure 45: Recovery Plan Configuration: Network Settings

Failover and Failback Operations (Leap)

You can perform test failover, planned failover, and unplanned failover of the guest VMs protected with Asynchronous replication schedule across different Nutanix clusters at the same or different on-prem availability zones (sites). The steps to perform test, planned, and unplanned failover are largely the same irrespective of the replication schedules that protect the guest VMs.

Performing a Test Failover (Leap)

After you create a recovery plan, you can run a test failover periodically to ensure that the failover occurs smoothly when required. To perform a test failover, do the following procedure at the recovery site. If you have two recovery sites for DR, perform the test at the site where you want to recover the guest VMs.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select the recovery plan that you want to test.

4. Click **Test** from the **Actions** drop-down menu.

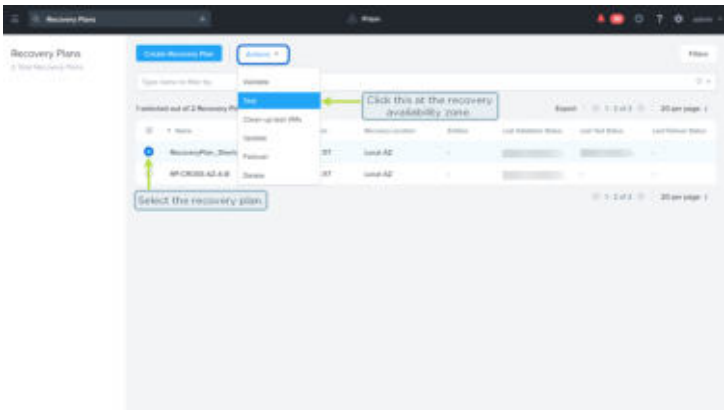


Figure 46: Test Failover (Drop-down)

Test Recovery Plan window shows. The window auto-populates the **Failover From** and **Failover To** locations from the recovery plan you select in step 3. **Failover To** location is Local AZ by default and is unavailable for editing.

Figure 47: Test Recovery Plan

5. Click **+ Add target clusters** if you want to failover to specific Nutanix clusters at the recovery site.

   If you do not add target clusters, the recovery plan recovers the guest VMs to any eligible cluster at the recovery site.

6. Click **Test**.
   The **Test Recovery Plan** dialog box lists the errors and warnings, if any, and allows you to stop or continue the test operation. If there are no errors or you resolve the errors in step 7, the guest VMs failover to the recovery cluster.

7. If you see errors, do the following.
   - To review errors or warnings, click **View Details** in the description.

     Resolve the error conditions and then restart the test procedure.
   - Select one of the following.
     - To stop the failover operation, click **Abort**.
     - To continue the failover operation despite the warnings, click **Execute Anyway**.

       Note: You cannot continue the failover operation when the validation fails with errors.

Cleaning up Test VMs (Leap)

After testing a recovery plan, you can remove the test VMs that the recovery plan creates in the recovery test network. To clean up the test VMs, do the following at the recovery site where the test failover created the test VMs.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select the recovery plans whose test VMs you want to remove.
4. Click **Clean Up Test VMs** from the **Actions** drop-down menu.

   Clean Up Test VMs dialog box shows with the name of the recovery plan you selected in step 3.
5. Click **Clean**.

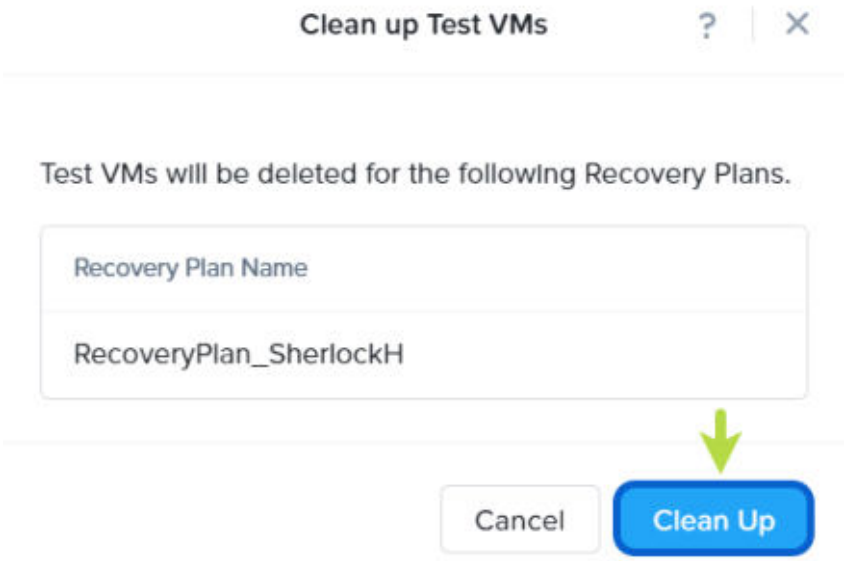


Figure 48: Clean Up Test VMs

Performing a Planned Failover (Leap)

If there is a planned event (for example, scheduled maintenance of guest VMs) at the primary availability zone (site), perform a planned failover to the recovery site. To perform a planned failover, do the following procedure at the recovery site. If you have two recovery sites for DR, perform the failover at the site where you want to recover the guest VMs.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select a recovery plan for the failover operation.

4. Click **Failover** from the **Actions** drop-down menu.

> Note: If you select more than one recovery plan in step 3, the **Failover** action is available only when the selected recovery plans have the same primary and recovery locations.

Specify the following information in the **Failover from Recovery Plan** window. The window auto-populates the **Failover From** and **Failover To** locations from the recovery plan you select in step 3.

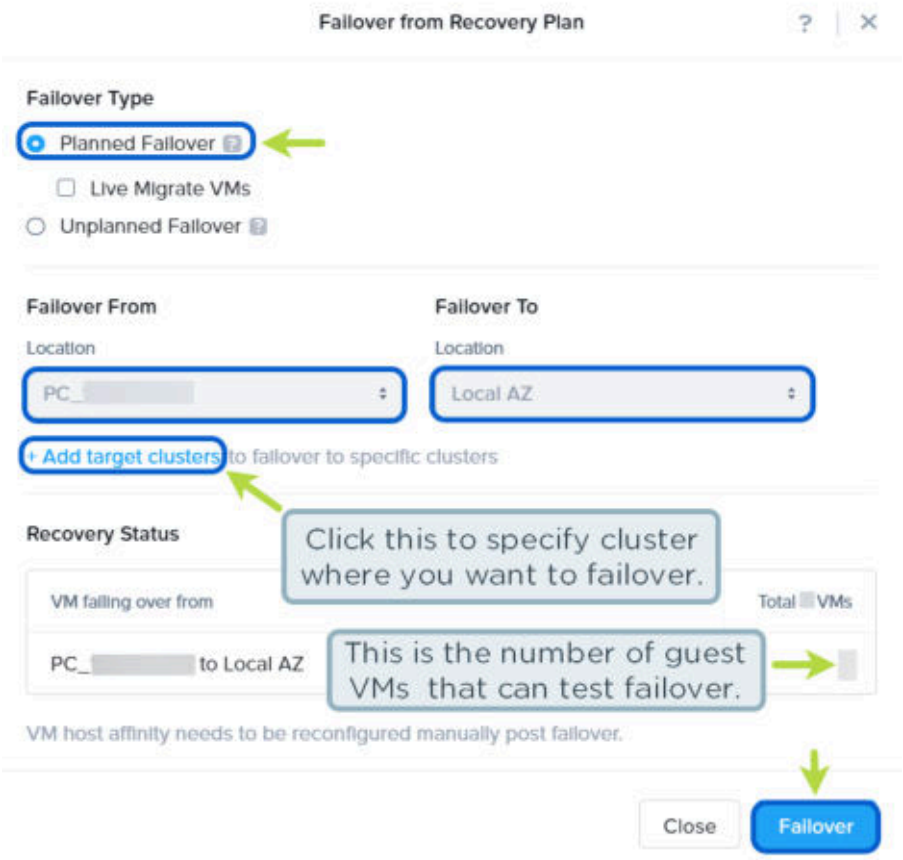


Figure 49: Planned Failover

a. **Failover Type**: Click **Planned Failover**.

> Warning: Do not check **Live Migrate VMs**. Live migration works only for the planned failover of the guest VMs protected in Synchronous replication schedule. If you check

Live Migrate VMs for the planned failover of the guest VMs protected in Asynchronous or NearSync replication schedule, the failover task fails.

b. Click **+ Add target clusters** if you want to failover to specific Nutanix clusters at the recovery site.

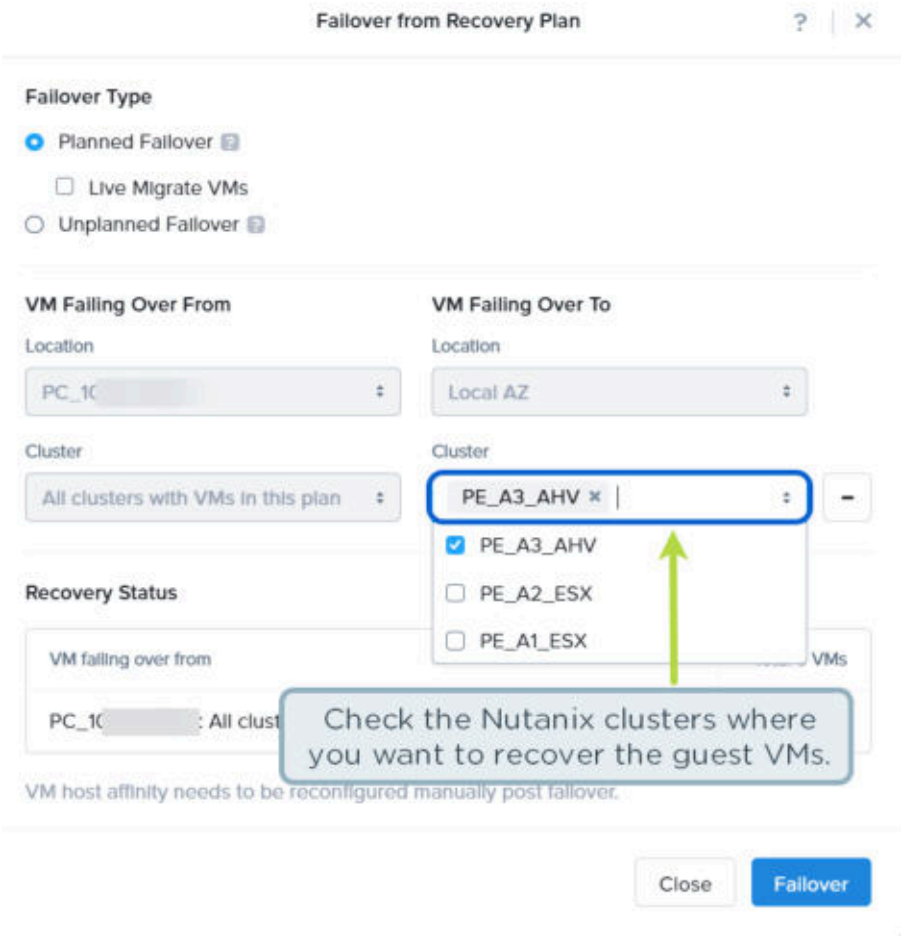


Figure 50: Planned Failover: Select Recovery Cluster

If you do not add target clusters, the recovery plan recovers the guest VMs to any eligible cluster at the recovery site.

5. Click **Failover**.
   The **Failover from Recovery Plan** dialog box lists the errors and warnings, if any, and allows you to stop or continue the failover operation. If there are no errors or you resolve the errors in step 6, the guest VMs failover to the recovery Nutanix cluster.

6. If you see errors, do the following.

   - To review errors or warnings, click **View Details** in the description.

     Resolve the error conditions and then restart the failover procedure.
   - Select one of the following.
     - To stop the failover operation, click **Abort**.
     - To continue the failover operation despite the warnings, click **Execute Anyway**.

       Note: You cannot continue the failover operation when the validation fails with errors.

   Note:

   The entities of AHV/ESXi clusters recover at a different path on the ESXi clusters if their files conflict with the existing files on the recovery ESXi cluster. For example, there is a file name conflict if a VM (VM1) migrates to a recovery cluster that already has a VM (VM1) in the same container.

   However, the entities recover at a different path with `VmRecoveredAtAlternatePath` alert only if the following conditions are met.

   - Prism Element running on both the primary and the recovery Nutanix clusters are of version 5.17 or newer.
   - A path for the entity recovery is not defined while initiating the failover operation.
   - The protected entities do not have shared disk/s.

   If these conditions are not satisfied, the failover operation fails.

Performing an Unplanned Failover (Leap)

If there is an unplanned event (for example, a natural disaster or network failure) at the primary availability zone (site), perform an unplanned failover to the recovery site. To perform an unplanned failover, do the following procedure at the recovery site. If you have two recovery sites for DR, perform the failover at the site where you want to recover the guest VMs.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select a recovery plan for the failover operation.

4. Click **Failover** from the **Actions** drop-down menu.

> Note: If you select more than one recovery plan in step 3, the **Failover** action is available only when the selected recovery plans have the same primary and recovery locations.

Specify the following information in the **Failover from Recovery Plan** window. The window auto-populates the **Failover From** and **Failover To** locations from the recovery plan you select in step 3.

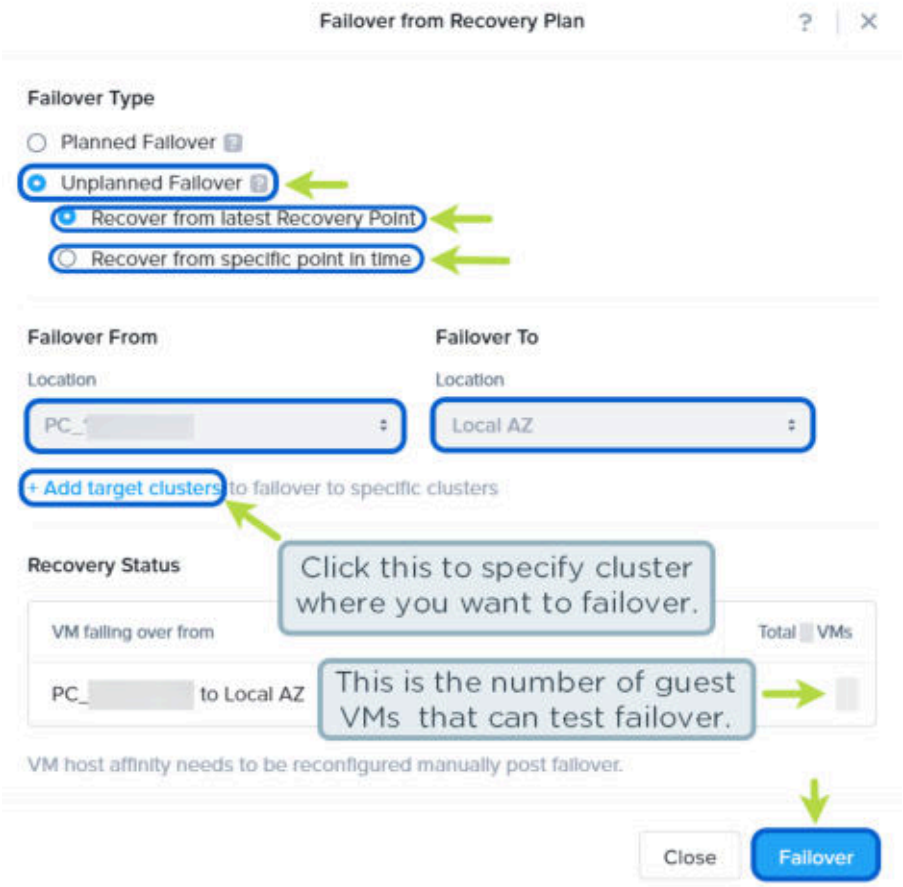


Figure 51: Unplanned Failover

a. **Failover Type**: Click **Unplanned Failover** and do one of the following.

   » Click **Recover from latest Recovery Point** to use the latest recovery point for recovery.

   » Click **Recover from specific point in time** to use a recovery point taken at a specific point in time for recovery.

> Note: If you click **Recover from specific point in time**, select a Nutanix cluster that hosts the specific point in time recovery point (step 4.b). If you do not select a cluster, or select multiple clusters where the same recovery points exist, the guest VMs fail to recover efficiently because the system encounters more than one recovery point at the recovery site. For example, if a primary site *AZ1* replicates the same recovery points to two clusters *CLA* and *CLB* at site *AZ2*, select either the cluster *CLA* or the cluster *CLB*

as the target cluster when you click to recover from a specific point in time. If you select both *CLA* and *CLB*, the guest VMs fail to recover.

b. Click **+ Add target clusters** if you want to failover to specific Nutanix clusters at the recovery site.

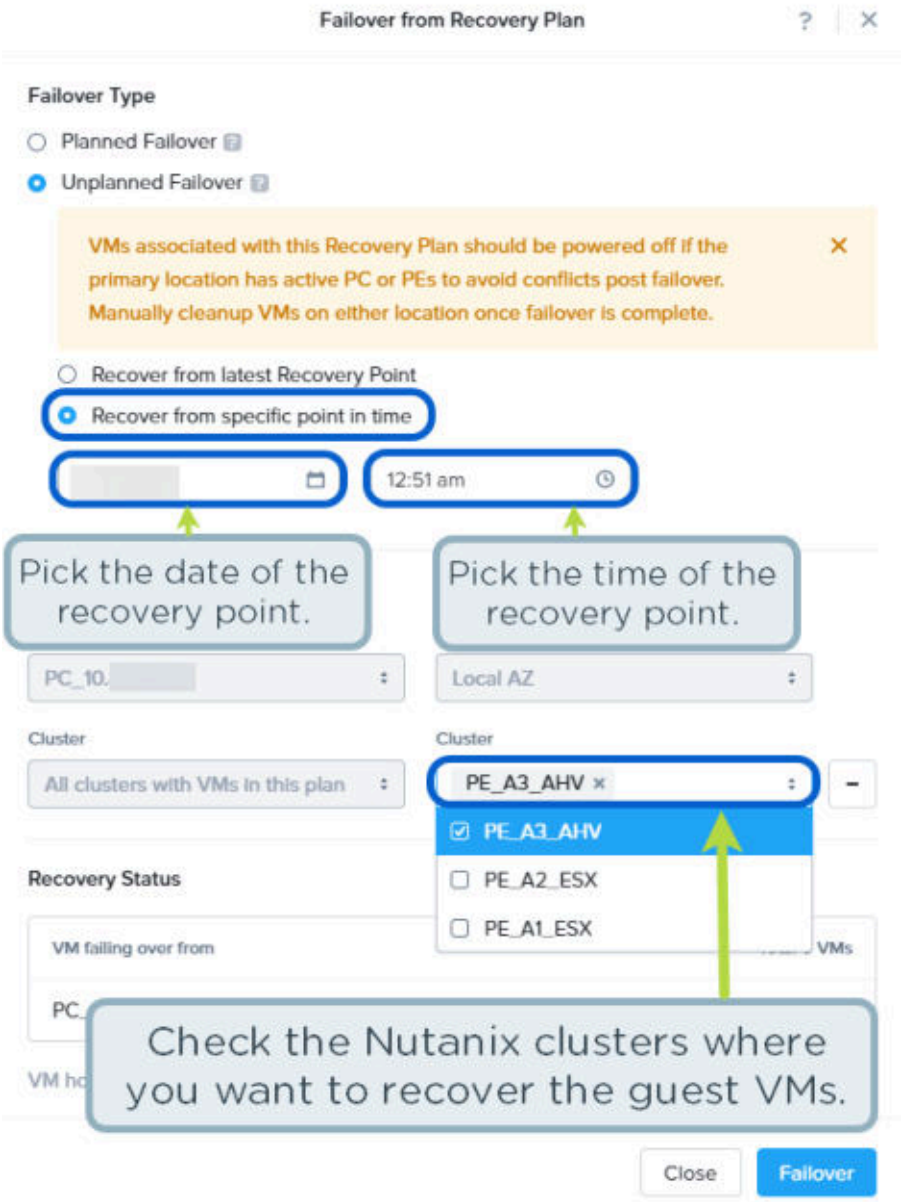


Figure 52: Unplanned Failover: Select Recovery Cluster

If you do not add target clusters, the recovery plan recovers the guest VMs to any eligible cluster at the recovery site.

Note: If recovery plans contain VM categories, the VMs from those categories recover in the same category after an unplanned failover to the recovery site. Also, the recovery points keep generating at the recovery site for those recovered VMs. Since the VM count represents the number of recoverable VMs (calculated from recovery points), the recovered VMs and their newly created recovery points sum up. Their sum gives double the count of the originally

recovered VMs on the recovery plans page. Now, if some VMs belonging to the given category at the primary or recovery site are deleted, the VM count at both sites still stay the same until the recovery points of deleted VMs expire. For example, when two VMs have failed over, the recovery plans page at the recovery site shows four VMs (two replicated recovery points from source and two newly generated recovery points). The page shows four VMs even if the VMs are deleted from the primary or recovery site. The VM count synchronizes and becomes consistent in the subsequent RPO cycle conforming to the retention policy set in the protection policy (due to the expiration of recovery points).

5. Click **Failover**.
   The **Failover from Recovery Plan** dialog box lists the errors and warnings, if any, and allows you to stop or continue the failover operation. If there are no errors or you resolve the errors in step 6, the guest VMs failover to the recovery Nutanix cluster.

6. If you see errors, do the following.

   - To review errors or warnings, click **View Details** in the description.

     Resolve the error conditions and then restart the failover procedure.

   - Select one of the following.

     - To stop the failover operation, click **Abort**.
     - To continue the failover operation despite the warnings, click **Execute Anyway**.

       Note: You cannot continue the failover operation when the validation fails with errors.

Note: To avoid conflicts when the primary site becomes active after the failover, shut down the guest VMs associated with this recovery plan. Manually power off the guest VMs on either primary or recovery site after the failover is complete. You can also block the guest VMs associated with this recovery plan through the firewall.

Note:

The entities of AHV/ESXi clusters recover at a different path on the ESXi clusters if their files conflict with the existing files on the recovery ESXi cluster. For example, there is a file name conflict if a VM (VM1) migrates to a recovery cluster that already has a VM (VM1) in the same container.

However, the entities recover at a different path with `VmRecoveredAtAlternatePath` alert only if the following conditions are met.

- Prism Element running on both the primary and the recovery Nutanix clusters are of version 5.17 or newer.
- A path for the entity recovery is not defined while initiating the failover operation.
- The protected entities do not have shared disk/s.

If these conditions are not satisfied, the failover operation fails.

Performing Failback (Leap)

A failback is failover of the guest VMs from the recovery availability zone (site) back to the primary site. The same recovery plan applies to both the failover and the failback operations. Only, for failover, you must perform the failover operations on the recovery plan at the recovery

site while for failback, you must perform the failover operations on the recovery plan at the primary site.

About this task

To perform a failback, do the following procedure at the primary site.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select a recovery plan for the failover operation.

4. Click **Failover** from the **Actions** drop-down menu.

> Note: If you select more than one recovery plan in step 3, the **Failover** action is available only when the selected recovery plans have the same primary and recovery locations.

Specify the following information in the **Failover from Recovery Plan** window. The window auto-populates the **Failover From** and **Failover To** locations from the recovery plan you select in step 3.

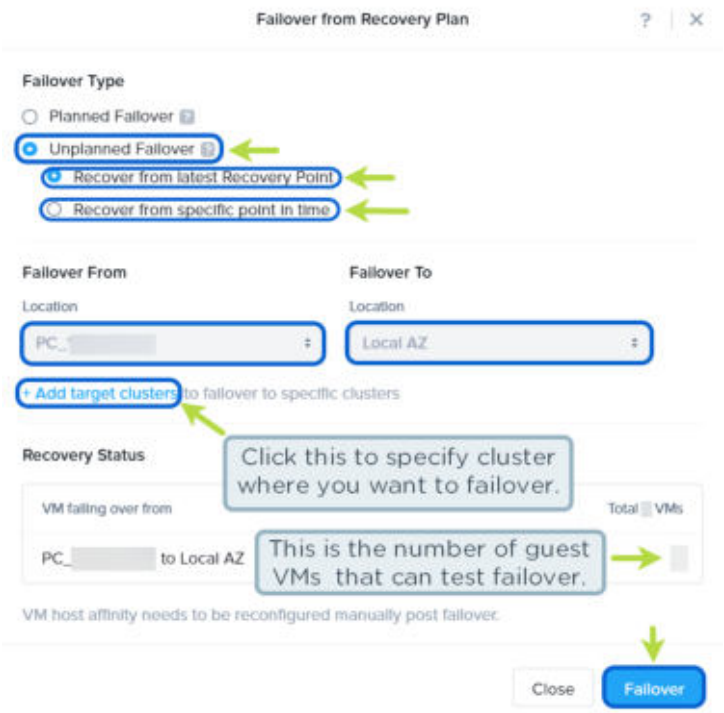


Figure 53: Unplanned Failover

a. **Failover Type**: Click **Unplanned Failover** and do one of the following.

> Tip: You can also click **Planned Failover** to perform planned failover procedure for a failback.

» Click **Recover from latest Recovery Point** to use the latest recovery point for recovery.

» Click **Recover from specific point in time** to use a recovery point taken at a specific point in time for recovery.

b. Click **+ Add target clusters** if you want to failover to specific Nutanix clusters at the primary site.

If you do not add target clusters, the recovery plan recovers the guest VMs to any eligible cluster at the primary site.

> Note: If recovery plans contain VM categories, the VMs from those categories recover in the same category after an unplanned failover to the recovery site. Also, the recovery points keep generating at the recovery site for those recovered VMs. Since the VM count represents the number of recoverable VMs (calculated from recovery points), the recovered VMs and their newly created recovery points sum up. Their sum gives double the count of the originally recovered VMs on the recovery plans page. Now, if some VMs belonging to the given

category at the primary or recovery site are deleted, the VM count at both sites still stay the same until the recovery points of deleted VMs expire. For example, when two VMs have failed over, the recovery plans page at the recovery site shows four VMs (two replicated recovery points from source and two newly generated recovery points). The page shows four VMs even if the VMs are deleted from the primary or recovery site. The VM count synchronizes and becomes consistent in the subsequent RPO cycle conforming to the retention policy set in the protection policy (due to the expiration of recovery points).

5. Click **Failover**.
   The **Failover from Recovery Plan** dialog box lists the errors and warnings, if any, and allows you to stop or continue the failover operation. If there are no errors or you resolve the errors in step 6, the guest VMs failover to the recovery cluster.

6. If you see errors, do the following.

   - To review errors or warnings, click **View Details** in the description.

     Resolve the error conditions and then restart the failover procedure.

   - Select one of the following.

     - To stop the failover operation, click **Abort**.
     - To continue the failover operation despite the warnings, click **Execute Anyway**.

       Note: You cannot continue the failover operation when the validation fails with errors.

   Note: To avoid conflicts when the primary site becomes active after the failover, shut down the guest VMs associated with this recovery plan. Manually power off the guest VMs on either primary or recovery site after the failover is complete. You can also block the guest VMs associated with this recovery plan through the firewall.

   Note:

   The entities of AHV/ESXi clusters recover at a different path on the ESXi clusters if their files conflict with the existing files on the recovery ESXi cluster. For example, there is a file name conflict if a VM (VM1) migrates to a recovery cluster that already has a VM (VM1) in the same container.

   However, the entities recover at a different path with `VmRecoveredAtAlternatePath` alert only if the following conditions are met.

   - Prism Element running on both the primary and the recovery Nutanix clusters are of version 5.17 or newer.
   - A path for the entity recovery is not defined while initiating the failover operation.
   - The protected entities do not have shared disk/s.

   If these conditions are not satisfied, the failover operation fails.

Monitoring a Failover Operation (Leap)

After you trigger a failover operation, you can monitor failover-related tasks. To monitor a failover, perform the following procedure at the recovery site. If you have two recovery sites for DR, perform the procedure at the site where you trigger the failover.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Click the name of the recovery plan for which you triggered failover.
4. Click the **Tasks** tab.
   The left pane displays the overall status. The table in the details pane lists all the running tasks and their individual statuses.

Self-Service Restore

The self-service restore (also known as file-level restore) feature allows you to do a self-service data recovery from the Nutanix data protection recovery points with minimal intervention.

You must deploy NGT 2.0 on the VM to enable self-service restore from Prism Central. For more information about enabling and mounting NGT, see *Enabling and Mounting Nutanix Guest Tools* in the *Prism Web Console Guide*. When you enable self-service restore and attach a disk by logging into the VM, you can recover files within the guest OS. If you fail to detach the disk from the VM, the disk is detached automatically from the VM after 24 hours.

> Note:
>
> - You can enable self-service restore for a VM through a web interface or nCLI.
> - NGT performs the in-guest actions.
> - Self-service restore supports only Asynchronous replication schedules (1 hour or greater RPO). For more information about Asynchronous replication schedules, see Protection with Asynchronous Replication Schedule and DR (Leap) on page 40.
> - You can enable self-service restore for data recovery in the same or different Prism Centrals.

> Note: Xi Cloud Services do not support self-service restore.

Self-Service Restore Requirements

The requirements of self-service restore of Windows and Linux VMs are as follows.

General Requirements of Self-Service Restore

The following are the general requirements of self-service restore. Ensure that you meet the requirements before configuring self-service restore for guest VMs.

License Requirements

AOS Ultimate. For more information about the features available with AOS Starter license, see Software Options.

Hypervisor Requirements

Two AHV or ESXi clusters, each registered to the same or different Prism Centrals.

- AHV (running AOS 5.18 or newer)

  The on-prem clusters must be running the version of AHV that comes bundled with the supported version of AOS.

- ESXi (running AOS 5.18 or newer)

  The on-prem clusters must be running on version ESXi 6.5 GA or newer.

Nutanix Software Requirements

Prism Centrals and their registered on-prem clusters (Prism Elements) must be running the following versions of AOS.

- AOS 5.18 or newer with AHV.
- AOS 5.18 or newer with ESXi.

- You have installed NGT 2.0 or newer. For more information about enabling and mounting NGT, see *Enabling and Mounting Nutanix Guest Tools* in the *Prism Web Console Guide*.
- You have set disk.enableUUID=true in the .vmx file for the guest VMs running on ESXi.
- You have configured Nutanix recovery points by adding guest VM to an Asynchronous protection policy.
- You have attached an IDE/SCSI or SATA disk

Requirements for Guest VMs Running Windows OS

The following are the specific requirements of self-service restore for guest VMs running Windows OS. Ensure that you meet the requirements before proceeding.

- You have enough logical drive letters to bring the disk online.
- You have one of the following Windows OS as the guest OS.
  - Windows Server 2008 R2 or newer
  - Windows 7 through Windows 10

Requirements for Guest VMs Running Linux OS

The following are the specific requirements of self-service restore for guest VMs running Linux OS. Ensure that you meet the requirements before proceeding.

- You have appropriate file systems to recover. Self-service restore supports only extended file systems (ext2, ext3, and ext4) and XFS file systems.
- Logical Volume Manager (LVM) disks for which the volume group corresponds to only a single physical disk are mounted.

- You have one of the following Linux OS as the guest OS.
  - CentOS 6.5 through 6.9 and 7.0 through 7.3
  - Red Hat Enterprise Linux (RHEL) 6.5 through 6.9 and 7.0 through 7.3
  - Oracle Linux 6.5 and 7.0
  - SUSE Linux Enterprise Server (SLES) 11 SP1 through 11 SP4 and 12 SP1 through 12 SP3
  - Ubuntu 14.04 for both AHV and ESXi
  - Ubuntu 16.10 for AHV only

Self-Service Restore Limitations

The limitations of self-service restore of Windows and Linux VMs are as follows.

General Limitations of Self-Service Restore

The following are the general limitations of self-service restore.

- Volume groups are not supported.
- Snapshots created in AOS 4.5 or later releases are only supported.
- PCI and delta disks are not supported.

Limitations of VMs Running Windows OS

The following are the specific limitations of self-service restore for guest VMs running Windows OS.

- File systems. Self-service restore does not support dynamic disks consisting of NTFS on simple volumes, spanned volumes, striped volumes, mirrored volumes, and RAID-5 volumes.
- Only 64-bit OSes are supported.
- Self-service restore does not support disks created as Microsoft Storage Space devices by using Microsoft Windows Server 2016 or newer.

Limitations of VMs Running Linux OS

Whenever the snapshot disk has an inconsistent filesystem (as indicated by the fsck check), the disk is only attached and not mounted.

Enabling Self-Service Restore

After enabling NGT for a guest VM, you can enable the self-service restore for that guest VM. Also, you can enable the self-service restore for a guest VM while you are installing NGT on that guest VM. For more information, see *Enabling and Mounting Nutanix Guest Tools* in the *Prism Web Console Guide*.

Before you begin

Ensure that you have installed and enabled NGT 2.0 on the guest VM.

About this task

To enable self-service restore, perform the following procedure.

Procedure

1. Log on to Prism Central.
2. Click the hamburger icon at the top-left corner of the window. Go to **Virtual Infrastructure** > **VMs** in the left pane.
3. Select the guest VM where you want to enable self-service restore.
4. Click **Manage NGT Applications** from the **Actions** drop-down menu.

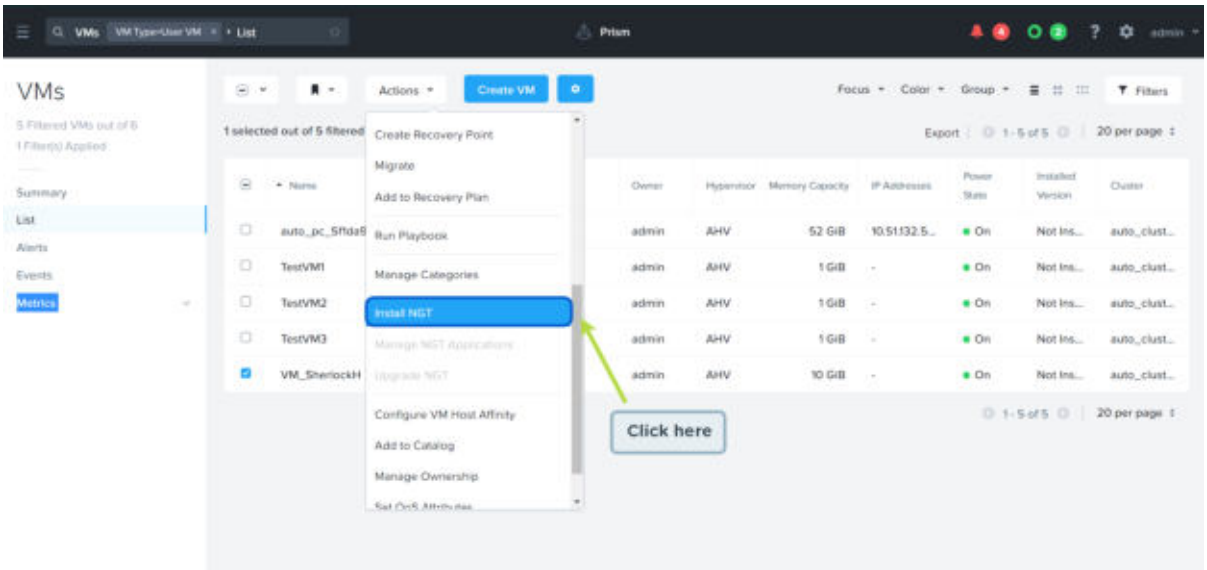


Figure 54: Enabling Self-Service Restore

> Note: If the guest VM does not have NGT installed, click **Install NGT** from the **Actions** drop-down menu and select to enable **Self Service Restore (SSR)**.

5. Click **Enable** below the **Self Service Restore (SSR)** panel.
6. Click **Confirm**.
   Self-service restore feature is enabled on the guest VM. You can now restore the desired files from the VM.

Self-Service Restore for Windows VMs

You can restore the desired files from the VM through the web interface or by using the ngtcli utility of self-service restore.

Restoring a File through Web Interface (Windows VM)

After you install NGT in the Windows guest VM, you can restore the desired files from the VM through the web interface.

Before you begin

Ensure that you have configured your Windows VM to use NGT. For more information, see *Installing NGT on Windows Machines* in the *Prism Web Console Guide*.

About this task

To restore a file in Windows guest VMs by using web interface, perform the following.

Procedure

1. Log in to the guest Windows VM by using administrator credentials.
2. Click the **Nutanix SSR** icon on the desktop.
3. Type the administrator credentials of the VM.

> Note: If you use:
>
> - NETBIOS domain name in username field (for example, domain\username), then you will be able to log on to SSR only if your account is explicitly added to Administrators group on the server. If username is added to any domain group, which is then added to Administrators group, then logon will fail. Also, you must type NETBIOS domain name in capital letters (domain name has to be written in the same way as you see in the output of command net localgroup administrators).
> - FQDN in username (for example domain.com\username), then you will only be able to logon if username user is a member of the domain admins group.

> Note: The snapshots that are taken for that day are displayed. You also have an option to select the snapshots for the week, month, and the year. In addition, you can also define a custom range of dates and select the snapshot.

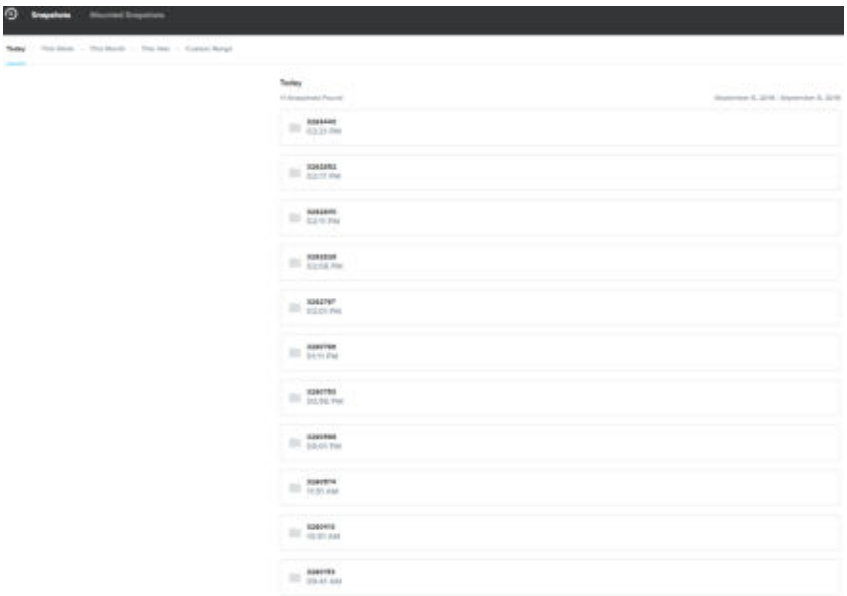


Figure 55: Snapshot Selection

4. Select the appropriate tab, **This Week**, **This Month**, **This Year**.
   You can also customize the selection by clicking **Custom Range** tab and selecting the date range in the **From** and **To** fields.
5. Select the check box of the disks that you want to attach from the snapshot.

6. Select **Mount** from the **Disk Action** drop-down menu.

Figure 56: Mounting of Disks

The selected disk or disks are mounted and the relevant disk label is displayed.

7. Go to the attached disk label drive in the VM and restore the desired files.

8. To view the list of all the mounted snapshots, select **Mounted Snapshots**.
   This page displays the original snapshot drive letters and its corresponding current drive letters. The original drive letters get assigned to the disk at the time of the snapshot. Mounted drive letters are on which the snapshotted disk is mounted right now.

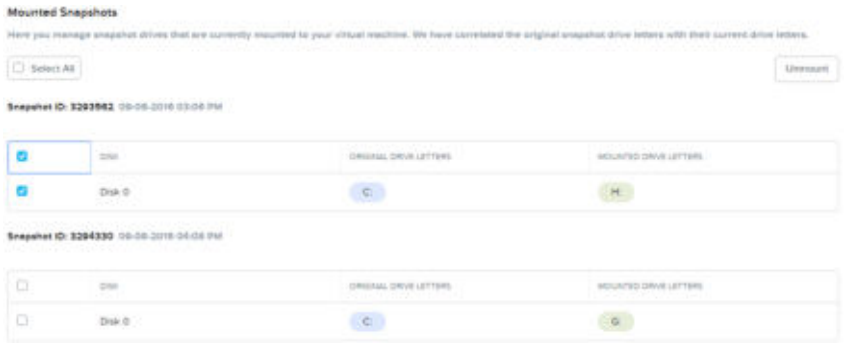


Figure 57: List of Mounted Snapshots

   a. To detach a disk, click the disk label and click **Unmount**.
      You can unmount all the disks at once by clicking **Select All** and then clicking **Unmount**.

9. To detach a disk, select the check box of the disk that you want to unmount and then from the **Disk Action** drop-down menu, select **Unmount**.

Restoring a File through Ngtcli (Windows VM)

After you install NGT in the Windows guest VM, you can restore the desired files from the VM through the ngtcli utility.

Before you begin

Ensure that you have configured your Windows VM to use NGT. For more information, see *Installing NGT on Windows Machines* in the *Prism Web Console Guide*.

About this task

To restore a file in Windows guest VMs by using ngtcli, perform the following.

Procedure

1. Log in to the guest Windows VM by using administrator credentials.
2. Open the command prompt as an administrator.
3. Go to the ngtcli directory in **Program Files** > **Nutanix**.

   ```
   > cd c:\Program Files\Nutanix\ngtcli
   ```

   > Tip:
   >
   > `> python ngtcli.py`
   >
   > creates a terminal with auto-complete.

4. Run the `ngtcli.cmd` command.
5. List the snapshots and virtual disks that are present for the guest VM.

   ```
   ngtcli> ssr ls-snaps
   ```

   The snapshot ID, disk labels, logical drives, and create time of the snapshot is displayed. You can use this information and take a decision to restore the files from the relevant snapshot that has the data.

   List the snapshots with a specific number.

   ```
   ngtcli> ssr ls-snaps snapshot-count=count_value
   ```

   Replace *`count_value`* with the number that you want to list.

6. Attach the disk from the snapshots.

   ```
   ngtcli> ssr attach-disk disk-label=disk_label snapshot-id=snap_id
   ```

   Replace *`disk_label`* with the name of the disk that you want to attach.

   Replace *`snap_id`* with the snapshot ID of the disk that you want to attach.

   For example, to attach a disk with snapshot ID 16353 and disk label scsi0:1, type the folllowing command.

   ```
   ngtcli> ssr attach-disk snapshot-id=16353 disk-label=scsi0:1
   ```

   After successfully running the command, a new disk with label gets attached to the guest VM.

   > Note: If sufficient logical drive letters are not present, bringing disks online action fails. In this case, you should detach the current disk, create enough free slots by detaching other self-service disks and reattach the disk again.

7. Go to the attached disk label drive and restore the desired files.
8. Detach a disk.

   ```
   ngtcli> ssr detach-disk attached-disk-label=attached_disk_label
   ```

   Replace *`attached_disk_label`* with the name of the disk that you want to attach.

   > Note: If the disk is not removed by the guest VM administrator, the disk is automatically removed after 24 hours.

9. View all the attached disks to the VM.

   ```
   ngtcli> ssr list-attached-disks
   ```

Self-Service Restore for Linux VMs

The Linux guest VM user with sudo privileges can restore the desired files from the VM through the web interface or by using the ngtcli utility.

Restoring a File through Web Interface (Linux VM)

After you install NGT in the Linux guest VM, you can restore the desired files from the VM through the web interface.

Before you begin

- Mount NGT for a VM. For more information about enabling NGT, see *Enabling and Mounting Nutanix Guest Tools* in the *Prism Web Console Guide*.
- Ensure that you have configured your Linux VM to use NGT. For more information, see *Prism Web Console Guide*.

About this task

To restore a file in Linux guest VMs by using web interface, perform the following.

Procedure

1. Log in to the guest Linux VM as a user with sudo privileges.
2. Click the **Nutanix SSR** icon on the desktop.
3. Type the root or sudo user credentials of the VM.
   The snapshots that are taken for that day is displayed. You also have an option to select the snapshots for the week, month, and the year. In addition, you can also define a custom range of dates and select the snapshot. For example, in the following figure snapshot taken on this month is displayed.

Figure 58: Snapshot Selection

4. Select the appropriate tab, **This Week**, **This Month**, **This Year**.
   You can also customize the selection by clicking **Custom Range** tab and selecting the date range in the **From** and **To** fields.
5. Select the check box of the disks that you want to attach from the snapshot.

6. Select **Mount** from the **Disk Action** drop-down menu.

   The selected disk or disks are mounted and the relevant disk label is displayed.

Figure 59: Mounting of Disks

7. Go to the attached disk label partitions in the VM and restore the desired files.

   > Note: If the disk gets updated between the snapshots, the restore process may not work as expected. If this scenario occurs, you need to contact support to help with the restore process.

8. To view the list of all the mounted snapshots, select **Mounted Snapshots**.

   This page displays the original snapshot drive letters and its corresponding current drive letters. The original drive letters gets assigned to the disk at the time of the snapshot. Mounted drive letters are on which the snapshotted disk is mounted right now.

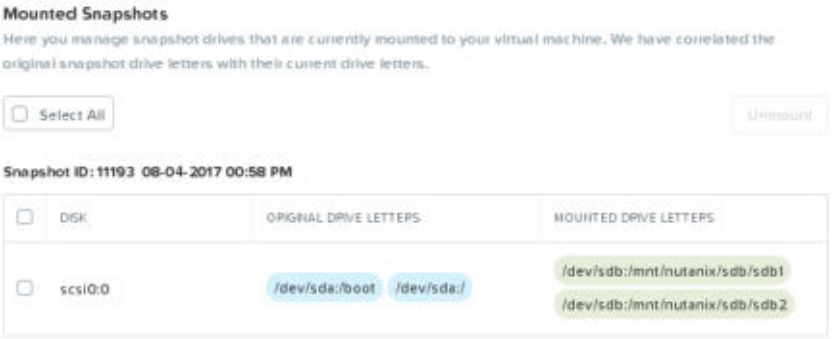


Figure 60: List of Mounted Snapshots

   a. To detach a disk, click the disk label and click **Unmount**.

      You can unmount all the disks at once by clicking **Select All** and then clicking **Unmount**.

9. To detach a disk, select the check box of the disk that you want to unmount and then from the **Disk Action** drop-down menu, select **Unmount**.

Restoring a File through Ngtcli (Linux VM)

After you install NGT in the Linux guest VM, you can restore the desired files from the VM through the ngtcli utility.

Before you begin

- Mount NGT for a VM. For more information about enabling NGT, see *Enabling and Mounting Nutanix Guest Tools* in the *Prism Web Console Guide*.

- Ensure that you have configured your Linux VM to use NGT. For more information, see *Prism Web Console Guide* in the *Prism Web Console Guide*.

About this task

To restore a file in Linux guest VMs by using ngtcli, perform the following.

Procedure

1. Log in to the guest Linux VM with sudo or root user credentials.
2. Go to the ngtcli directory.

   ```
   > cd /usr/local/nutanix/ngt/ngtcli
   ```

3. Run the `python ngtcli.py` command.

   > Tip: This command creates a terminal with auto-complete.

4. List the snapshots and virtual disks that are present for the guest VM.

   ```
   ngtcli> ssr ls-snaps
   ```

   The snapshot ID, disk labels, logical drives, and create time of the snapshot is displayed. You can use this information and take a decision to restore the files from the relevant snapshot that has the data.

   To list the snapshots with a specific number.

   ```
   ngtcli> ssr ls-snaps snapshot-count=count_value
   ```

   Replace *`count_value`* with the number that you want to list.

5. Attach the disk from the snapshots.

   ```
   ngtcli> ssr attach-disk disk-label=disk_label snapshot-id=snap_id
   ```

   Replace *`disk_label`* with the name of the disk that you want to attach.

   Replace *`snap_id`* with the snapshot ID of the disk that you want to attach.

   For example, to attach a disk with snapshot ID 1343 and disk label scsi0:2,

   ```
   ngtcli> ssr attach-disk snapshot-id=1343 disk-label=scsi0:2
   ```

   After successfully running the command, a new disk with new label is attached to the guest VM.

6. Go to the attached disk label partition and restore the desired files.

   > Note: If the disk gets updated between the snapshots, the restore process may not work as expected. If this scenario occurs, you need to contact support to help with the restore process.

7. Detach a disk.

```
ngtcli> ssr detach-disk attached-disk-label=attached_disk_label
```

Replace *attached_disk_label* with the name of the disk that you want to attach.

For example, to remove the disk with disk label scsi0:3, type the following command.

```
ngtcli> ssr detach-disk attached-disk-label=scsi0:3
```

> Note: If the disk is not removed by the guest VM administrator, the disk is automatically removed after 24 hours.

8. View all the attached disks to the VM.

```
ngtcli> ssr list-attached-disks
```

Protection with NearSync Replication Schedule and DR (Leap)

NearSync replication enables you to protect your guest VMs with an RPO of as low as 1 minute. A protection policy with a NearSync replication creates a recovery point in a minutely time interval (between 1–15 minutes), and replicates it to the recovery availability zones (sites) for High Availability. For guest VMs protected with NearSync replication schedule, you can perform disaster recovery (DR) to a different Nutanix cluster at same or different sites. In addition to DR to Nutanix clusters of the same hypervisor type, you can also perform cross-hypervisor disaster recovery (CHDR)—disaster recovery from AHV clusters to ESXi clusters, or from ESXi clusters to AHV clusters.

> Note: Nutanix provides multiple DR solutions to secure your environment. See Nutanix Disaster Recovery Solutions on page 11 for the detailed representation of the DR offerings of Nutanix.

The following are the advantages of protecting your guest VMs with a NearSync replication schedule.

- Protection for the mission-critical applications. Securing your data with minimal data loss if there is a disaster, and providing you with more granular control during the recovery process.
- No minimum network latency or distance requirements.
- Low stun time for guest VMs with heavy I/O applications.

  Stun time is the time of application freeze when the recovery point is taken.
- Allows resolution to a disaster event in minutes.

To implement the NearSync feature, Nutanix has introduced a technology called lightweight snapshots (LWSs). LWS recovery points are created at the metadata level only, and they continuously replicate incoming data generated by workloads running on the active clusters. LWS recovery points are stored in the LWS store, which is allocated on the SSD tier. When you configure a protection policy with a NearSync replication schedule, the system allocates the LWS store automatically.

> Note: The maximum LWS store allocation for each node is 360 GB. For the hybrid systems, it is 7% of the SSD capacity on that node.

Transitioning in and out of NearSync

When you create a NearSync replication schedule, the schedule remains an hourly schedule until its transition into a minutely schedule is complete.

To transition into NearSync (minutely) replication schedule, initial seeding of the recovery site with the data is performed, the recovery points are taken on an hourly basis, and replicated to the recovery site. After the system determines that the recovery points containing the seeding data have replicated within a specified amount of time (default is an hour), the system automatically transitions the replication schedule into NearSync schedule depending on the bandwidth and the change rate. After you transition into the NearSync replication schedule, you can see the configured minutely recovery points in the web interface.

The following are the characteristics of the process.

- Until you are transitioned into NearSync replication schedule, you can see only the hourly recovery points in Prism Central.
- If for any reason, a guest VM transitions out of NearSync replication schedule, the system raises alerts in the **Alerts** dashboard, and the minutely replication schedule transitions out to the hourly replication schedule. The system continuously tries to get into the minutely replication schedule that you have configured. If the transition is successful, the replication schedule automatically transitions back into NearSync, and alerts specific to this condition are raised in the **Alerts** dashboard.

To transition out of the NearSync replication schedule, you can do one of the following.

- Delete the NearSync replication schedule that you have configured.
- Update the NearSync replication schedule to use an hourly RPO.
- Unprotect the guest VMs.

> Note: There is no transitioning out of the NearSync replication schedule on the addition or deletion of a guest VM.

Repeated transitioning in and out of NearSync replication schedule can occur because of the following reasons.

- LWS store usage is high.
- The change rate of data is high for the available bandwidth between the primary and the recovery sites.
- Internal processing of LWS recovery points is taking more time because the system is overloaded.

Retention Policy

Depending on the RPO (1–15 minutes), the system retains the recovery points for a specific time period. For a NearSync replication schedule, you can configure the retention policy for days, weeks, or months on both the primary and recovery sites instead of defining the number of recovery points you want to retain. For example, if you desire an RPO of 1 minute and want to retain the recovery points for 5 days, the retention policy works in the following way.

- For every 1 minute, a recovery point is created and retained for a maximum of 15 minutes.

> Note: The recent 15 recovery points are only visible in Prism Central and are available for the recovery operation.

- For every hour, a recovery point is created and retained for 6 hours.
- One daily recovery point is created and retained for 5 days.

You can also define recovery point retention in weeks or months. For example, if you configure a 3-month schedule, the retention policy works in the following way.

- For every 1 minute, a recovery point is created and retained for 15 minutes.
- For every hour, a recovery point is created and retained for 6 hours.
- One daily recovery point is created and retained for 7 days.
- One weekly recovery point is created and retained for 4 weeks.
- One monthly recovery point is created and retained for 3 months.

> Note:
>
> - You can define different retention policies on the primary and recovery sites.
> - The system retains subhourly and hourly recovery points for 15 minutes and 6 hours respectively. Maximum retention time for days, weeks, and months is 7 days, 4 weeks, and 12 months respectively.
> - If you change the replication schedule from an hourly schedule to a minutely schedule (Asynchronous to NearSync), the first recovery point is not created according to the new schedule. The recovery points are created according to the start time of the old hourly schedule (Asynchronous). If you want to get the maximum retention for the first recovery point after modifying the schedule, update the start time accordingly for NearSync.

NearSync Replication Requirements (Leap)

The following are the specific requirements for protecting your guest VMs with NearSync replication schedule. Ensure that you meet the following requirements in addition to the general requirements of Leap.

For more information about the general requirements of Leap, see Leap Requirements on page 20.

For information about node, disk and Foundation configurations required to support NearSync replication schedules, see On-Prem Hardware Resource Requirements on page 15.

Hypervisor Requirements

AHV or ESXi

- The AHV clusters must be running on version 20190916.189 or newer.
- The ESXi clusters must be running on version ESXi 6.5 GA or newer.

Nutanix Software Requirements

Each on-prem site must have a Leap enabled Prism Central instance.

The primary and recovery Prism Centrals and their registered Nutanix clusters must be running the following versions of AOS.

- AOS 5.17.1 or newer for DR to different Nutanix clusters at the same site.
- AOS 5.17 or newer for DR to Nutanix clusters at the different sites.

Cross Hypervisor Disaster Recovery (CHDR) Requirements

Guest VMs protected with NearSync replication schedule support cross-hypervisor disaster recovery. You can perform failover (DR) to recover guest VMs from AHV clusters to ESXi clusters or guest VMs from ESXi clusters to AHV clusters by considering the following requirements.

- Both the primary and the recovery Nutanix clusters must be running AOS 5.18 or newer.
- Install and configure Nutanix Guest Tools (NGT) on all the guest VMs. For more information, see *Enabling and Mounting Nutanix Guest Tools* in *Prism Web Console Guide*.

  NGT configures the guest VMs with all the required drivers for VM portability. For more information about general NGT requirements, see *Nutanix Guest Tools Requirements and Limitations* in *Prism Web Console Guide*.
- CHDR supports guest VMs with flat files only.
- CHDR supports IDE/SCSI disks only.

  > Tip: From AOS 5.19.1, CHDR supports SATA disks also.

- For all the non-boot SCSI disks of Windows guest VMs, set the SAN policy to OnlineAll so that they come online automatically.
- In vSphere 6.7, guest VMs are configured with UEFI secure boot by default. Upon CHDR to an AHV cluster, these guest VMs do not start if the host does not support the UEFI secure boot feature. For more information about supportability of UEFI secure boot on Nutanix clusters, see https://portal.nutanix.com/page/documents/compatibility-interoperability-matrix/guestos.
- For information about operating systems that support UEFI and Secure Boot, see UEFI and Secure Boot Support for CHDR on page 227.
- Nutanix does not support vSphere inventory mapping (for example, VM folder and resource pools) when protecting workloads between VMware clusters.
- Nutanix does not support vSphere snapshots or delta disk files. If you have delta disks attached to a VM and you proceed with failover, you get a validation warning and the VM does not recover. Contact Nutanix Support for assistance.

Table 22: Operating Systems Supported for CHDR (Asynchronous Replication)

Operating System	Version	Requirements and limitations
Windows	• Windows 2008 R2 or newer versions • Windows 7 or newer versions	• Only 64-bit operating systems are supported.
Linux	• CentOS 6.5 and 7.0 • RHEL 6.5 or newer and RHEL 7.0 or newer. • Oracle Linux 6.5 and 7.0 • Ubuntu 14.04	• SLES operating system is not supported.

Additional Requirements

- Both the primary and the recovery Nutanix clusters must be of minimum three-nodes.

- The recovery site container must have as much space as the protected VMs working size set of the primary site. For example, if you are protecting a VM that is using 30 GB of space on the container of the primary site, the same amount of space is required on the recovery site container.

NearSync Replication Limitations (Leap)

Consider the following specific limitations before protecting your guest VMs with NearSync replication schedule. These limitations are in addition to the general limitations of Leap.

For information about the general limitations of Leap, see Leap Limitations on page 24.

- All files associated with the VMs running on ESXi must be located in the same folder as the VMX configuration file. The files not located in the same folder as the VMX configuration file might not recover on a recovery cluster. On recovery, the guest VM with such files fails to start with the following error message. `Operation failed: InternalTaskCreationFailure: Error creating host specific VM change power state task. Error: NoCompatibleHost: No host is compatible with the virtual machine`
- Deduplication enabled on storage containers having guest VMs protected with NearSync replication schedule lowers the replication speed.
- Cross hypervisor disaster recovery (CHDR) does not preserve hypervisor-specific properties (for example, multi-writer flags, independent persistent and non-persistent disks, changed block tracking (CBT), PVSCSI disk configurations).
- On CHDR, NearSync replication schedules do not support retrieving recovery points from the recovery sites.

  For example, if you have 1 day retention at the primary site and 5 days retention at the recovery site, and you want to go back to a recovery point from 5 days ago. NearSync replication schedule does not support replicating 5 days retention back from the recovery site to the primary site.

Creating a Protection Policy with a NearSync Replication Schedule (Leap)

To protect the guest VMs in a minutely replication schedule, configure a NearSync replication schedule while creating the protection policy. The policy takes recovery points of the protected guest VMs in the specified time intervals (1–15 minutes) and replicates them to the recovery availability zone (site) for High Availability. To maintain the efficiency of minutely replication, the protection policy allows you to configure a NearSync replication schedule to only one recovery site. When creating a protection policy, you can specify only VM categories. If you want to include VMs individually, you must first create the protection p policy—which can also include VM categories and then include the VMs individually in the protection policy from the **VMs** page.

Before you begin

Ensure that the primary and the recovery AHV or ESXi clusters at the same or different sites are NearSync capable. A cluster is NearSync capable if the capacity of each SSD in the cluster is at least 1.2 TB.

See NearSync Replication Requirements (Leap) on page 102 and NearSync Replication Limitations (Leap) on page 104 before you start.

About this task

To create a protection policy with a NearSync replication schedule, do the following at the primary site. You can also create a protection policy at the recovery site. Protection policies you create or update at a recovery site synchronize back to the primary site.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Protection Policies** in the left pane.

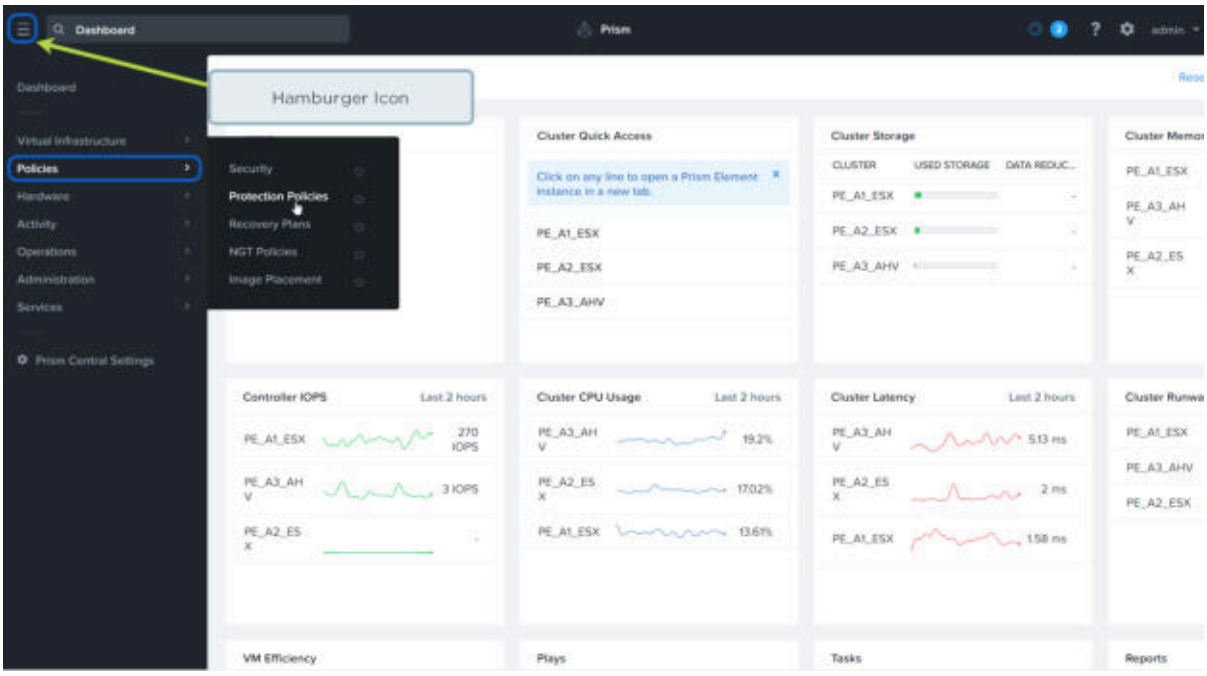


Figure 61: Protection Policy Configuration: Protection Policies

3. Click **Create Protection Policy**.
   Specify the following information in the Create Protection Policy window.

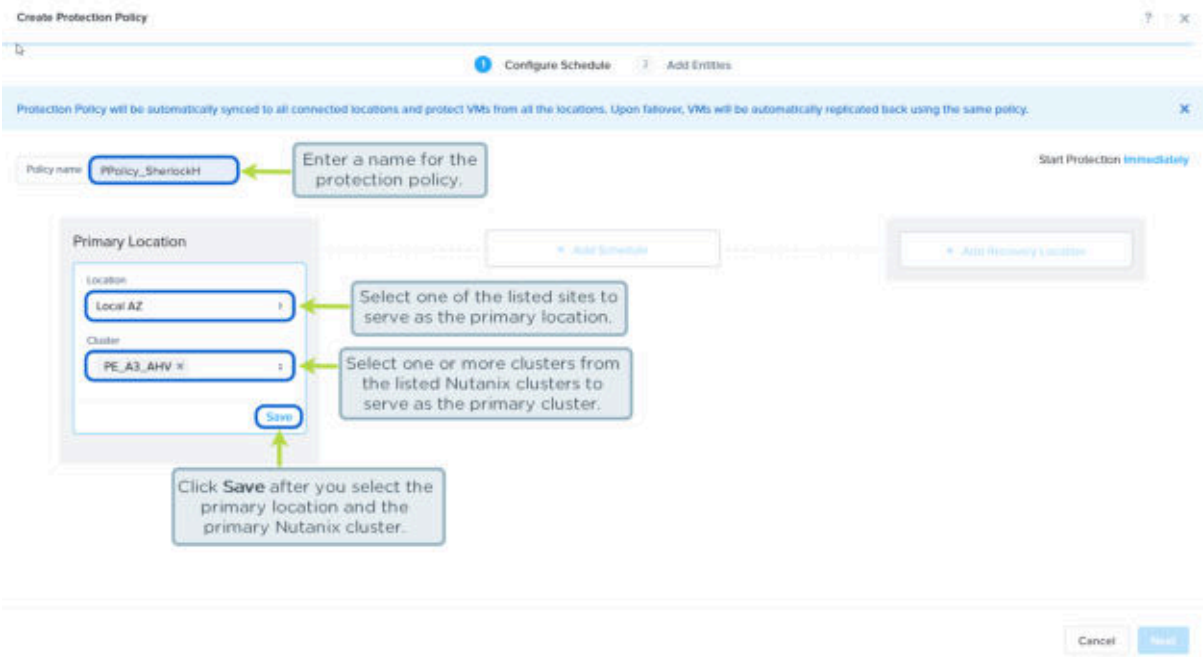


Figure 62: Protection Policy Configuration: Select Primary Location

a. **Policy name**: Enter a name for the protection policy.

> Caution: The name can be of only alphanumeric, dot, dash, and underscore characters.

b. In the **Primary Location** pane, specify the following information.

- 1. **Location**: From the drop-down list, select an availability zone (site) that hosts the guest VMs to protect.

  The drop-down lists all the sites paired with the local site. Local AZ represents the local site (Prism Central). For your primary site, you can check either the local site or a non-local site.

  2. **Cluster**: From the drop-down list, select the cluster that hosts the guest VMs to protect.

     The drop-down lists all the Nutanix clusters registered to Prism Central representing the selected site. If you want to protect the guest VMs from multiple Nutanix clusters in the same protection policy, select the clusters that host those guest VMs. All Clusters protects the guest VMs of all Nutanix clusters registered to Prism Central.

  3. Click **Save**.

     Clicking **Save** activates the **Recovery Location** pane. After saving the primary site configuration, you can optionally add a local schedule to retain the recovery points at the primary site.

  4. Click **+ Add Local Schedule** if you want to retain recovery points locally in addition to retaining recovery points in a replication schedule (step d.iv). For example, you can create a local schedule to retain 15 minute recovery points locally and also

an hourly replication schedule to retain recovery points and replicate them to a recovery site every 2 hours. The two schedules apply differently on the guest VMs.

Specify the following information in the Add Schedule window.

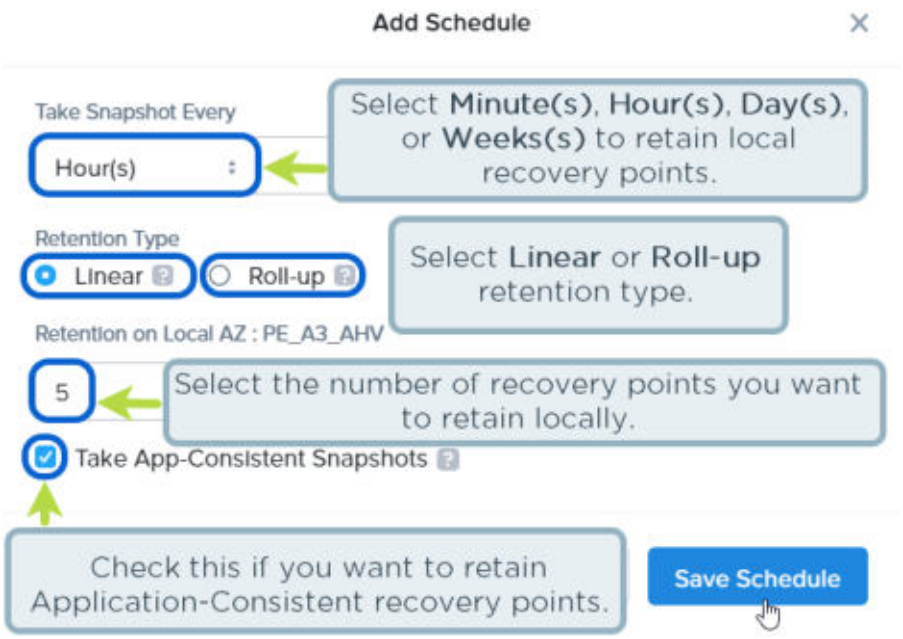


Figure 63: Protection Policy Configuration: Add Local Schedule

1. **Take Snapshot Every**: Specify the frequency in minutes, hours, days, or weeks at which you want the recovery points to be taken locally.
2. **Retention Type**: Specify one of the following two types of retention policy.

   - **Linear**: Implements a simple retention scheme at the local site. If you set the retention number to n, the local site retains the n recent recovery points.

     When you enter the frequency in minutes, the system selects the **Roll-up** retention type by default because minutely recovery points do not support **Linear** retention types.

   - **Roll-up**: Rolls up the recovery points into a single recovery point at the local site.

     For more information about the roll-up recovery points, see step d.iii.
3. **Retention on Local AZ:PE_A3_AHV**: Specify the retention number for the local site.
4. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Irrespective of the local or replication schedules, the recovery points are of the specified type. If you check **Take App-Consistent Recovery Point**, the recovery points generated are application-consistent and if you do not check **Take App-Consistent Recovery Point**, the recovery points generated are crash-consistent.

If the time in the local schedule and the replication schedule match, the single recovery point generated is application-consistent.

> Note: See Application-consistent Recovery Point Conditions and Limitations on page 55 before you take application-consistent snapshot.

5. Click **Save Schedule**.

c. In the **Recovery Location** pane, specify the following information.

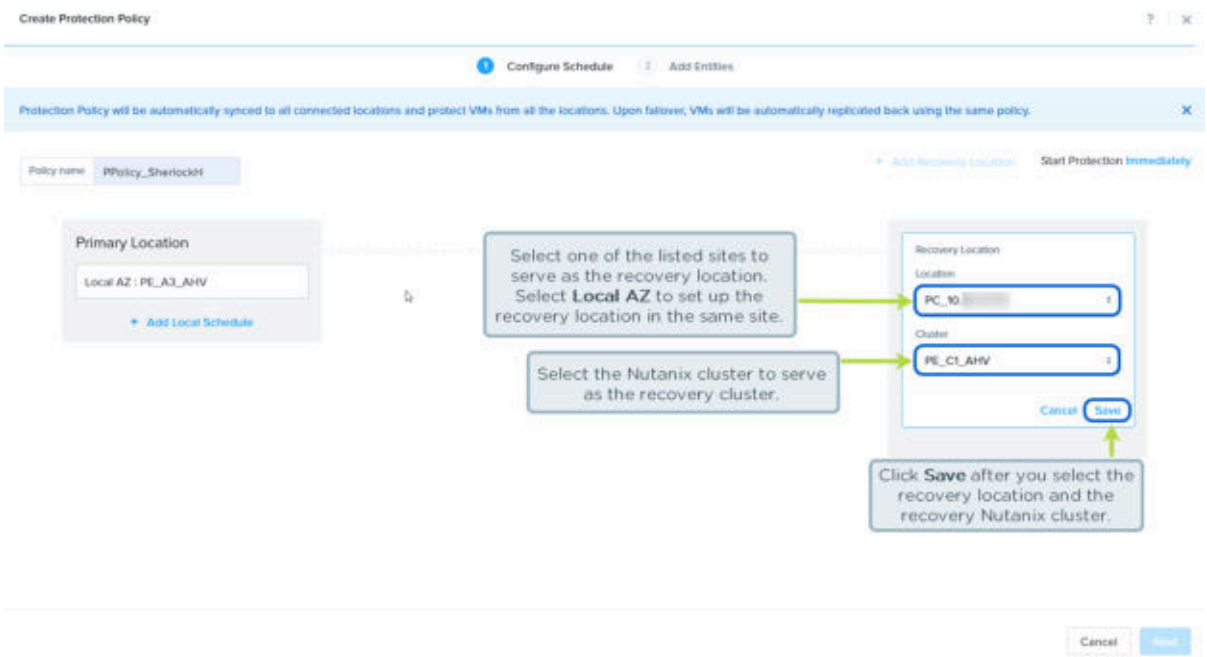


Figure 64: Protection Policy Configuration: Select Recovery Location

- 1. **Location**: From the drop-down list, select the availability zone (site) where you want to replicate the recovery points.

  The drop-down lists all the sites paired with the local site. Local AZ represents the local site (Prism Central). Select Local AZ if you want to configure DR to a different Nutanix cluster at the same site.

  If you do not select a site, local recovery points that are created by the protection policy do not replicate automatically. You can, however, replicate the recovery points manually and use recovery plans to recover the guest VMs. For more information, see Protection and Manual DR (Leap) on page 143.

  2. **Cluster**: From the drop-down list, select the cluster where you want to replicate the recovery points.

  The drop-down lists all the Nutanix clusters registered to Prism Central representing the selected site. You can select one cluster at the recovery site. To maintain the efficiency of minutely replication, a protection policy allows you to configure only one recovery site for a NearSync replication schedule. However, you can add another Asynchronous replication schedule for replicating recovery points to the

same or different sites. For more information to add another recovery site with a replication schedule, see step e.

> Note: Selecting auto-select from the drop-down list replicates the recovery points to any available cluster at the recovery site. Select auto-select from the drop-down list only if all the clusters at the recovery site are NearSync capable and are up and running. A cluster is NearSync capable if the capacity of each SSD in the cluster is at least 1.2 TB. All-flash clusters do not have any specific SSD sizing requirements.

> Caution: If the primary Nutanix cluster contains an IBM POWER Systems server, you can replicate recovery points to an on-prem site only if that on-prem site contains an IBM Power Systems server.

3. Click **Save**.

   Clicking **Save** activates the **+ Add Schedule** button between the primary and the recovery site. After saving the recovery site configuration, you can optionally add a local schedule to retain the recovery points at the recovery site.

4. Click **+ Add Local Schedule** if you want to retain recovery points locally in addition to retaining recovery points in a replication schedule (step d.iv). For example, you can create a local schedule to retain one hourly recovery points locally to supplement the hourly replication schedule. The two schedules apply differently on the guest VMs after failover, when the recovery points replicate back to the primary site.

   Specify the following information in the Add Schedule window.

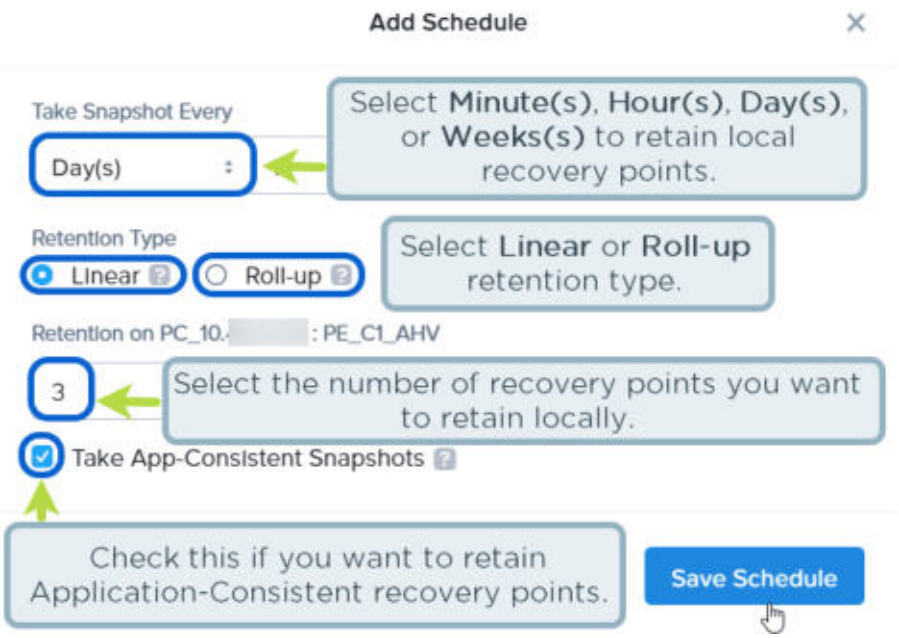


Figure 65: Protection Policy Configuration: Add Local Schedule

1. **Take Snapshot Every**: Specify the frequency in minutes, hours, days, or weeks at which you want the recovery points to be taken locally.

2. **Retention Type**: Specify one of the following two types of retention policy.

   - **Linear**: Implements a simple retention scheme at the local site. If you set the retention number to n, the local site retains the n recent recovery points.

     When you enter the frequency in minutes, the system selects the **Roll-up** retention type by default because minutely recovery points do not support **Linear** retention types.

   - **Roll-up**: Rolls up the recovery points into a single recovery point at the local site.

     For more information about the roll-up recovery points, see step d.iii.

3. **Retention on 10.xx.xx.xxx:PE_C1_AHV**: Specify the retention number for the local site.
4. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Irrespective of the local or replication schedules, the recovery points are of the specified type. If you check **Take App-Consistent Recovery Point**, the recovery points generated are application-consistent and if you do not check **Take App-Consistent Recovery Point**, the recovery points generated are crash-consistent.

If the time in the local schedule and the replication schedule match, the single recovery point generated is application-consistent.

5. Click **Save Schedule**.

d. Click **+ Add Schedule** to add a replication schedule between the primary and the recovery site.

   Specify the following information in the Add Schedule window. The window auto-populates the Primary Location and Recovery Location that you have selected in step b and step c.

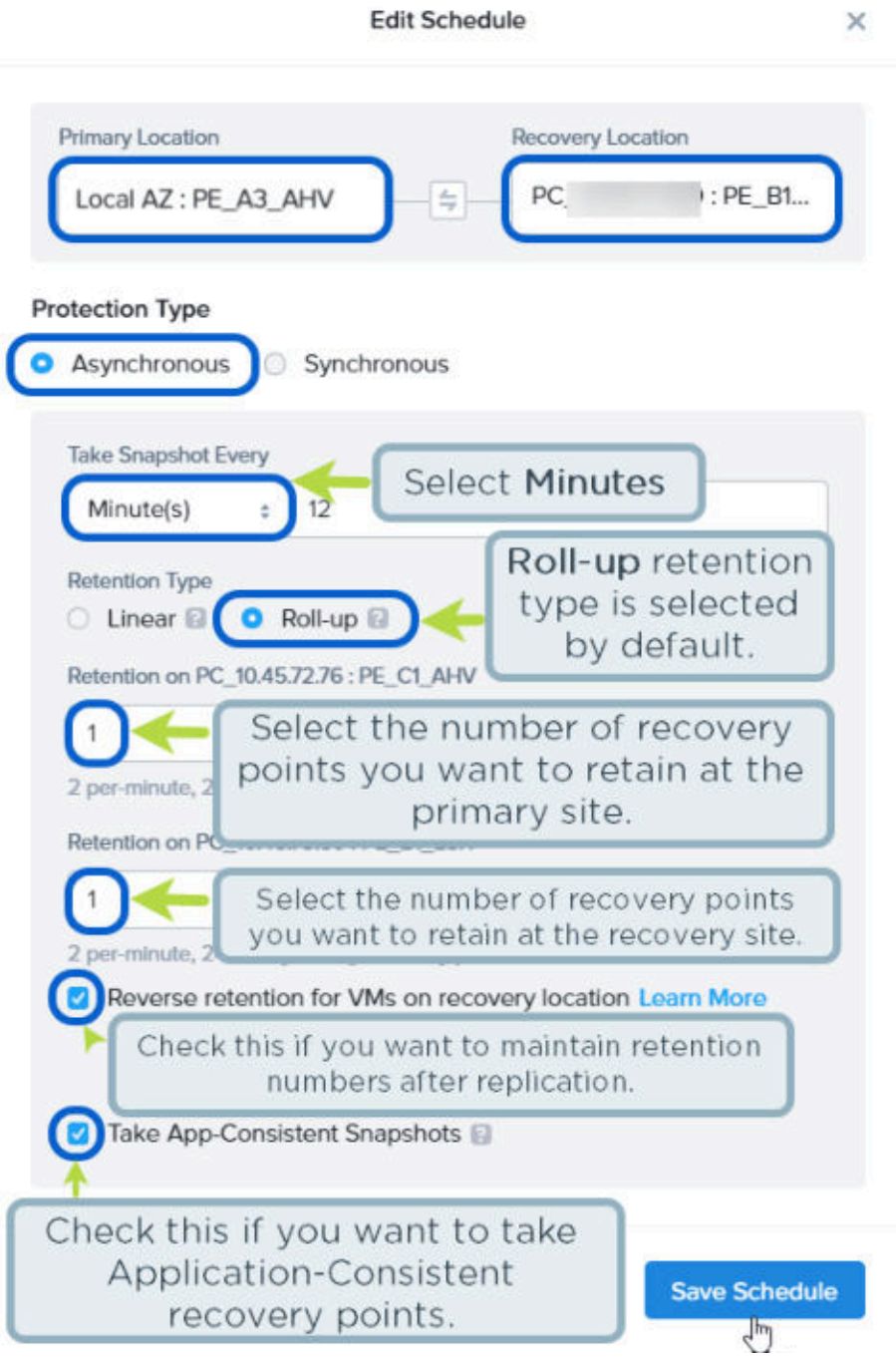


Figure 66: Protection Policy Configuration: Add Schedule (NearSync)

- 1. **Protection Type**: Click **Asynchronous**.
  2. **Take Snapshot Every**: Specify the frequency in minutes (anywhere between 1-15 minutes) at which you want the recovery points to be taken.

     The specified frequency is the RPO. For more information about RPO, see Leap Terminology on page 8.
  3. **Retention Type**: When you enter the frequency in minutes in step ii, the system selects the **Roll-up** retention type by default because NearSync replication schedules do not support Linear retention types.

     Roll-up retention type rolls up the recovery points as per the RPO and retention period into a single recovery point at a site. For example, if you set the RPO to 1 hour, and the retention time to 5 days, the 24 oldest hourly recovery points roll up into a single daily recovery point (one recovery point = 24 hourly recovery points) after every 24 hours. The system keeps one day (of rolled-up hourly recovery points) and 4 days of daily recovery points.

     > Note:
     >
     > - If the retention period is n days, the system keeps 1 day of RPO (rolled-up hourly recovery points) and n-1 days of daily recovery points.
     > - If the retention period is n weeks, the system keeps 1 day of RPO, 1 week of daily and n-1 weeks of weekly recovery points.
     > - If the retention period is n months, the system keeps 1 day of RPO, 1 week of daily, 1 month of weekly, and n-1 months of monthly recovery points.
     > - If the retention period is n years, the system keeps 1 day of RPO, 1 week of daily, 1 month of weekly, and n-1 months of monthly recovery points.

     > Note: The recovery points that are used to create a rolled-up recovery point are discarded.

     > Tip: Use roll-up retention policies for anything with a longer retention period. Roll-up policies are more flexible and automatically handle recovery point aging/pruning while still providing granular RPOs for the first day.
  4. To specify the retention number for the primary and recovery sites, do the following.

     - **Retention on Local AZ: PE_A3_AHV**: Specify the retention number for the primary site.

       This field is unavailable if you do not specify a recovery location.
     - **Retention on 10.xx.xx.xxx:PE_C1_AHV**: Specify the retention number for the recovery site.
  5. If you want to enable reverse retention of the recovery points, check **Reverse retention for VMs on recovery location**.

     Reverse retention maintains the retention numbers of recovery points even after failover to a recovery site in the same or different availability zones. For example, if you retain two recovery points at the primary site and three recovery points at the recovery site, and you enable reverse retention, a failover event does not change the initial retention numbers when the recovery points replicate back to the primary site. The recovery site still retains two recovery points while the primary site retains three recovery points. If you do not enable reverse retention, a failover event changes the

initial retention numbers when the recovery points replicate back to the primary site. The recovery site retains three recovery points while the primary site retains two recovery points.

Maintaining the same retention numbers at a recovery site is required if you want to retain a particular number of recovery points, irrespective of where the guest VM is after its failover.

6. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Application-consistent recovery points ensure that application consistency is maintained in the replicated recovery points. For application-consistent recovery points, install NGT on the guest VMs running on AHV clusters. For guest VMs running on ESXi clusters, you can take application-consistent recovery points without installing NGT, but the recovery points are hypervisor-based, and leads to VM stuns (temporary unresponsive VMs) after failover to the recovery sites.

   > Caution: Application-consistent recovery points fail for EFI-boot enabled Windows 2019 VM running on ESXi when NGT is not installed. Nutanix recommends installing NGT on guest VMs running on ESXi also.

7. Click **Save Schedule**.

e. Click **+ Add Recovery Location** if you want to add an additional recovery site for the guest VMs in the protection policy.

   » To add an on-prem site for recovery, see Protection and DR between On-Prem Sites (Leap) on page 18

   » To add Xi Cloud Services for recovery, see Protection and DR between On-Prem Site and Xi Cloud Service (Xi Leap) on page 154.

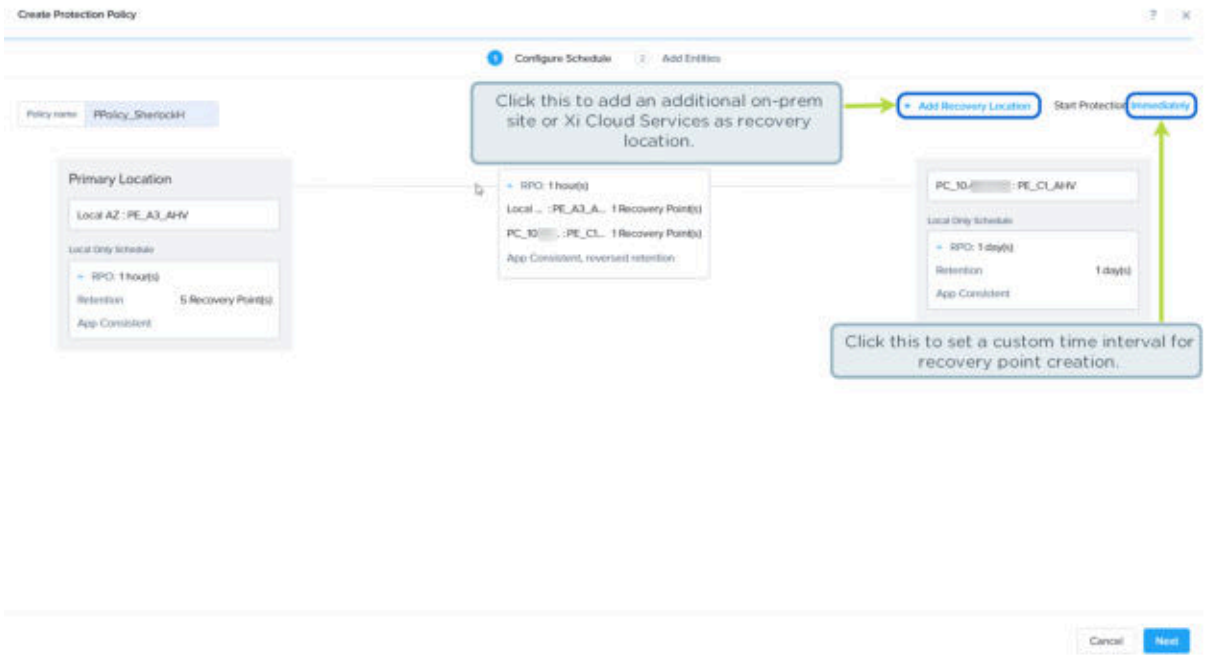


Figure 67: Protection Policy Configuration: Additional Recovery Location

f. Click **+ Add Schedule** to add a replication schedule between the primary site and the additional recovery site you specified in step e.

   The Add Schedule window shows that auto-populates the Primary Location and the additional Recovery Location. Perform step d again to add the replication schedule.

   By default, recovery point creation begins immediately after you create the protection policy. If you want to specify when recovery point creation must begin, click **Immediately** at the top-right corner, and then, in the **Start Time** dialog box, do the following.

   - 1. Click **Start protection at specific point in time**.
     2. Specify the time at which you want to start taking recovery points.
     3. Click **Save**.

g. Click **Next**.

   Clicking **Next** shows a list of VM categories where you can optionally check one or more VM categories to protect in the protection policy. DR configurations using Leap allows you to protect a guest VM by using only one protection policy. Therefore, VM categories specified in another protection policy are not in the list. If you protect a guest VM in another protection policy by specifying the VM category of the guest VM (category-based inclusion), and if you protect the guest VM from the **VMs** page in this policy (individual inclusion), the individual inclusion supersedes the category-based inclusion.

Effectively, only the protection policy that protected the individual guest VM protects the guest VM.

For example, the guest VM *VM_SherlockH* is in the category *Department:Admin*, and you add this category to the protection policy named *PP_AdminVMs*. Now, if you add *VM_SherlockH* from the **VMs** page to another protection policy named *PP_VMs_UK*, *VM_SherlockH* is protected in *PP_VMs_UK* and unprotected from *PP_AdminVMs*.

h. If you want to protect the guest VMs category wise, check the VM categories that you want to protect from the list and click **Add**.

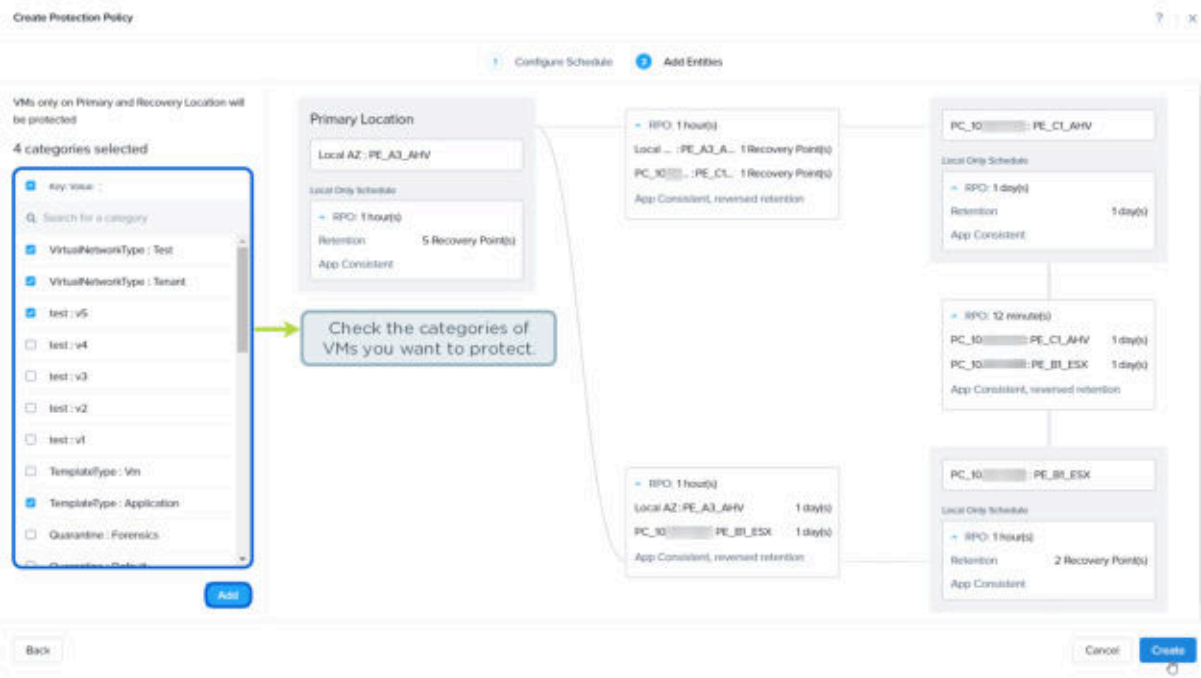


Figure 68: Protection Policy Configuration: Add VM Categories

Prism Central includes built-in VM categories for frequently encountered applications (for example, MS Exchange and Oracle). If the VM category or value you want is not available, first create the category with the required values, or update an existing category so that it has the values you require. Doing so ensures that the VM categories and values are available for selection. You can add VMs to the category either before or after you configure the protection policy. If the guest VMs have a common characteristic, such as belonging to a specific application or location, create a VM category and add the guest VMs into the category.

If you do not want to protect the guest VMs category wise, proceed to the next step without checking VM categories. You can add the guest VMs individually to the protection policy later from the **VMs** page (see Adding Guest VMs Individually to a Protection Policy on page 136).

i. Click **Create**.
The protection policy with a NearSync replication schedule is created. To verify the protection policy, see the **Protection Policies** page. If you check VM categories in step h, the protection policy starts generating recovery points of the guest VMs in those VM categories. To see the generated recovery points, click the hamburger icon at the top-left corner of the window and go to **VM Recovery Points**. Click the recovery points for its

information. You can see the time estimated for the very first replication (seeding) to the recovery sites.

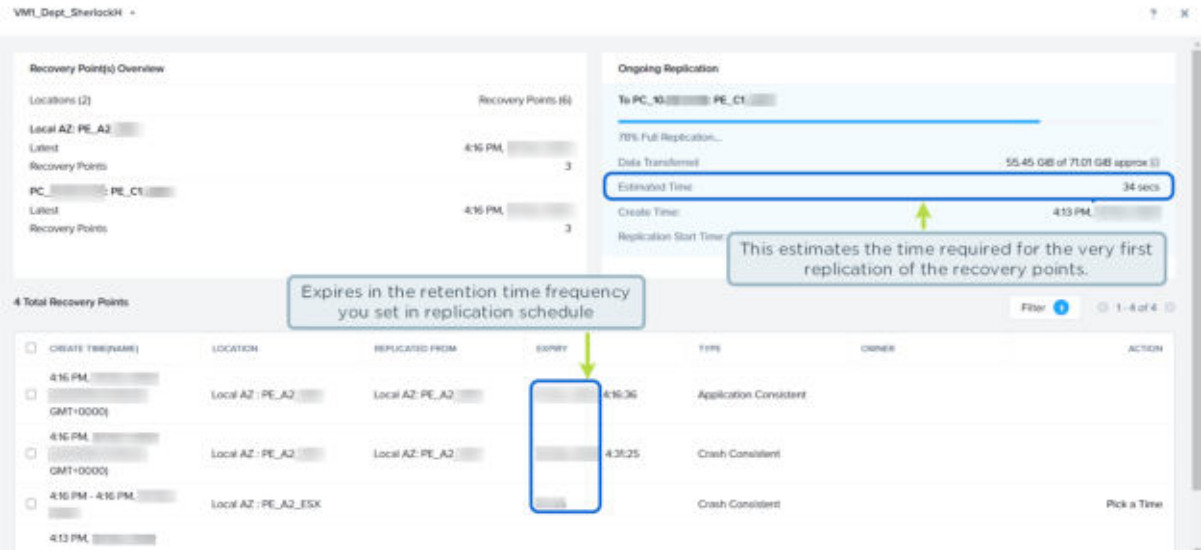


Figure 69: Recovery Points Overview

> Tip: DR using Leap with a NearSync replication schedule also allows you to recover the data of the minute just before the unplanned failover. For example, with a 10 minutely protection policy, you can use the internal lightweight snapshots (LWS) to recover the data of the ninth minute when there is an unplanned failover.

Creating a Recovery Plan (Leap)

To orchestrate the failover of the protected guest VMs to the recovery site, create a recovery plan. After a failover, a recovery plan recovers the protected guest VMs to the recovery site. If you have configured two recovery sites in a protection policy, create two recovery plans for DR —one for recovery to each recovery site. The recovery plan synchronizes continuously to the recovery site in a bidirectional way.

For more information about creating a recovery plan, see Creating a Recovery Plan (Leap) on page 60.

Failover and Failback Operations (Leap)

You can perform test failover, planned failover, and unplanned failover of the guest VMs protected with NearSync replication schedule across different Nutanix clusters at the same or different on-prem availability zone (site). The steps to perform test, planned, and unplanned failover are largely the same irrespective of the replication schedules that protect the guest VMs.

Refer Failover and Failback Management on page 70 for test, planned, and unplanned failover procedures.

Protection with Synchronous Replication Schedule (0 RPO) and DR

Synchronous replication enables you to protect your guest VMs with a zero recovery point objective (0 RPO). A protection policy with Synchronous replication schedule replicates all the writes on the protected guest VMs synchronously to the recovery availability zone (sites) for High Availability. The policy also takes recovery points of those protected VMs every 6 hours—the first snapshot is taken immediately—for raw node (HDD+SSD) size up to 120 TB. Since the replication is synchronous, the recovery points are crash-consistent only. For guest VMs (AHV) protected with Synchronous replication schedule, you can perform DR only to an AHV cluster

at the same or different site. Replicating writes synchronously and also generating recovery points helps to eliminate data losses due to:

- Unplanned failure events (for example, natural disasters and network failure).
- Planned failover events (for example, scheduled maintenance).

Nutanix recommends that the round-trip latency (RTT) between AHV clusters be less than 5 ms for optimal performance of Synchronous replication schedules. Maintain adequate bandwidth to accommodate peak writes and have a redundant physical network between the clusters.

To perform the replications synchronously yet efficiently, the protection policy limits you to configure only one recovery site if you add a Synchronous replication schedule. If you configure Synchronous replication schedule for a guest VM, you cannot add an Asynchronous or NearSync schedule to the same guest VM. Similarly, if you configure an Asynchronous or a NearSync replication schedule, you cannot add a Synchronous schedule to the same guest VM.

If you unpair the sites while the guest VMs in the Nutanix clusters are still in synchronization, the Nutanix cluster becomes unstable. Therefore, disable Synchronous replication and clear stale stretch parameters if any on both the primary and recovery Prism Element before unpairing the sites. For more information about disabling Synchronous replication, see Synchronous Replication Management on page 125.

> Note: Nutanix provides multiple disaster recovery (DR) solutions to secure your environment. See Nutanix Disaster Recovery Solutions on page 11 for the detailed representation of the DR offerings of Nutanix.

Synchronous Replication Requirements

The following are the specific requirements for protecting your AHV guest VMs with Synchronous replication schedule. Ensure that you meet the following requirements in addition to the general requirements of Leap.

For information about the general requirements of Leap, see Leap Requirements on page 20.

For information about node, disk and Foundation configurations required to support Synchronous replication schedules, see On-Prem Hardware Resource Requirements on page 15.

Hypervisor Requirements

AHV

The AHV clusters must be running on version 20190916.189 or newer.

> Note: Synchronous replication schedules support only AHV.

Nutanix Software Requirements

Each on-prem availability zone (site) must have a Leap enabled Prism Central instance.

The primary and recovery Prism Central and Prism Element on the registered Nutanix clusters must be running on the same AOS version.

- AOS 5.17 or newer.
- AOS 5.17.1 or newer to support Synchronous replications of UEFI secure boot enabled guest VMs.
- AOS 5.19.2 or newer for DR to an AHV cluster at the same site. For DR to an AHV cluster at the same site, Prism Central must be running version 2021.3 or newer.

Additional Requirements

- For optimal performance, maintain the round trip latency (RTT) between Nutanix clusters to less than 5 ms. Also, maintain adequate bandwidth to accommodate peak writes and have a redundant physical network between the clusters.
- The storage container name of the protected guest VMs must be the same on both the primary and recovery clusters. Therefore, a storage container must exist on the recovery cluster with the same name as the one on the primary cluster. For example, if the protected guest VMs are in the *SelfServiceContainer* storage container on the primary cluster, there must also be a *SelfServiceContainer* storage container on the recovery cluster.
- For hardware and Foundation configurations required to support Synchronous replication schedules, see On-Prem Hardware Resource Requirements on page 15.
- The clusters on the primary site and the recovery site communicate over the ports 2030, 2036, 2073, and 2090. Ensure that these ports have open access between both the primary and the recovery clusters (Prism Element). For the complete list of required ports, see Port Reference.
- If the primary and the recovery clusters (Prism Element) are in different subnets, open the ports manually for communication.

  > Tip: If the primary and the recovery clusters (Prism Element) are in the same subnet, you need not open the ports manually.

  - To open the ports for communication to the recovery cluster, run the following command on all CVMs of the primary cluster.

    ```
    nutanix@cvm$ allssh 'modify_firewall -f -r remote_cvm_ip,remote_virtual_ip -p
     2030,2036,2073,2090 -i eth0'
    ```

    Replace `remote_cvm_ip` with the IP address of the recovery cluster CVM. If there are multiple CVMs, replace `remote_cvm_ip` with the IP addresses of the CVMs separated by comma.

    Replace `remote_virtual_ip` with the virtual IP address of the recovery cluster.

  - To open the ports for communication to the primary cluster, run the following command on all CVMs of the recovery cluster.

    ```
    nutanix@cvm$ allssh 'modify_firewall -f -r source_cvm_ip,source_virtual_ip -p
     2030,2036,2073,2090 -i eth0'
    ```

    Replace `source_cvm_ip` with the IP address of the primary cluster CVM. If there are multiple CVMs, replace `source_cvm_ip` with the IP addresses of the CVMs separated by comma.

    Replace `source_virtual_ip` with the virtual IP address of the primary cluster.

Synchronous Replication Limitations

Consider the following specific limitations before protecting your guest VMs with Synchronous replication schedule. These limitations are in addition to the general limitations of Leap.

For information about the general limitations of Leap, see Leap Limitations on page 24.

- You cannot restore guest VMs with incompatible GPUs at the recovery cluster.
- You cannot protect guest VMs configured as part of a network function chain.
- You cannot protect guest VMs with affinity policies.

- You cannot resize a guest VM disk while the guest VM is in replication. See *KB-9986* for more information.

Creating a Protection Policy with the Synchronous Replication Schedule (Leap)

To protect the guest VMs in an instant replication schedule, configure a Synchronous replication schedule while creating the protection policy. The policy replicates all the writes on the protected guest VMs synchronously to the recovery availability zone (site) for High Availability. For a raw node (HDD+SSD) size up to 120 TB, the policy also takes crash-consistent recovery points of those guest VMs every 6 hours and replicates them to the recovery site—the first snapshot is taken immediately. To maintain the efficiency of synchronous replication, the protection policy allows you to add only one recovery site for the protected VMs. When creating a protection policy, you can specify only VM categories. If you want to protect guest VMs individually, you must first create the protection policy—which can also include VM categories, and then include the guest VMs individually in the protection policy from the **VMs** page.

Before you begin

See Synchronous Replication Requirements on page 117 and Synchronous Replication Limitations on page 118 before you start.

About this task

To create a protection policy with the Synchronous replication schedule, do the following at the primary site. You can also create a protection policy at the recovery site. Protection policies you create or update at a recovery site synchronize back to the primary site.

Procedure

1. Log on to the Prism Central web console.

2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Protection Policies** in the left pane.

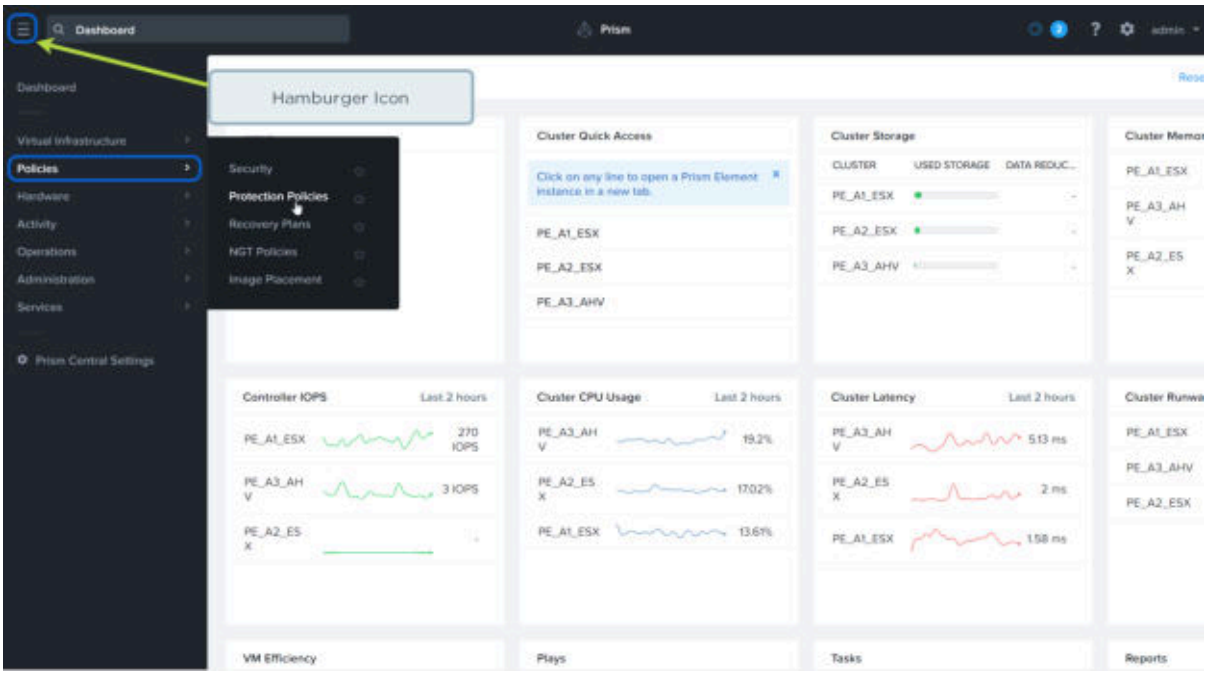


Figure 70: Protection Policy Configuration: Protection Policies

3. Click **Create Protection Policy**.
   Specify the following information in the Create Protection Policy window.

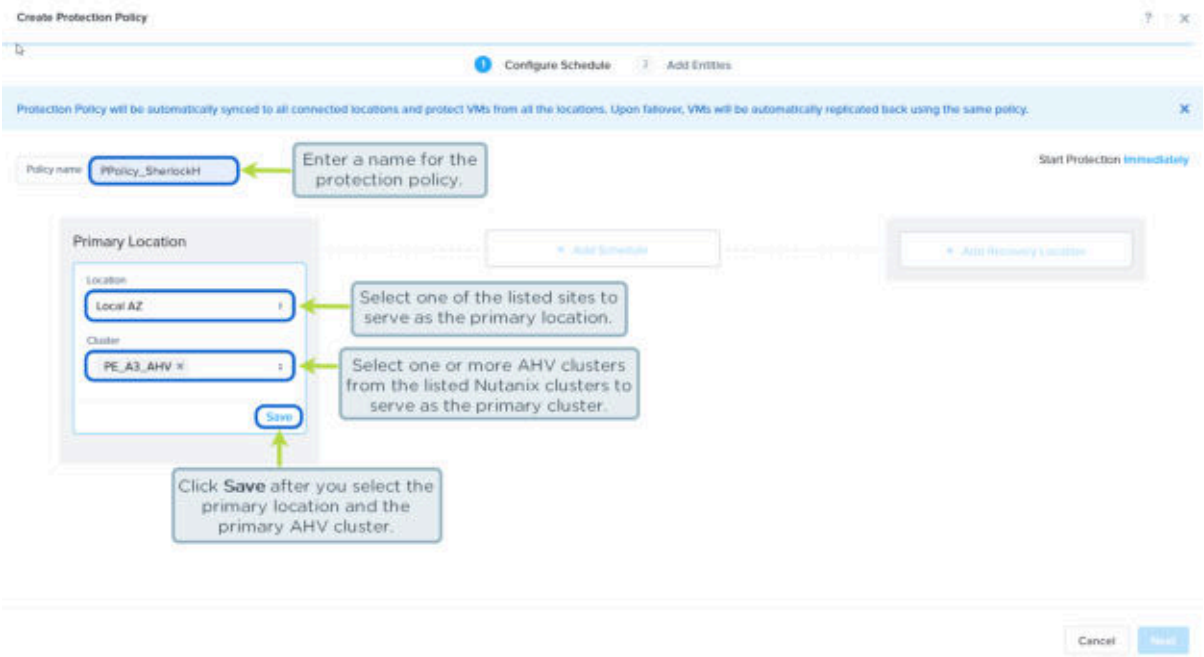


Figure 71: Protection Policy Configuration: Select Primary Location

a. **Policy name**: Enter a name for the protection policy.

> Caution: The name can be of only alphanumeric, dot, dash, and underscore characters.

b. In the **Primary Location** pane, specify the following information.

- 1. **Location**: From the drop-down list, select an availability zone (site) that hosts the guest VMs to protect.

  The drop-down lists all the sites paired with the local site. Local AZ represents the local site (Prism Central). For your primary site, you can check either the local site or a non-local site.

  2. **Cluster**: From the drop-down list, select the AHV cluster that hosts the VMs to protect.

  The drop-down lists all the Nutanix clusters registered to Prism Central representing the selected site. If you want to protect the guest VMs from multiple AHV clusters in the same protection policy, select the AHV clusters that host those guest VMs. All Clusters protects the guest VMs of all Nutanix clusters registered to Prism Central. Select All Clusters only if all the clusters are running AHV.

  3. Click **Save**.

  Clicking **Save** activates the **Recovery Location** pane. Do not add a local schedule to retain the recovery points locally. To maintain the replication efficiency,

Synchronous replication allows only the replication schedule. If you add a local schedule, you cannot click **Synchronous** in step d.

c. In the **Recovery Location** pane, specify the following information.

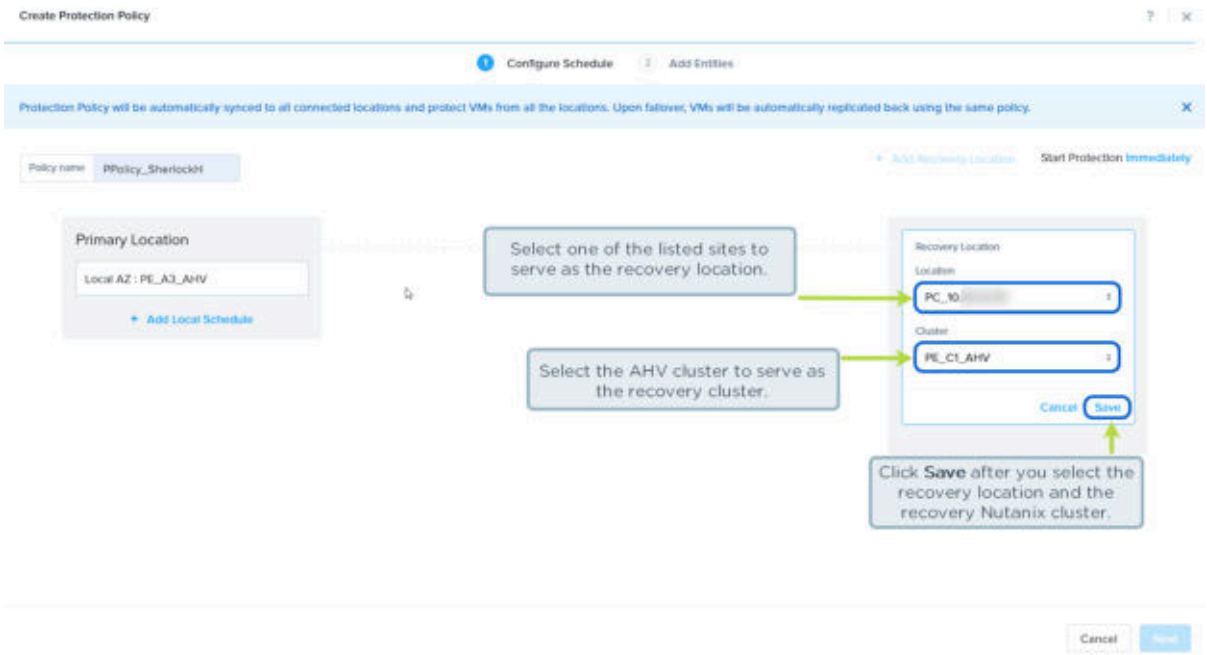


Figure 72: Protection Policy Configuration: Select Recovery Location

- 1. **Location**: From the drop-down list, select the availability zone (site) where you want to replicate the recovery points.

  The drop-down lists all the sites paired with the local site. Local AZ represents the local site (Prism Central). Select Local AZ if you want to configure DR to a different AHV cluster at the same site.

  If you do not select a site, local recovery points that are created by the protection policy do not replicate automatically. You can, however, replicate the recovery points manually and use recovery plans to recover the guest VMs. For more information, see Protection and Manual DR (Leap) on page 143.

  2. **Cluster**: From the drop-down list, select the AHV cluster where you want to replicate the guest VM writes synchronously and recovery points.

     The drop-down lists all the Nutanix clusters registered to Prism Central representing the selected site. You can select one AHV cluster at the recovery site. Do not select an ESXi cluster because DR configurations using Leap support only AHV cluster. If you select an ESXi cluster and configure a Synchronous replication schedule, replications fail.

     > Note: Selecting auto-select from the drop-down menu replicates the recovery points to any available cluster at the recovery site. Select auto-select from the drop-down list only if all the clusters at the recovery site are running on AHV and are up and running.

  3. Click **Save**.

     Clicking **Save** activates the **+ Add Schedule** button between the primary and the recovery site. Do not add a local schedule to retain the recovery points locally.

To maintain the replication efficiency, Synchronous replication allows only the replication schedule. If you add a local schedule, you cannot click **Synchronous** in step d.

d. Click **+ Add Schedule** to add a replication schedule between the primary and the recovery site.

   Specify the following information in the Add Schedule window. The window auto-populates the Primary Location and Recovery Location that you have selected in step b and step c.

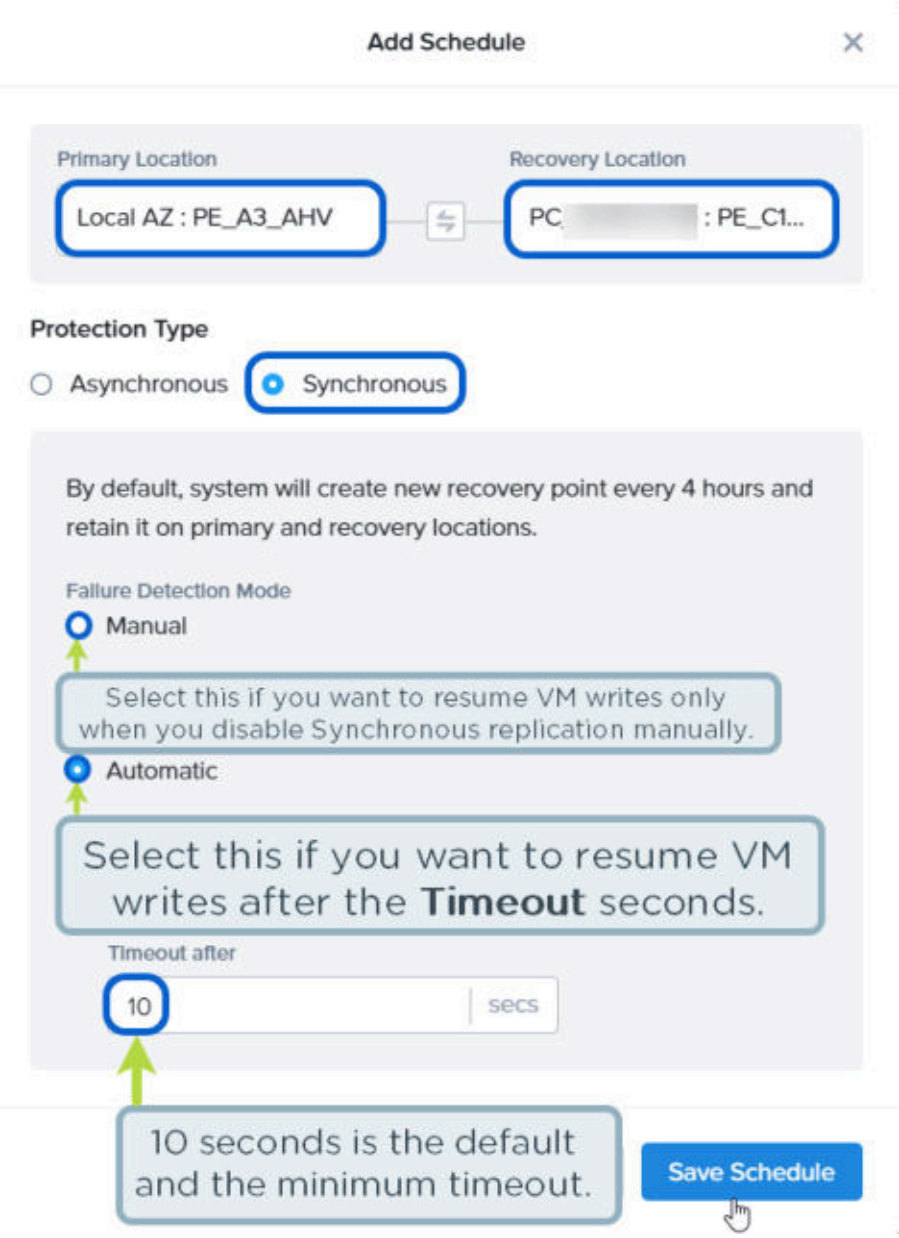


Figure 73: Protection Policy Configuration: Add Schedule (Synchronous)

- 1. **Protection Type**: Click **Synchronous**.

2. **Failure Handling**: Select one of the following options to handle failure. For example, if the connection between the primary and the recovery site breaks and VM writes on the primary cluster stops replicating.

   - **Manual**: Select this option if you want to resume the VM writes on the primary site only when you manually disable Synchronous replication.
   - **Automatic**: Select this option to resume VM writes on the primary site automatically after the specified **Timeout after** seconds.

     Note: The minimum timeout is 10 seconds.

3. Click **Save Schedule**.

   Clicking **Save Schedule** disables the **+ Add Recovery Location** button at the top-right because to maintain the efficiency of synchronous replication, the policy allows you to add only one recovery site.

By default, recovery point creation begins immediately after you create the protection policy. If you want to specify when recovery point creation must begin, click **Immediately** at the top-right corner, and then, in the **Start Time** dialog box, do the following.

- 1. Click **Start protection at specific point in time**.
  2. Specify the time at which you want to start taking recovery points.
  3. Click **Save**.

e. Click **Next**.

   Clicking **Next** shows a list of VM categories where you can optionally check one or more VM categories to protect in the protection policy. DR configurations using Leap allows you to protect a guest VM by using only one protection policy. Therefore, VM categories specified in another protection policy are not in the list. If you protect a guest VM in another protection policy by specifying the VM category of the guest VM (category-based inclusion), and if you protect the guest VM from the **VMs** page in this policy (individual inclusion), the individual inclusion supersedes the category-based inclusion. Effectively, only the protection policy that protected the individual guest VM protects the guest VM.

   For example, the guest VM *VM_SherlockH* is in the category *Department:Admin*, and you add this category to the protection policy named *PP_AdminVMs*. Now, if you add *VM_SherlockH* from the **VMs** page to another protection policy named *PP_VMs_UK*, *VM_SherlockH* is protected in *PP_VMs_UK* and unprotected from *PP_AdminVMs*.

f. If you want to protect the guest VMs category wise, check the VM categories that you want to protect from the list and click **Add**.

   Prism Central includes built-in VM categories for frequently encountered applications (for example, MS Exchange and Oracle). If the VM category or value you want is not available, first create the category with the required values, or update an existing category so that it has the values you require. Doing so ensures that the VM categories and values are available for selection. You can add VMs to the category either before or after you configure the protection policy. If the guest VMs have a common characteristic, such as belonging to a specific application or location, create a VM category and add the guest VMs into the category.

   If you do not want to protect the guest VMs category wise, proceed to the next step without checking VM categories. You can add the guest VMs individually to the protection

policy later from the **VMs** page (see Adding Guest VMs Individually to a Protection Policy on page 136).

g. Click **Create**.
The protection policy with Synchronous replication schedule is created. To verify the protection policy, see the **Protection Policies** page. If you check VM categories in step f, the protection policy starts generating recovery points of the guest VMs in those VM categories. To see the generated recovery points, click the hamburger icon at the top-left corner of the window and go to **VM Recovery Points**. Click the recovery points for its information. You can see the time estimated for the very first replication (seeding) to the recovery sites.

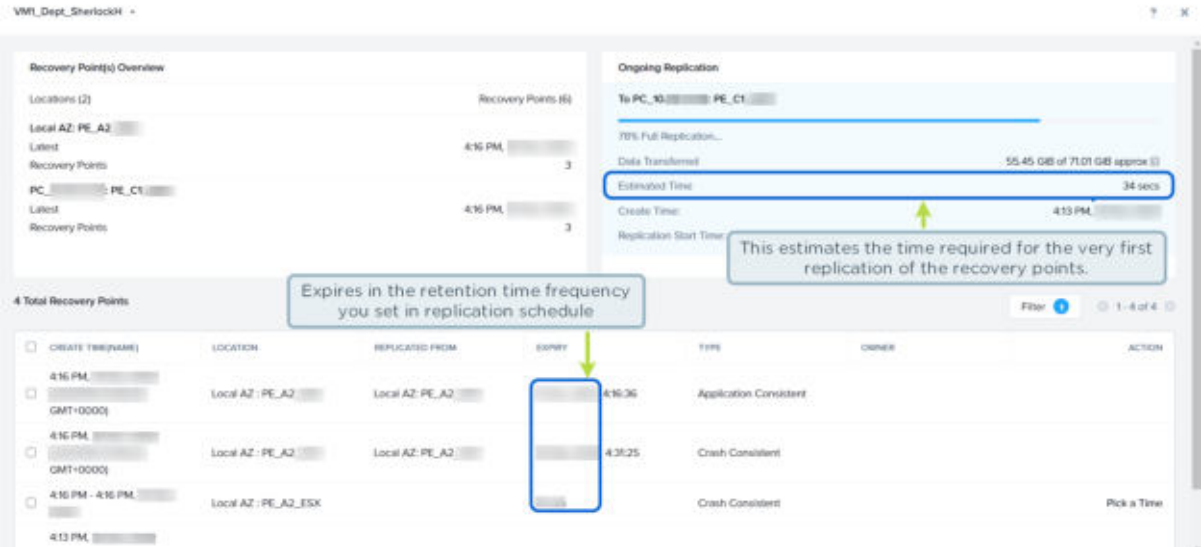


Figure 74: Recovery Points Overview

Creating a Recovery Plan (Leap)

To orchestrate the failover of the protected guest VMs to the recovery site, create a recovery plan. After a failover, a recovery plan recovers the protected guest VMs to the recovery site. If you have configured two recovery sites in a protection policy, create two recovery plans for DR —one for recovery to each recovery site. The recovery plan synchronizes continuously to the recovery site in a bidirectional way.

For more information about creating a recovery plan, see Creating a Recovery Plan (Leap) on page 60.

Synchronous Replication Management

Synchronous replication instantly replicates all writes on the protected guest VMs to the recovery cluster. Replication starts when you configure a protection policy and add the guest VMs to protect. You can manage the replication by enabling, disabling, pausing, or resuming the Synchronous replication on the protected guest VMs from the Prism Central.

Enabling Synchronous Replication

When you configure a protection policy with Synchronous replication schedule and add guest VMs to protect, the replication is enabled by default. However, if you have disabled the Synchronous replication on a guest VM, you have to enable it to start replication.

About this task

To enable Synchronous replication on a guest VM, perform the following procedure at the primary availability zone (site). You can also perform the following procedure at the recovery site. The operations you perform at a recovery site synchronize back to the primary site.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Virtual Infrastructure** > **VMs** > **List** in the left pane.
3. Select the guest VMs on which you want to enable Synchronous replication.
4. Click **Protect** from the **Actions** drop-down menu.
5. Select the protection policy in the table to include the guest VMs in the protection policy.
6. Click **Protect**.

Pausing Synchronous Replication

The protected guest VMs on the primary cluster stop responding when the recovery cluster is disconnected abruptly (for example, due to network outage or internal service crash). To come out of the unresponsive state, you can pause Synchronous replication on the guest VMs. Pausing Synchronous replication temporarily suspends the replication state of the guest VMs without completely disabling the replication relationship.

About this task

To pause Synchronous replication on a guest VM, perform the following procedure.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Virtual Infrastructure** > **VMs** > **List** in the left pane.
3. Select the guest VMs on which want to pause the Synchronous replication.
4. Click **Pause Synchronous Replication** from the **Actions** drop-down menu.

Resuming Synchronous Replication

You can resume the Synchronous replication that you had paused to come out of the unresponsive state of the primary cluster. Resuming Synchronous replication restores the replication status and reconciles the state of the guest VMs. To resume Synchronous replication on a guest VM, perform the following procedure.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Virtual Infrastructure** > **VMs** > **List** in the left pane.
3. Select the guest VMs on which want to resume Synchronous replication.

4. Click **Resume Synchronous Replication** from the **Actions** drop-down menu.

Failover and Failback Operations (Leap)

You can perform test failover, planned failover, and unplanned failover of the guest VMs protected with Synchronous replication schedule across the AHV clusters at the different on-prem availability zone (site). The steps to perform test, planned, and unplanned failover are largely the same irrespective of the replication schedules that protects the guest VMs. Additionally, a planned failover of the guest VMs protected with Synchronous replication schedule also allows for live migration of the protected guest VMs.

Refer Failover and Failback Management on page 70 for test, planned, and unplanned failover procedures.

Cross-Cluster Live Migration

Planned failover of the guest VMs protected with Synchronous replication schedule supports live migration to another AHV cluster. Live migration offers zero downtime for your applications during a planned failover event to the recovery cluster (for example, during scheduled maintenance).

Cross-Cluster Live Migration Requirements

The following are the specific requirements to successfully migrate your guest VMs with Live Migration.

Ensure that you meet the following requirements in addition to the requirements of Synchronous replication schedule (Synchronous Replication Requirements on page 117) and general requirements of Leap (Leap Requirements on page 20).

- Stretch L2 networks across the primary and recovery sites.

  Network stretch spans your network across different sites. A stretched L2 network retains the IP addresses of guest VMs after their Live Migration to the recovery site.

- Both the primary and recovery Nutanix clusters must have identical CPU types.

  The primary and recovery Nutanix clusters must have identical CPU feature set. If the CPU feature sets (set of CPU flags) are unidentical, Live Migration fails.

- Both the primary and recovery Nutanix clusters must run on the same AHV version.
- If the primary and the recovery Nutanix clusters (Prism Element) are in different subnets, open the ports 49250–49260 for communication. For the complete list of required ports, see Port Reference.

Cross-Cluster Live Migration Limitations

Consider the following limitation in addition to the limitations of Synchronous replication schedule (Synchronous Replication Limitations on page 118) and general limitations of Leap (Leap Limitations on page 24) before performing live migration of your guest VMs.

- Live migration of guest VMs fails if the guest VMs are part of Flow security policies.

> Tip: To enable the guest VMs to retain the Flow security policies after the failover (live migration), revoke the policies on the guest VMs and **Export** them to the recovery site. At the recovery site, **Import** the policies. The guest VMs read the policies automatically after recovery.

Performing Cross-Cluster Live Migration

If due to a planned event (for example, scheduled maintenance of guest VMs) at the primary availability zone (site), you want to migrate your applications to another AHV cluster without downtime, perform a planned failover with Live Migration to the recovery site.

About this task

To live migrate the guest VMs, do the following procedure at the recovery site.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select a recovery plan for the failover operation.

> Caution: The **Recovery Plans** page displays many recovery plans. Select the recovery plan that has **Stretch Networks**. If you select a recovery plan having **Non-stretch networks**, the migration fails. For more information about selection of stretch and non-stretch networks, see Creating a Recovery Plan (Leap) on page 60.

4. Click **Failover** from the **Actions** drop-down menu.

   Specify the following information in the **Failover from Recovery Plan** window. The window auto-populates the **Failover From** and **Failover To** locations from the recovery plan you select in step 3.

   > Note: If you select more than one recovery plan in step 3, the **Failover** action is available only when the selected recovery plans have the same primary and recovery locations.

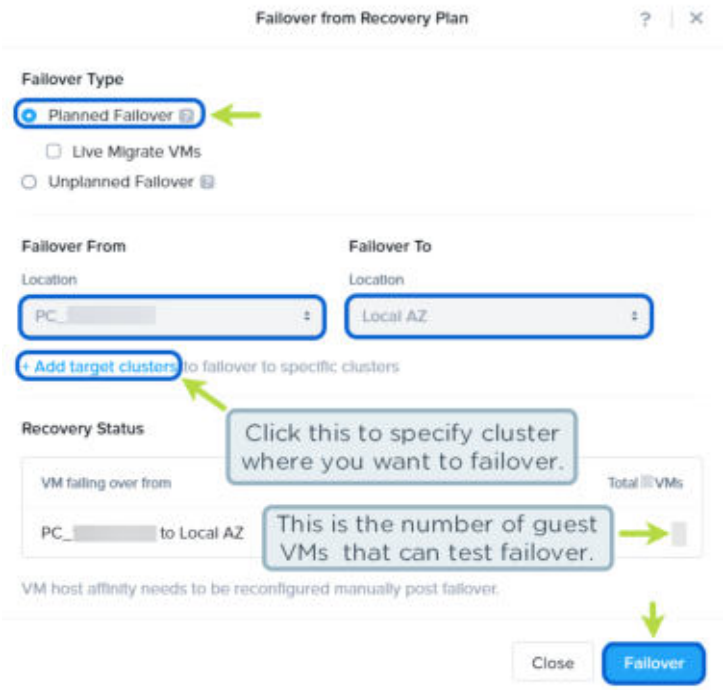


   Figure 75: Planned Failover

   a. **Failover Type**: Click **Planned Failover** and check **Live Migrate VMs**.

   b. Click **+ Add target clusters** if you want to failover to specific clusters at the recovery site.

      If you do not add target clusters, the recovery plan migrates the guest VMs to any AHV cluster at the recovery site.

5. Click **Failover**.

   The **Failover from Recovery Plan** dialog box lists the errors and warnings, if any, and allows you to stop or continue the test operation. If there are no errors or you resolve the errors in step 6, the guest VMs migrate and start at the recovery cluster. The migration might show a network latency of 300-600 ms. You cannot see the migrated guest VMs on the primary cluster because those VMs come up at the recovery cluster after the migration.

6. If you see errors, do the following.

   - To review errors or warnings, click **View Details** in the description.

     Resolve the error conditions and then restart the failover procedure.
   - Select one of the following.
     - To stop the failover operation, click **Abort**.
     - To continue the failover operation despite the warnings, click **Execute Anyway**.

       Note: You cannot continue the failover operation when the validation fails with errors.

AHV Metro (Witness Option)

Witness option extends the capability of Metro Availability to AHV clusters in Leap. Witness is a service within Prism Central (including scaleout deployment) that monitors your disaster recovery (DR) configuration. The service enables you to automate pausing and recovering your guest VMs when the communication between the primary and recovery AHV clusters is unavailable. To automate the recovery process, the recovery plan configuration is enhanced to automatically handle the failure execution.

Witness continuously reads the health status from the primary and recovery clusters. If the communication between the two clusters is unavailable for a set time period, the service breaks the replication between the clusters. If the primary cluster is unavailable, the service can also trigger an unplanned failover automatically to start the guest VMs at the recovery cluster. The following table details the cluster failure scenarios, and the witness response behavior.

Table 23: Witness Response Behaviors

Failure scenario	Witness mode (automatic)	Manual mode
Primary cluster outage or complete network failure in primary cluster	• Replication breaks after timeout • Guest VMs fail over automatically to recovery cluster	Administrative intervention required to: • Break the replication • Fail over the guest VMs to recovery cluster
Recovery cluster outage or complete network failure in recovery cluster	• Replication breaks after timeout • Guest VMs continue to run on primary cluster	Administrative intervention required to break the replication.
Connection loss between primary and recovery cluster	• Replication breaks after timeout • Guest VMs continue to run on primary cluster	Administrative intervention required to break the replication.
Witness failure	No impact. However, ability to recover automatically is lost.	n/a

Failure scenario	Witness mode (automatic)	Manual mode
Connection loss between witness and both primary and recovery clusters	Guest VMs continue to run on primary cluster.	n/a
Failure at both primary and recovery clusters	Witness option stops. Administrative intervention required.	Witness option stops. Administrative intervention required.
Storage-only outage on primary cluster	The witness brings the storage up on the recovery cluster. The storage on the primary cluster goes into an inaccessible state.	All replications break until, either the storage outage is fixed, or the storage is manually brought up on recovery cluster using Prism Element web console. After bringing storage up on recovery cluster, guest VMs have to be migrated manually.
Storage-only outage on recovery cluster	Primary cluster remains active and automatically breaks replication to recovery cluster to be able to continue serving I/O requests locally. The storage on recovery cluster goes into a read-only state and any guest VMs that were running on hosts on recovery cluster goes into a paused state.	All replications break until the storage outage is fixed.

You can configure witness only for protection and DR to the AHV cluster at the same availability zone (site). Nutanix recommends deploying the single Prism Central at a different AHV or ESXi cluster (different fault domain) to help avoid a single point of failure. Prism Central residing in a separate fault domain provides an outside view that can distinguish a cluster failure from a network interruption between the Metro Availability clusters.

If your cluster is running ESXi and you want the capability of witness, see the *Witness Option* topic in *Data Protection and Recovery with Prism Element* guide.

AHV Metro Requirements

These requirements help you to successfully configure a witness to automate the recovery (failover) when the primary AHV cluster is unavailable. Ensure that you meet these requirements in addition to the requirements of Synchronous replication and the applicable general requirements of Leap before you start.

See Synchronous Replication Requirements on page 117 and Leap Requirements on page 20.

Hypervisor Requirements

AHV

Nutanix Software Requirements

- Prism Central must run version pc.2021.7 or newer.

  For Prism Central and AOS version compatibility, see Software Interoperability

- The primary and recovery AHV clusters must be registered to the same Prism Central deployment.
- The primary and recovery AHV clusters registered to Prism Central must be running AOS version 6.0 or newer.

Recovery Plan Requirements

- A guest VM must be added in a single recovery plan with witness configuration.

  The system generates an alert when a guest VM is in multiple recovery plans with witness configuration.

- Only the guest VMs protected with Synchronous replication must be in the recovery plans with witness configuration.

  The system generates an alert when guest VMs protected with Asynchronous and NearSync replication are in the recovery plans with witness configuration.

Additional Requirements

- When the recovery AHV cluster becomes available after the break in the synchronous replication between the AHV clusters, manually resume the synchronous replication of the guest VMs from the primary cluster. For more information about resuming synchronous replication, see Resuming Synchronous Replication on page 126.
- When the primary AHV cluster becomes available after the unplanned failover of guest VMs, delete the guest VMs on the primary cluster if the unplanned failover was successful.

AHV Metro Limitations

Consider the following limitations before configuring AHV Metro to automate the failover recovery when the primary AHV cluster is unavailable. These limitations are in addition to limitations of Synchronous replication and applicable general limitations of Leap.

See Synchronous Replication Limitations on page 118 and Leap Limitations on page 24.

- You cannot use witness in a configuration with two or more Prism Central deployments across availability zones.
- You cannot use witness to automate the unplanned failover recovery of the guest VMs protected with Asynchronous or NearSync replication.
- You cannot use witness to automate the unplanned failover recovery to Xi Cloud Services.
- You can only create the recovery plan with witness configuration after creating a protection policy, and having the entities in Synced status.

AHV Metro Recommendations

Nutanix recommends the following best practices for configuring a witness to automate an unplanned failover when the primary AHV cluster fails.

See Leap Recommendations on page 25.

- If the single Prism Central that you use for protection and disaster recovery to AHV clusters at the same availability zone (site) becomes unavailable, the witness service also becomes unavailable. To help avoid a single point of failure, Nutanix recommends deploying the single Prism Central in a different availability zone (different fault domain).
- To avoid conflicts after the unplanned failover of the guest VMs, shut down the guest VMs associated with this recovery plan if the primary site becomes active. Manually power off the guest VMs on either primary or recovery site after the failover is complete.

Configuring AHV Metro

Before you begin

See AHV Metro Requirements on page 131 and AHV Metro Limitations on page 132.

About this task

Procedure

1. Log on to the Prism Central web console.
2. Click the settings button (gear icon) at the top-right corner of the window.
3. Click **Witness** in the **General** section on the left pane.

   Witness is enabled by default in Prism Central. The cluster information is visible if the cluster hosting Prism Central is also registered to Prism Central. Click **View Usage History** to see unplanned failover events triggered by the witness in the past.
4. Create a recovery plan to automatically trigger an unplanned failover to recover the AHV cluster.

   For more information about creating a recovery plan, see Creating a Recovery Plan (Leap) on page 60. The procedure to create a recovery plan with witness configuration is identical except as follows.

   a. In the General tab, select the **Primary Location** and the **Recovery Location** as Local AZ.

   b. Select Automatic in **Failure Execution Mode** to enable the witness to break the replication and trigger an unplanned failover to the recovery AHV cluster in the specified timeout seconds.

   > Note: Set **Execute failover after disconnectivity of** to value between 30 to 300 seconds. If a guest VM protected in a Synchronous replication schedule with auto-break timeout is in a witness configured recovery plan, the auto-break timeout specified in the recovery plan supersedes.

   > Note: **Failure Execution Mode** appears only when the **Primary Location** and the **Recovery Location** are the same.

   A recovery plan with witness configuration is created. To confirm the witness configuration, go to settings button (gear icon) at the top-right corner of the window and click **Witness** in the **General** section on the left pane. The page shows the number of recovery plans using the witness under **Monitoring (On this PC)**. Click the number to see the list of recovery plans using the witness.

Converting a Multisite Deployment to Single-Site

To use disaster recovery features that support only single Prism Central (site) managed deployments, you can convert your multisite deployment to single-site deployment. For

example, in two site deployment where each Prism Central *(Prism Central A, Prism Central B)* hosts one Prism Element *(Prism Element A, Prism Element B)*, you can perform the following procedure to convert into single-site deployment *(Prism Central A managing both Prism Element A, Prism Element B)*.

Before you begin

This procedure converts deployments protected with Synchronous replication schedules. The result of this procedure is single Prism Central managed deployment. Therefore, see the Synchronous Replication Requirements on page 117 for the supported Prism Central and AOS versions.

> Note: The conversion procedure for deployments protected in Asynchronous and NearSync replications schedules are identical except that the protection status (step 2) of Asynchronous and NearSync replications schedules is available only in **Focus** > **Data Protection** at the top-right corner of Prism Central web console.

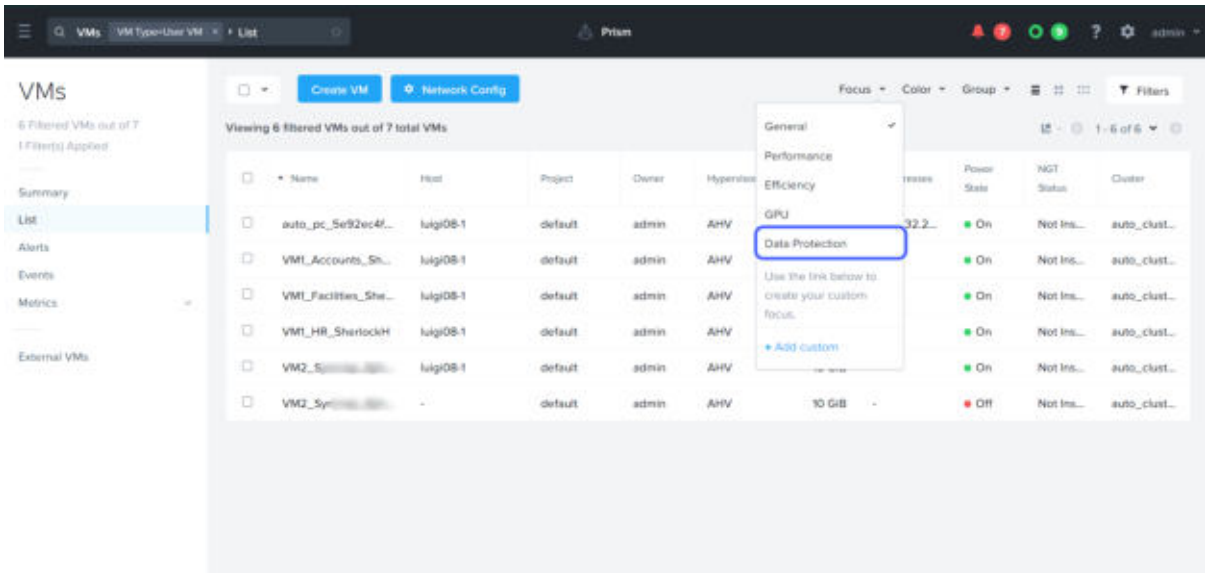


Figure 76: Focus

About this task

Procedure

1. Log on to the web console of *Prism Central A*.

2. Modify all the protection policies and recovery plans that have reference to *Prism Element B* and *Prism Central B*.

   a. Modify the protection policies to either remove all the references to *Prism Element B* and *Prism Central B* or remove all the guest VMs from the policy.

      For more information about updating a protection policy, see Updating a Protection Policy on page 139.

   b. Modify the recovery plans to remove all the references to *Prism Element B* and *Prism Central B*.

      If you do not modify the recovery plans, the recovery plans become invalid after the unregistration of *Prism Element B* with *Prism Central A* in step 2. For more information about updating a recovery plan, see Updating a Recovery Plan on page 143.

   c. Ensure that there are no issues with the modified protection policies and recovery plans.

      Before unregistering *Prism Element B* from *Prism Central B* in step 2, ensure that no guest VM is protected to and from *Prism Element B*.

3. Unprotect all the guest VMs replicating to and from *Prism Element B* and *Prism Central B*.

   To see the unprotect status of the guest VMs, click **Data Protection** from the **Focus** menu at the top-right corner.

4. Ensure that the guest VMs unprotect completely.

   - To ensure all the stretch states are deleted, log on to *Prism Element B* through SSH and run the following command on.

     ```
     nutanix@cvm$ stretch_params_printer
     ```

   - To ensure all the stretch states between *Prism Central B* and *Prism Element B* are deleted, log on to *Prism Central B* through SSH and run the following command.

     ```
     mcli> mcli dr_coordinator.list
     ```

5. Unregister *Prism Element B* from *Prism Central B*.

   After unregistering *Prism Element B* from *Prism Central B*, all *Prism Central B* attributes and policies applied to guest VMs on *Prism Element B* (for example, VM categories) gets removed.

6. Register *Prism Element B* to *Prism Central A*.

   After registering *Prism Element B* to *Prism Central A*, add all *Prism Central B* attributes and policies applied to entities on the *Prism Element B* (for example, VM categories) back.

7. Modify the protection policies and recovery plans to refer to *Prism Central A* and *Prism Element B*.

8. Unpair *Prism Central B*.

   To ensure all the stretch states between *Prism Central A* and *Prism Central B* are deleted, log on to both *Prism Central A* and *Prism Central B* through SSH and run the following command..

   ```
   mcli dr_coordinator.list
   ```

Protection Policy Management

A protection policy automates the creation and replication of recovery points. When creating a protection policy, you specify replication schedules, retention policies for the recovery points,

and the guest VMs you want to protect. You also specify a recovery availability zone (maximum 2) if you want to automate recovery point replication to the recovery availability zones (sites).

When you create, update, or delete a protection policy, it synchronizes to the recovery sites and works bidirectionally. The recovery points generated at the recovery sites replicate back to the primary site when the primary site starts functioning. For information about how Leap determines the list of sites for synchronization, see Entity Synchronization Between Paired Availability Zones on page 151.

Adding Guest VMs Individually to a Protection Policy

You can also protect guest VMs individually in a protection policy from the **VMs** page, without the use of a VM category. To protect guest VMs individually in a protection policy, perform the following procedure.

About this task

> Note: If you protect a guest VM individually, you can remove the guest VM from the protection policy only by using the procedure in Removing Guest VMs Individually from a Protection Policy on page 138.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Virtual Infrastructure** > **VMs** in the left pane.

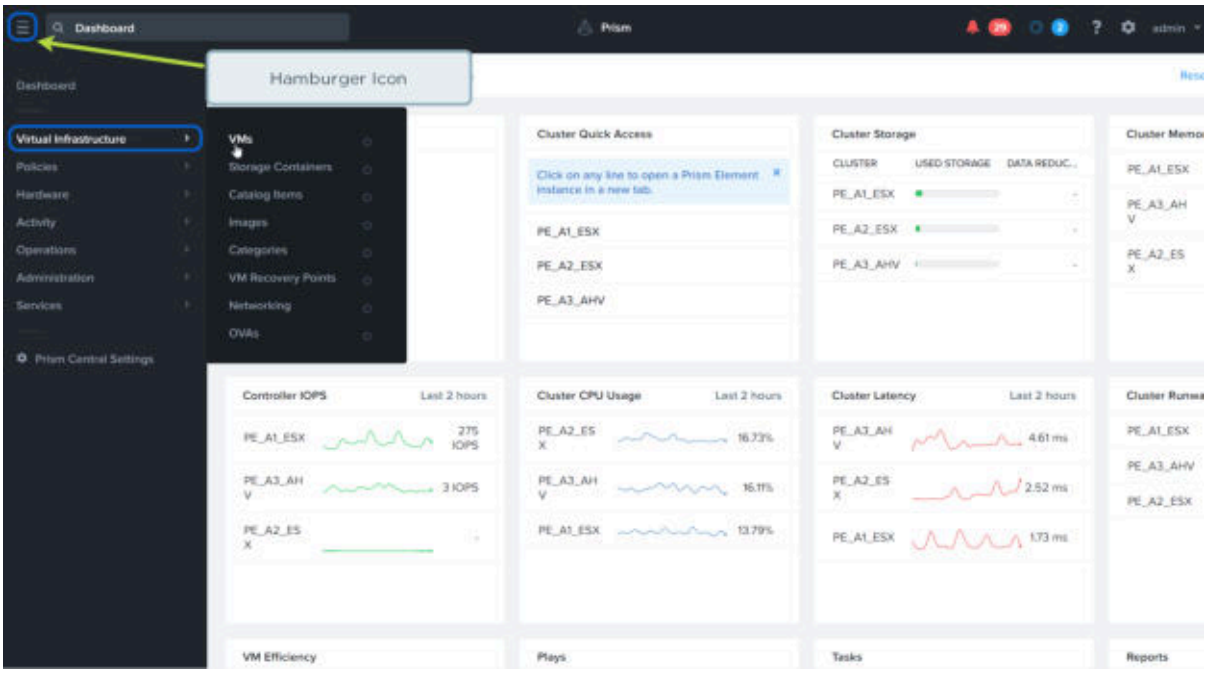


Figure 77: Virtual Infrastructure: VMs

3. Select the guest VMs that you want to add to a protection policy.

4. Click **Protect** from the **Actions** drop-down menu.

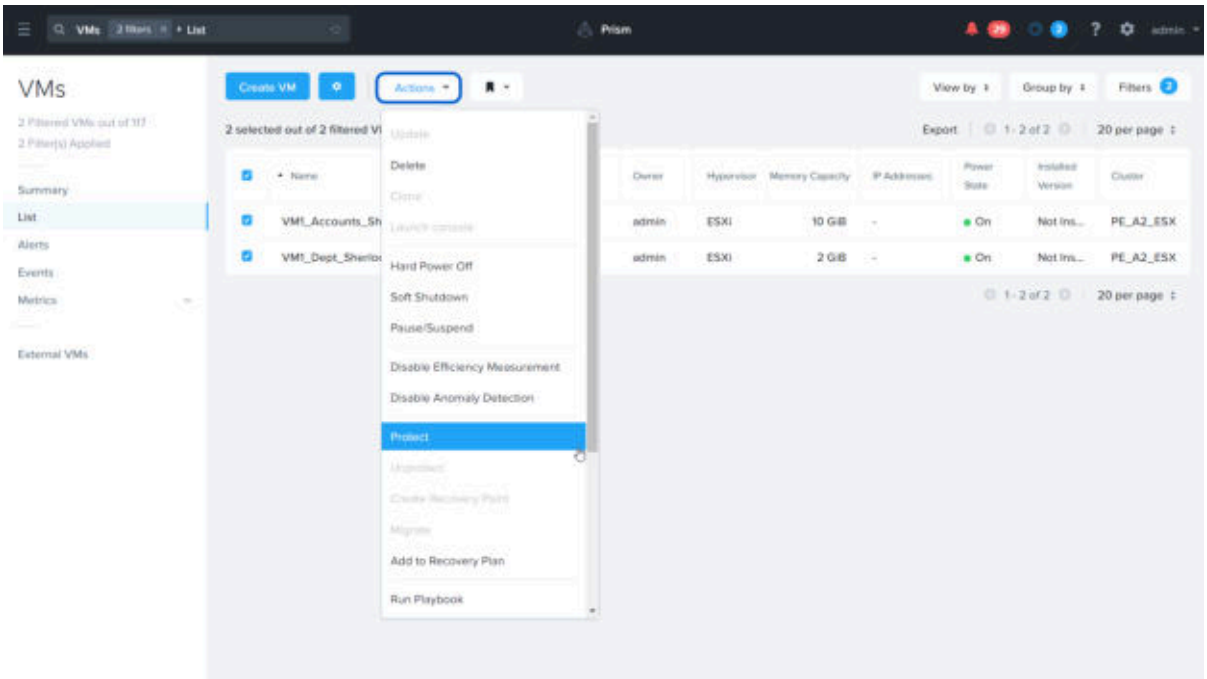


Figure 78: Protect Guest VMs Individually: Actions

5. Select the protection policy in the table to protect the selected guest VMs.

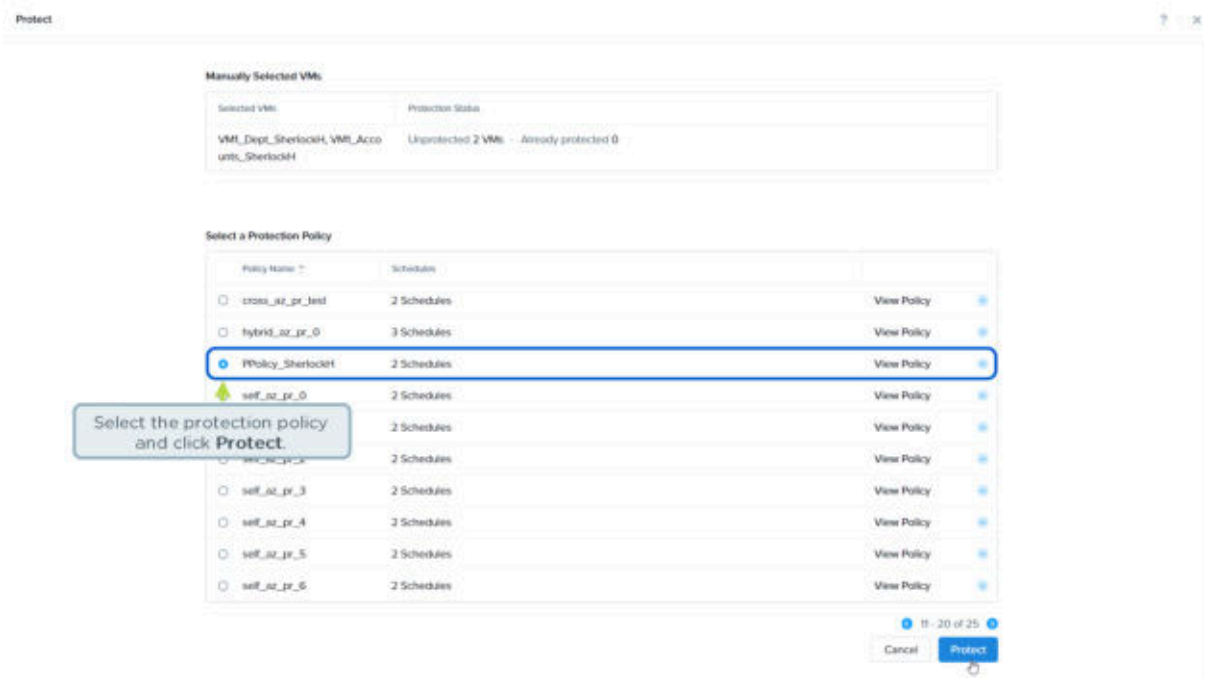


Figure 79: Protect Guest VMs Individually: Protection Policy Selection

6. Click **Protect**.

Removing Guest VMs Individually from a Protection Policy

You can remove guest VMs individually from a protection policy from the **VMs** page. To remove guest VMs individually from a protection policy, perform the following procedure.

About this task

> Note: If a guest VM is protected under a VM category, you cannot remove the guest VM from the protection policy by the following procedure. You can remove the guest VM from the protection policy only by dissociating the guest VM from the VM category.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Virtual Infrastructure** > **VMs** in the left pane.

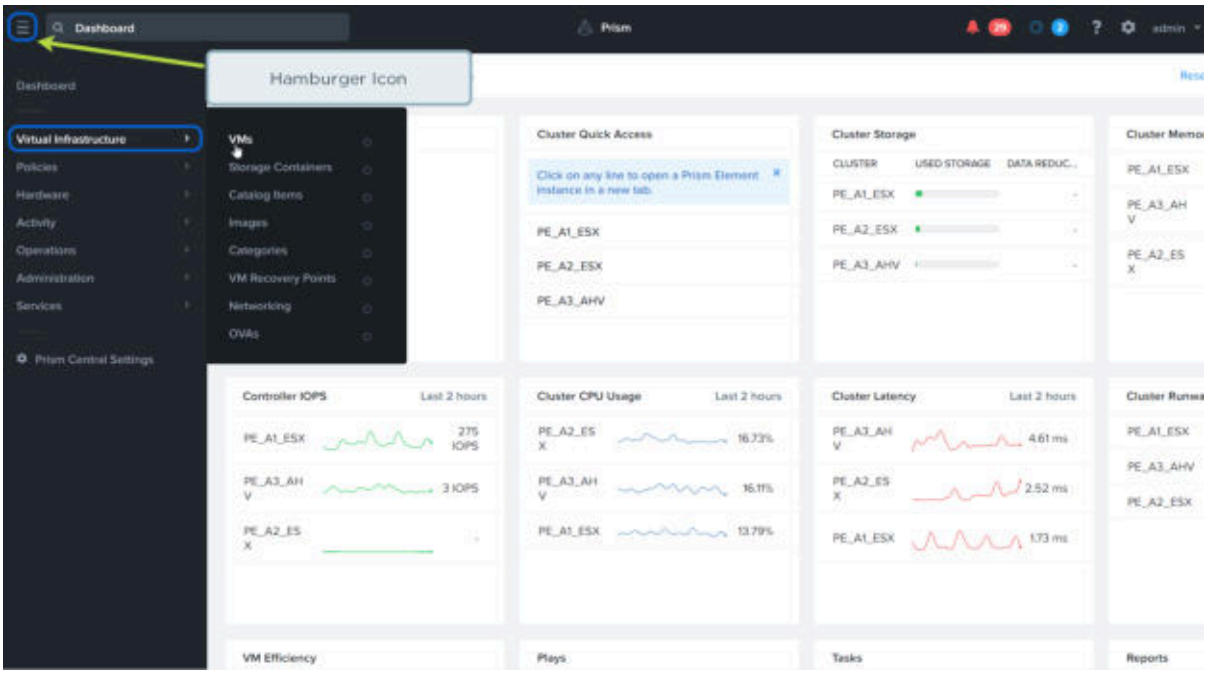


Figure 80: Virtual Infrastructure: VMs

3. Select the guest VMs that you want to remove from a protection policy.
4. Click **UnProtect** from the **Actions** drop-down menu.

Cloning a Protection Policy

If the requirements of the protection policy that you want to create are similar to an existing protection policy, you can clone the existing protection policy and update the clone. To clone a protection policy, perform the following procedure.

About this task

Procedure

1. Log on to the Prism Central web console.

2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Protection Policies** in the left pane.

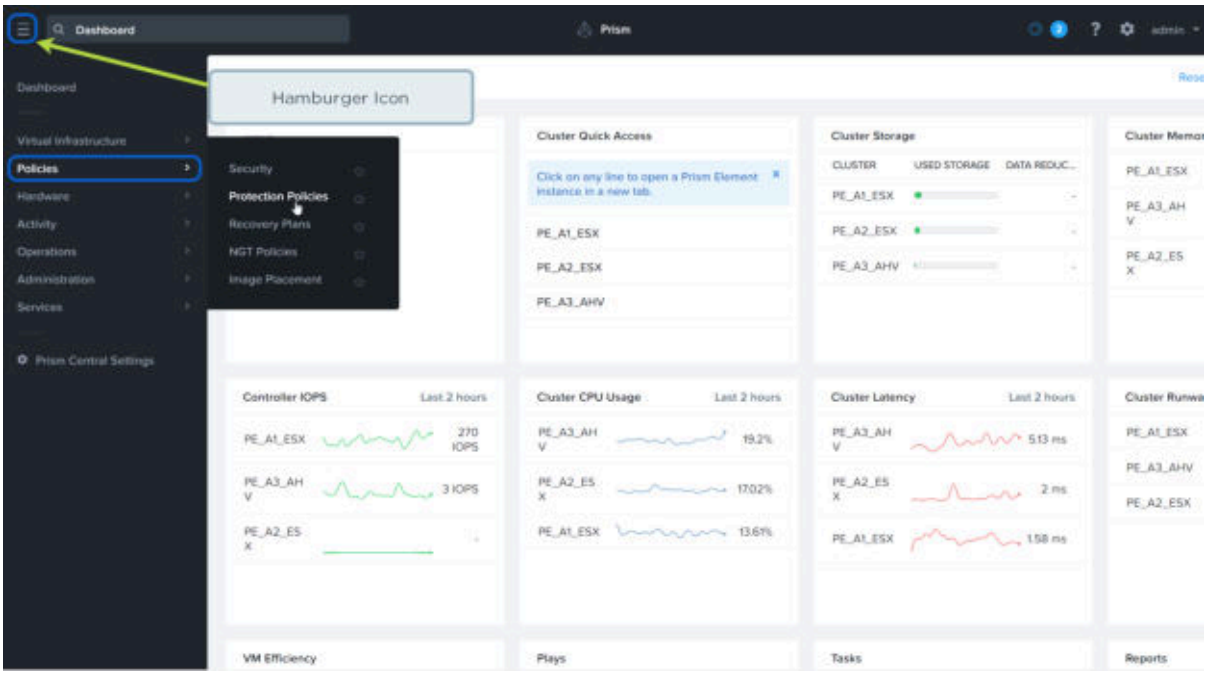


Figure 81: Protection Policy Configuration: Protection Policies

3. Select the protection policy that you want to clone.
4. Click **Clone** from the **Actions** drop-down menu.
5. Make the required changes on the **Clone Protection Policy** page.
   For information about the fields on the page, see:
   - Creating a Protection Policy with an Asynchronous Replication Schedule (Leap) on page 42
   - Creating a Protection Policy with a NearSync Replication Schedule (Leap) on page 104
   - Creating a Protection Policy with the Synchronous Replication Schedule (Leap) on page 119
6. Click **Save**.

Updating a Protection Policy

You can modify an existing protection policy in Prism Central. To update an existing protection policy, perform the following procedure.

Procedure

1. Log on to the Prism Central web console.

2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Protection Policies** in the left pane.

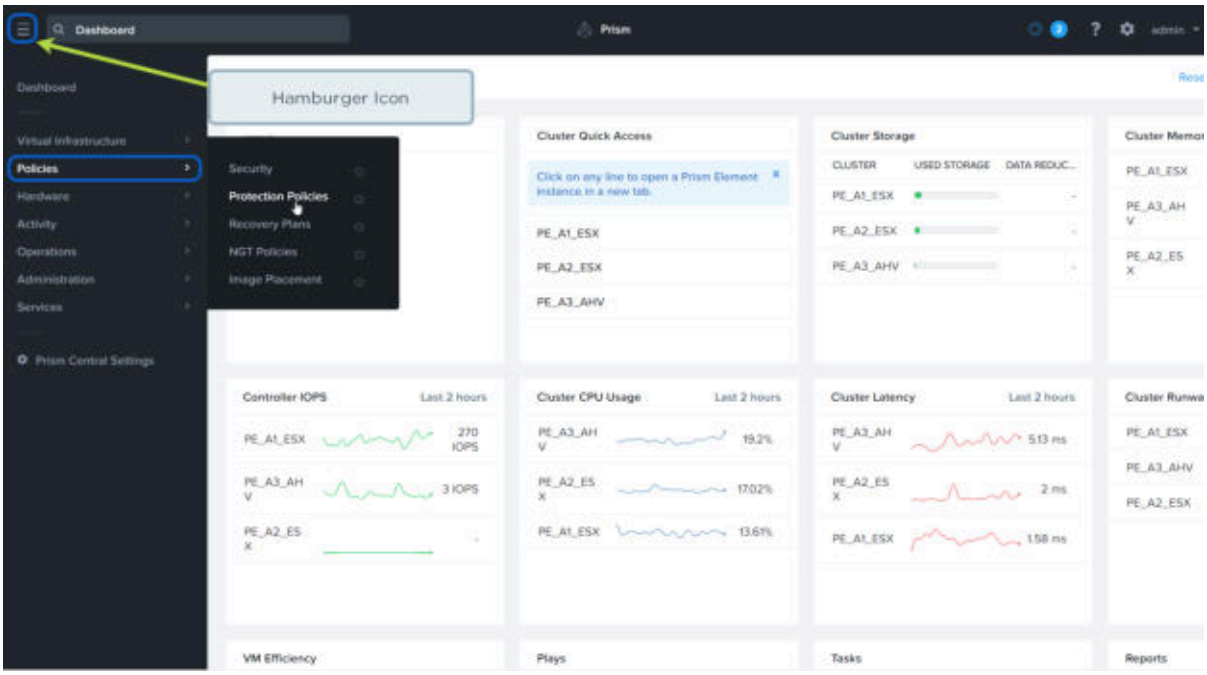


Figure 82: Protection Policy Configuration: Protection Policies

3. Select the protection policy that you want to update.

4. Click **Update** from the **Actions** drop-down menu.

5. Make the required changes on the **Update Protection Policy** page.
   For information about the fields on the page, see:

   - Creating a Protection Policy with an Asynchronous Replication Schedule (Leap) on page 42
   - Creating a Protection Policy with a NearSync Replication Schedule (Leap) on page 104
   - Creating a Protection Policy with the Synchronous Replication Schedule (Leap) on page 119

6. Click **Save**.

Finding the Protection Policy of a Guest VM

You can use the data protection focus on the **VMs** page to determine the protection policies to which a guest VM belongs. To determine the protection policy, perform the following procedure.

Procedure

1. Log on to the Prism Central web console.

2. Click the hamburger icon at the top-left corner of the window. Go to **Virtual Infrastructure** > **VMs** > **List** in the left pane.

3. Click **Data Protection** from the **Focus** menu at the top-right corner.
   The **Protection Policy** column that is displayed shows the protection policy to which the guest VMs belong.

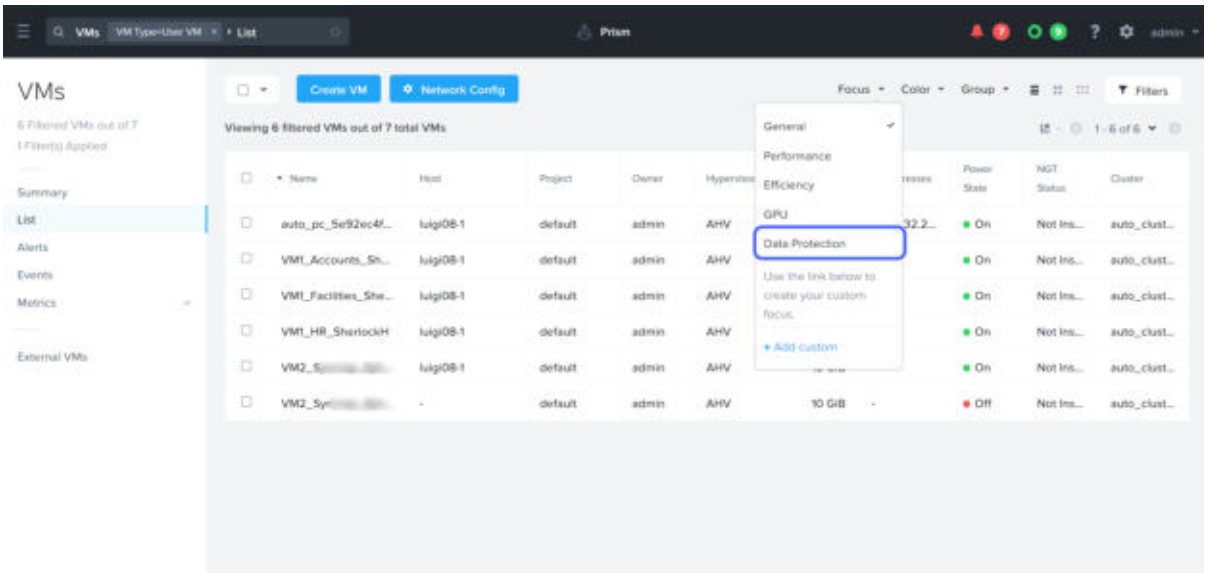


Figure 83: Focus

Recovery Plan Management

A recovery plan orchestrates the recovery of protected VMs at the recovery site. Recovery plans are predefined procedures (runbooks) that use stages to enforce VM power-on sequence. You can also specify the inter-stage delays to recover applications.

When you create, update, or delete a recovery plan, it synchronizes to the recovery sites and works bidirectionally. For information about how Leap determines the list of sites for synchronization, see Entity Synchronization Between Paired Availability Zones on page 151. After a failover from the primary site to a recovery site, you can failback to the primary site by using the same recovery plan.

Recovery plans are independent of protection policies and do not reference protection policies in their configuration information. Also, they do not create recovery points. While the process of planned failover includes the creation of a recovery point so that the latest data can be used for recovery, unplanned and test failovers rely on the availability of the required recovery points at the recovery site. A recovery plan therefore requires the guest VMs in the recovery plan to also be associated with a protection policy.

Adding Guest VMs Individually to a Recovery Plan

You can also add guest VMs individually to a recovery plan from the **VMs** page, without the use of a VM category. To add VMs individually to a recovery plan, perform the following procedure.

Procedure

1. Log on to the Prism Central web console.

2. Click the hamburger icon at the top-left corner of the window. Go to **Virtual Infrastructure** > **VMs** in the left pane.

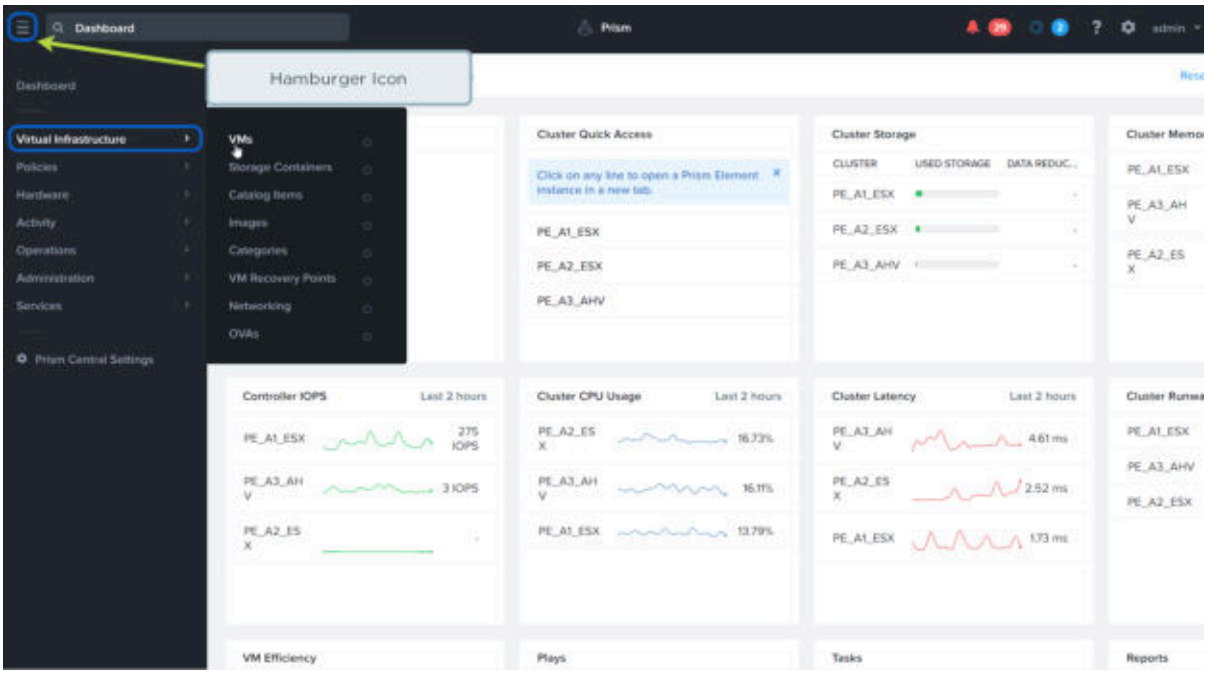


Figure 84: Virtual Infrastructure: VMs

3. Select the guest VMs that you want to add to a recovery plan.

4. Click **Add to Recovery Plan** from the **Actions** drop-down menu.

5. Select the recovery plan where you want to add the guest VMs in the **Add to Recovery Plan** page.

6. Click **Add**.
   The **Update Recovery Plan** page appears. Make the required changes to the recovery plan. For information about the various fields and options, see Creating a Recovery Plan (Leap) on page 60.

Removing Guest VMs Individually from a Recovery Plan

You can also remove guest VMs individually from a recovery plan. To remove guest VMs individually from a recovery plan, perform the following procedure.

Procedure

1. Log on to the Prism Central web console.

2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.

3. Select the recovery plan from which you want to remove guest VM.

4. Click **Update** from the **Actions** drop-down menu.
   The **Update Recovery Plan** page appears. Make the required changes to the recovery plan. For information about the various fields and options, see Creating a Recovery Plan (Leap) on page 60.

Updating a Recovery Plan

You can update an existing recovery plan. To update a recovery plan, perform the following procedure.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select the recovery plan that you want to update.
4. Click **Update** from the **Actions** drop-down menu.
   The **Update Recovery Plan** dialog box appears. Make the required changes to the recovery plan. For information about the various fields and options, see Creating a Recovery Plan (Leap) on page 60.

Validating a Recovery Plan

You can validate a recovery plan from the recovery site. Recovery plan validation does not perform a failover like the test failover does, but reports warnings and errors. To validate a recovery plan, perform the following procedure.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select the recovery plan that you want to validate.
4. Click **Validate** from the **Actions** drop-down menu.
5. In the **Validate Recovery Plan** page, do the following.
   a. In **Primary Location**, select the primary location.
   b. In **Recovery Location**, select the recovery location.
   c. Click **Proceed**.
   The validation process lists any warnings and errors.
6. Click **Back**.
   A summary of the validation is displayed. You can close the dialog box.
7. To return to the detailed results of the validation, click the link in the **Validation Errors** column.
   The selected recovery plan is validated for its correct configuration. The updated recovery plan starts synchronizing to the recovery Prism Central.

Protection and Manual DR (Leap)

Manual data protection involves manually creating recovery points, replicating recovery points, and recovering the guest VMs at the recovery site. You can also automate some of these tasks. For example, the last step—that of manually recovering guest VMs at the recovery site—can be performed by a recovery plan while the underlying recovery point creation and replication can be performed by protection policies. Conversely, you can configure protection policies to

automate recovery point creation and replication and then recover guest VMs manually at the recovery site.

Creating Recovery Points Manually (Out-of-Band Snapshots)

About this task

To create recovery points manually, do the following.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Compute & Storage** > **VMs** > **List** in the left pane.
3. Select the guest VMs for which you want to create a recovery point.
4. Click **Create Recovery Point** from the **Actions** drop-down menu.
5. To verify that the recovery point is created, click the name of the VM, click the **Recovery Points** tab, and verify that a recovery point is created.

Replicating Recovery Points Manually

You can manually replicate recovery points only from the availability zone (site) where the recovery points exist.

About this task

To replicate recovery points manually, do the following.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Compute & Storage** > **VMs** > **List** in the left pane.
3. Click the guest VMs whose recovery point you want to replicate, and then click **Recovery Points** in the left.

   The Recovery Points view lists all the recovery points of the VM.
4. Select the recovery points that you want to replicate.
5. Click **Replicate** from the **Actions** drop-down menu.

   In the **Replicate** dialog box, do these steps.

   a. In **Recovery Location**, select the location where you want to replicate the recovery point.

   b. In **Target Cluster**, select the cluster where you want to replicate the recovery point.

   c. Click **Replicate Recovery Point**.

Manual Recovery of Guest VMs

After manually creating recovery points, and then replicating those recovery points, you can manually recover the guest VMs from those recovery points by either of the following methods.

> Note: You can manually recover guest VMs from both manually and automatically created recovery points.

Cloning the guest VM from a recovery point (out-of-place restore)

Out-of-place restore enables you to create a guest VM clone (with different UUIDs) from each selected recovery point of the guest VM. The operation creates a copy of the guest VM in the same Nutanix cluster without overwriting the original guest VM. The guest VM clone remain separate from the original guest VM. For example, if you select cloning of a guest VM , a new VM with different UUID is created on the specified site.

Reverting the guest VM to an older recovery point (in-place restore)

In-place restore enables you to revert the guest VM to its previous recovery point state. The operation recreates the guest VM in the same Nutanix cluster by overwriting the original guest VM. The new guest VM retains the properties like UUIDs, network routing, MAC addresses, hostnames of the original guest VM. For example, if you select reverting a guest VM, a new VM is created on the same site with the same properties.

Recovering a Guest VM Manually (Clone)

About this task

To recover a guest VM manually by cloning it from a recovery point, do the following.

> Note: This method is available as **Restore** in the previous versions of AOS.

Before you begin

Consider the following limitations before performing cloning operation on guest VMs.

- The guest VMs recover without a vNIC if the recovery is performed at the remote site.
- The guest VMs recover without VM categories.
- You must reconfigure Nutanix Guest Tools (NGT) on each recovered guest VM.

  For installing and configuring NGT and its requirements and limitations, see *Nutanix Guest Tools* in the *Prism Web Console Guide*.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Compute & Storage** > **VMs** > **List** in the left pane.
3. Click the guest VM that you want to recover (clone), and then click **Recovery Points**.

   The **Recovery Points** view lists all the recovery points of the guest VM.

   > Tip: You can also click the hamburger icon at the top-left corner and go to **Data Protection** > **VM Recovery Points**

4. Select the recovery point from which you want to recover (clone) the guest VM.
   If you want to recover (clone) multiple states of the guest VM, select multiple recovery points.

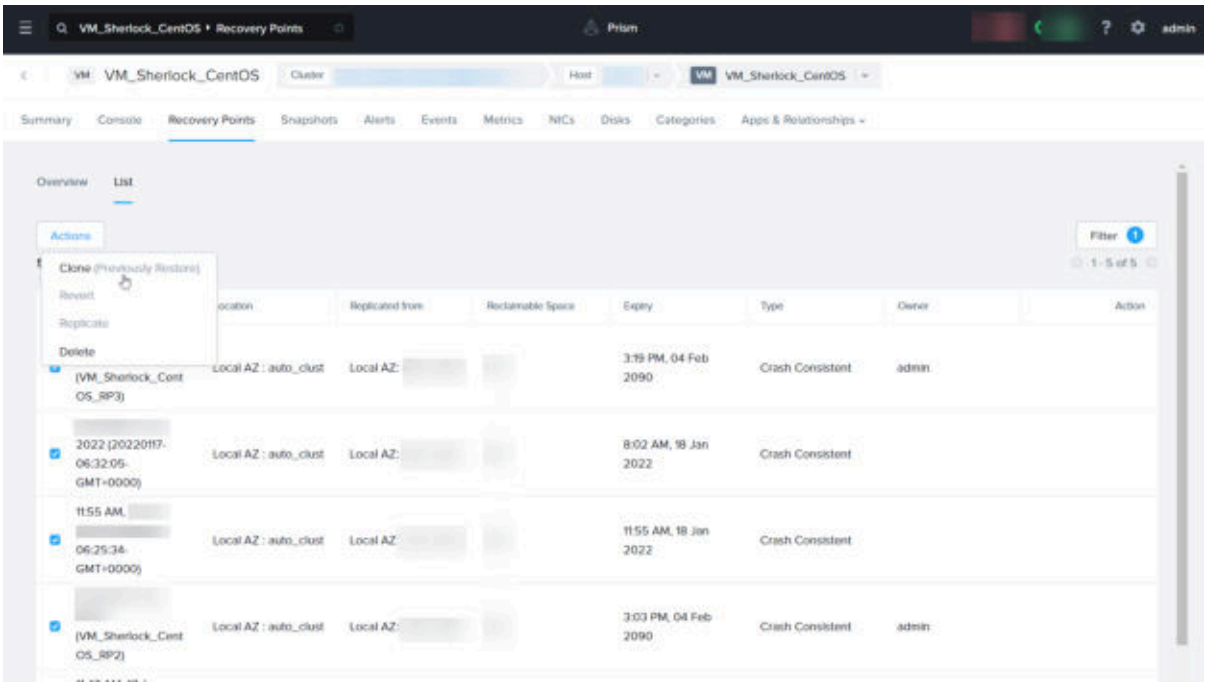


Figure 85: Select One or More Recovery Points

5. Click **Clone (Previously Restore)** from the **Actions** drop-down menu.
   In the Clone dialog box, do the following.

   a. In the text box provided for specifying a name for the guest VM clone, specify a new name or do nothing to use the automatically generated name.

   b. Click **Clone**.
      The guest VM is cloned from each selected recovery point. The guest VM clones list by the name you specified in step 5.a in **VMs** > **List**. The guest VM clones are turned off and you have to start the guest VMs manually.

Recovering a Guest VM Manually (Revert)

About this task

To recover a guest VM manually by reverting it to an older recovery point, do the following.

Before you begin

Before you can perform an in-place restore (revert) of a guest VM, ensure that you meet the following Nutanix Software requirements.

- The Nutanix cluster must be running AOS version 6.0 or newer.
- Prism Central is running on version pc.2022.1 or newer.

For successful and efficient in-place restore (revert) of the guest VM along with its configuration, refer to the following tables. Each of the following tables describes settings and the possible change scenarios that can fail the revert operation.

Table 24: NGT

This table describes the NGT configuration on the original guest VM and the selected recovery point. It describes the scenarios where changes in NGT configuration fail the revert operation or restore the guest VM without NGT configuration.

Entity	Original guest VM	Selected recovery point	Original guest VM after the recovery point (before revert)	Revert operation	Guest VM after revert
Guest VM	NGT configured	NGT configured. The NGT information is the same as that on the original guest VM.	No change to the NGT configuration	Succeeds	Need not reconfigure NGT on the restored guest VM.
	NGT configured	NGT not configured, or the NGT information is different than that on the original guest VM.	NGT configured	Fails because there is a mismatch between NGT information on the original guest VM and the selected recovery point.	Not applicable
	NGT configured	NGT configured	NGT not configured	Succeeds	NGT on the restored guest VM is disabled. You must enable it manually. For enabling NGT, see *Nutanix Guest Tools* in the *Prism Web Console Guide*.
	NGT not configured	NGT not configured, or the NGT information is different than that on the original guest VM.	No change to the NGT configuration	Succeeds	Not applicable

Entity	Original guest VM	Selected recovery point	Original guest VM after the recovery point (before revert)	Revert operation	Guest VM after revert
	NGT configured	NGT configured but the NGT information available is of an older NGT version.	Newer NGT version configured	Fails because there is a mismatch between NGT information on the original guest VM and the selected recovery point.	Not applicable

Table 25: VM Categories

This table describes the VM categories attached with the original guest VM and the selected recovery point. It describes the scenarios where changes in VM categories fail the revert operation or restore the guest VM without VM categories.

Entity	Original guest VM	Selected recovery point	Revert operation	Guest VM after revert
Guest VM	VM category configured (for example, catX:valX)	VM category configured but its information is unidentical (for example, catY:valY)	Succeeds	Retains the VM category of the original guest VM (catX:valX).
	VM category not configured.	VM category configured (for example, catY:valY)	Succeeds	Does not retain VM category because the original guest VM did not have VM category configured. You must attach a VM category manually.

Table 26: Virtual Networks

This table describes the virtual networks attached with the original guest VM and the selected recovery point. It describes the scenarios where changes in virtual networks fail the revert operation or restore the guest VM without virtual networks.

Entity	Original guest VM	Selected recovery point	Network available during revert operation	Revert operation	Guest VM after revert
Guest VM	Network configured (for example, vLANx)	Network configured (for example, vLANx)	The Nutanix cluster has vLANx configured.	Succeeds	Not applicable
	Network configured (for example, vLANx)	Network configured is different (for example, vLANy)	The Nutanix cluster has both vLANx and vLANy configured.	Succeeds	Retains the network configured on the selected recovery point (vLANy).
	No network configured	Network configured (for example, vLANy)	The Nutanix cluster has vLANy configured.	Succeeds	Not applicable
	No network configured	Network configured (for example, vLANy)	The Nutanix cluster has a diiferent network (for example, vLANx) configured.	Fails	Not applicable

Table 27: vNUMA nodes and vGPU AHV Features

This table describes the AHV Features (vNUMA nodes and vGPU) attached with the original guest VM and the selected recovery point. It describes the scenarios where changes in vNUMA nodes and vGPU configuration fail the revert operation or restore the guest VM without vNUMA nodes and vGPU configuration.

Entity	Original guest VM	Selected recovery point	Features available during revert operation	Revert operation	Guest VM after revert
Guest VM	vNUMA nodes and vGPU configured	vNUMA nodes and vGPU configured. The feature information is the same as that on the original guest VM.	The Nutanix cluster vNUMA nodes and vGPU configured	Succeeds	Not applicable

Entity	Original guest VM	Selected recovery point	Features available during revert operation	Revert operation	Guest VM after revert
	vNUMA nodes and vGPU configured	vNUMA nodes and vGPU configured. The feature information is different than that on the original guest VM.	The Nutanix cluster vNUMA nodes and vGPU configured	Succeeds	Does not retain vNUMA nodes and vGPU configuration.
	vNUMA nodes and vGPU not configured	vNUMA nodes and vGPU configured	Yes	Succeeds	Retains vNUMA nodes and vGPU configuration. If you encounter missing internal configuration, contact Nutanix Support.
	vNUMA nodes and vGPU not configured	vNUMA nodes and vGPU configured	No	Fails because the Nutanix cluster doesn't support GPU.	Not applicable

Table 28: Role Based Access Control (RBAC)

This table describes the user roles supporting in-place restore (revert) of a guest VM.

Entity	Legacy Administrator	Cluster Administrator (Prism Administrator)
Guest VM	Supported	Not Supported

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Compute & Storage** > **VMs** > **List** in the left pane.
3. Click the guest VM that you want to recover (revert) to a recovery point, and then click **Recovery Points**.

   The **Recovery Points** view lists all the recovery points of the guest VM.

   > Tip: You can also click the hamburger icon at the top-left corner and go to **Data Protection** > **VM Recovery Points**

4. Select the recovery point from which you want to recover (revert) the guest VM.

   > Note: Do not select more than one recovery point. If you select two or more recovery points, you cannot perform **Revert** operation.

5. Click **Revert** from the **Actions** drop-down menu.

   In the Revert dialog box, verify the recovery point details and then click **Revert**.

   The selected guest VM turns off. It is unregistered from the inventory, updated with files copied from the selected recovery point, and then registered to the inventory. The recovered (reverted) guest VM lists by the name (Revert-Recoverypoint) in **Compute & Storage** > **VMs** > **List**. If there are more than one guest VM by the name (Revert-Recoverypoint), see the timestamps of the recovered (reverted) guest VMs to identify the correct guest VMs. The recovered (reverted) guest VM is turned off and you have to start the VM manually.

   After reverting the guest VM to its previous state, some settings of the original guest VM needs reconfiguration for the recovered guest VMs to work efficiently. The following table describes the settings of the original guest VM and whether or not they need reconfiguration after revert.

Entity Synchronization Between Paired Availability Zones

When paired with each other, availability zones (sites) synchronize disaster recovery (DR) configuration entities. Paired sites synchronize the following DR configuration entities.

Protection Policies

A protection policy is synchronized whenever you create, update, or delete the protection policy.

Recovery Plans

A recovery plan is synchronized whenever you create, update, or delete the recovery plan. The list of availability zones (sites) to which the on-prem must synchronize a recovery plan is derived from the guest VMs that are included in the recovery plan. The guest VMs used to derive the availability zone list are VM categories and individually added guest VMs.

If you specify VM categories in a recovery plan, Leap determines which protection policies use those VM categories, and then synchronizes the recovery plans to the availability zones specified in those protection plans.

If you include guest VMs individually (without VM categories) in a recovery plan, Leap uses the recovery points of those guest VMs to determine which protection policies created those recovery points, and then synchronizes the recovery plans to the availability zones (sites) specified in those protection policies. If you create a recovery plan for VM categories or guest VMs that are not associated with a protection policy, Leap cannot determine the availability zone list and therefore cannot synchronize the recovery plan. If a recovery plan includes only individually added guest VMs and a protection policy associated with a guest VM has not yet created guest VM recovery points, Leap cannot synchronize the recovery plan to the availability zone specified in that protection policy. However, recovery plans are monitored every 15 minutes for the availability of recovery points that can help derive availability zone information. When recovery points become available, the paired on-prem site derives the availability zone by the process described earlier and synchronizes the recovery plan to the availability zone.

VM Categories used in Protection Policies and Recovery Plans

A VM category is synchronized when you specify the VM category in a protection policy or recovery plan.

Issues such as a loss of network connectivity between paired availability zones or user actions such as unpairing of availability zones followed by repairing of those availability zones can affect VM synchronization.

> Tip: Nutanix recommends to unprotect all the VMs on the availability zone before unpairing it to avoid getting into a state where the entities have stale configurations after repairing of availability zones.

If you update guest VMs in either or both availability zones before such issues are resolved or before unpaired availability zones are paired again, VM synchronization is not possible. Also, during VM synchronization, if a guest VM cannot be synchronized because of an update failure or conflict (for example, you updated the same VM in both availability zones during a network connectivity issue), no further VMs are synchronized. Entity synchronization can resume only after you resolve the error or conflict. To resolve a conflict, use the **Entity Sync** option, which is available in the web console. Force synchronization from the availability zone that has the desired configuration. Forced synchronization overwrites conflicting configurations in the paired availability zone.

> Note: Forced synchronization cannot resolve errors arising from conflicting values in guest VM specifications (for example, the paired availability zone already has a VM with the same name).

If you do not update entities before a connectivity issue is resolved or before you pair the availability zones again, the synchronization behavior described earlier resumes. Also, pairing previously unpaired availability zones trigger an automatic synchronization event. For recommendations to avoid facing such issues, see Entity Synchronization Recommendations (Leap) on page 152.

Entity Synchronization Recommendations (Leap)

Consider the following recommendations to avoid inconsistencies and the resulting synchronization issues.

- During network connectivity issues, do not update entities at both the availability zones (sites) in a pair. You can safely make updates at any one site. After the connectivity issue is resolved, force synchronization from the site in which you made updates. Failure to adhere to this recommendation results in synchronization failures.

  You can safely create entities at either or both the sites as long as you do not assign the same name to entities at the two sites. After the connectivity issue is resolved, force synchronization from the site where you created entities.

- If one of the sites becomes unavailable, or if any service in the paired site is down perform force synchronization from the paired availability zone after the issue is resolved.

Forcing Entity Synchronization (Leap)

Entity synchronization, when forced from an availability zone (site), overwrites the corresponding entities in paired sites. Forced synchronization also creates, updates, and removes those entities from paired sites.

About this task

The availability zone (site) to which a particular entity is forcefully synchronized depends on which site requires the entity (see Entity Synchronization Between Paired Availability Zones on page 151). To avoid inadvertently overwriting required entities, ensure to force VM synchronization from the site in which the entities have the desired configuration.

If a site is paired with two or more availability zones (sites), you cannot select one or more sites with which to synchronize entities.

To force entity synchronization, do the following.

Procedure

1. Log on to the Prism Central web console.
2. Click the settings button (gear icon) at the top-right corner of the window.
3. Click **Entity Sync** in the left pane.
4. In the **Entity Sync** dialog box, review the message at the top of the dialog box, and then do the following.
   a. To review the list of entities that will be synchronized to an **AVAILABILITY ZONE**, click the number of **ENTITIES** adjacent to an availability zone.
   b. After you review the list of entities, click **Back**.
5. Click **Sync Entities**.

3

PROTECTION AND DR BETWEEN ON-PREM SITE AND XI CLOUD SERVICE (XI LEAP)

Xi Leap is essentially an extension of Leap to Xi Cloud Services. Xi Leap protects your guest VMs and orchestrates their disaster recovery (DR) to Xi Cloud Services when events causing service disruption occur at the primary availability zone (site). For protection of your guest VMs, protection policies with Asynchronous and NearSync replication schedules generate and replicate recovery points to Xi Cloud Services. Recovery plans orchestrate DR from the replicated recovery points to Xi Cloud Services.

Protection policies create a recovery point—and set its expiry time—in every iteration of the specified time period (RPO). For example, the policy creates a recovery point every 1 hour for an RPO schedule of 1 hour. The recovery point expires at its designated expiry time based on the retention policy—see step 3 in Creating a Protection Policy with Asynchronous Replication Schedule (Xi Leap) on page 198. If there is a prolonged outage at a site, the Nutanix cluster retains the last recovery point to ensure you do not lose all the recovery points. For NearSync replication (lightweight snapshot), the Nutanix cluster retains the last full hourly snapshot. During the outage, the Nutanix cluster does not clean up the recovery points due to expiry. When the Nutanix cluster comes online, it cleans up the recovery points that are past expiry immediately.

For High Availability of a guest VM, Leap can enable replication of recovery points to one or more sites. A protection policy can replicate recovery points to maximum two sites. One of the two sites can be in cloud (Xi Cloud Services). For replication to Xi Cloud Services, you must add a replication schedule between the on-prem site and Xi Cloud Services. You can set up the on-prem site and Xi Cloud Services in the following arrangements.

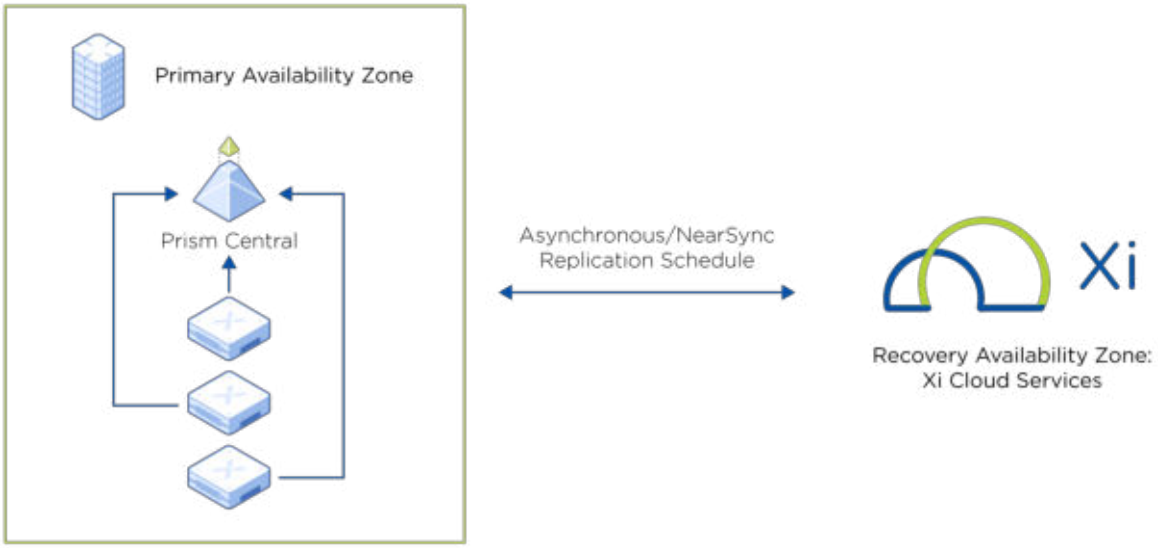


Figure 86: The Primary Nutanix Cluster at on-prem site and recovery Xi Cloud Services

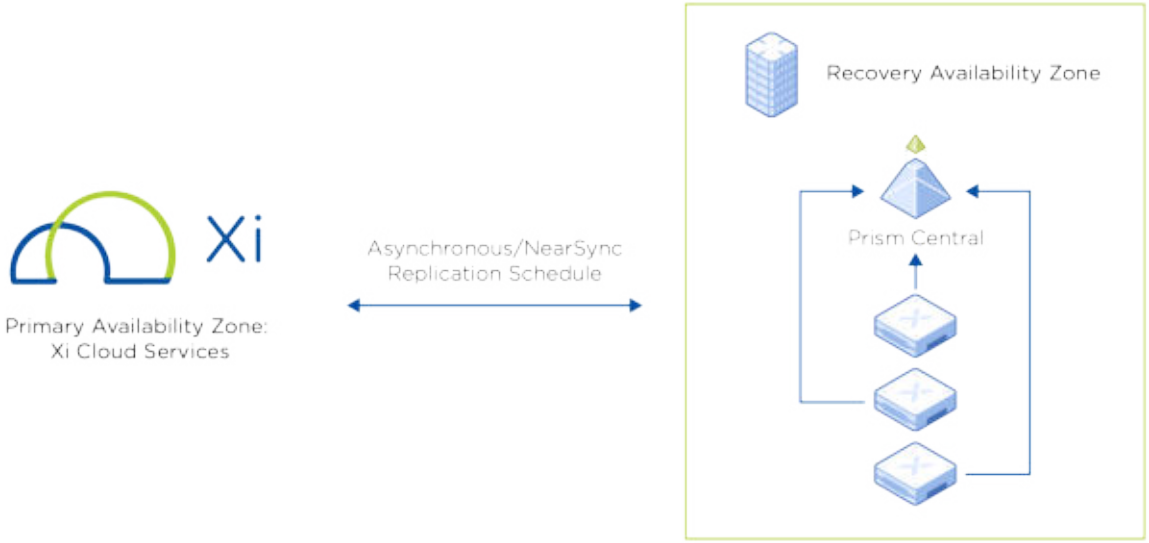


Figure 87: The Primary Xi Cloud Services and recovery Nutanix Cluster at on-prem site

The replication schedule between an on-prem site and Xi Cloud Services enables DR to Xi Cloud Services. To enable performing DR to Xi Cloud Services, you must create a recovery plan. In addition to performing DR from AHV clusters to Xi Cloud Services (only AHV), you can also perform cross-hypervisor disaster recovery (CHDR)—DR from ESXi clusters to Xi Cloud Services.

The protection policies and recovery plans you create or update synchronize continuously between the on-prem site and Xi Cloud Services. The reverse synchronization enables you to create or update entities (protection policies, recovery plans, and guest VMs) at either the primary or the recovery site.

This section describes protection of your guest VMs and DR from Xi Cloud Services to a Nutanix cluster at the on-prem site. In Xi Cloud Services, you can protect your guest VMs and DR to a Nutanix cluster at only one on-prem site. For information about protection of your

guest VMs and DR to Xi Cloud Services, see Protection and DR between On-Prem Sites (Leap) on page 18.

Xi Leap Requirements

The following are the general requirements of Xi Leap. Along with the general requirements, there are specific requirements for protection with the following supported replication schedules.

- For information about the on-prem node, disk and Foundation configurations required to support Asynchronous and NearSync replication schedules, see On-Prem Hardware Resource Requirements on page 15.
- For specific requirements of protection with Asynchronous replication schedule (1 hour or greater RPO), see Asynchronous Replication Requirements (Xi Leap) on page 196.
- For specific requirements of protection with NearSync replication schedule (1–15 minutes RPO), see NearSync Replication Requirements (Xi Leap) on page 231.

License Requirements

The AOS license required depends on the features that you want to use. For information about the features that are available with an AOS license, see Software Options.

Hypervisor Requirements

The underlying hypervisors required differ in all the supported replication schedules. For more information about underlying hypervisor requirements for the supported replication schedules, see:

- Asynchronous Replication Requirements (Xi Leap) on page 196
- NearSync Replication Requirements (Xi Leap) on page 231

Nutanix Software Requirements

- Each on-prem availability zone (site) must have a Leap enabled Prism Central instance. To enable Leap in Prism Central, see Enabling Leap for On-Prem Site on page 36.

  > Note: If you are using ESXi, register at least one vCenter Server to Prism Central. You can also register two vCenter Servers, each to a Prism Central at different sites. If you register both the Prism Central to the single vCenter Server, ensure that each ESXi cluster is part of different datacenter object in vCenter.

- The on-prem Prism Central and its registered Nutanix clusters (Prism Element) must be running on the supported AOS versions. For more information about the required versions for the supported replication schedules, see:
  - Asynchronous Replication Requirements (Xi Leap) on page 196
  - NearSync Replication Requirements (Xi Leap) on page 231

  > Tip:
  >
  > Nutanix supports replications between the all the latest supported LTS and STS released AOS versions. To check the list of the latest supported AOS versions, see

KB-5505. To determine if the AOS versions currently running on your clusters are EOL, see the EOL document .

Upgrade the AOS version to the next available supported LTS/STS release. To determine if an upgrade path is supported, check the Upgrade Paths page before you upgrade the AOS.

> Note: If both clusters have different AOS versions that are EOL, upgrade the cluster with lower AOS version to match the cluster with higher AOS version and then perform the upgrade to the next supported LTS version.

For example, the clusters are running AOS versions 5.5.x and 5.10.x respectively. Upgrade the cluster on 5.5.x to 5.10.x. After both the clusters are on 5.10.x, proceed to upgrade each cluster to 5.15.x (supported LTS). Once both clusters are on 5.15.x you can upgrade the clusters to 5.20.x or newer.

Nutanix recommends that both the primary and the replication clusters or sites run the same AOS version.

User Requirements

You must have one of the following roles in Xi Cloud Services.

- User admin
- Prism Central admin
- Prism Self Service admin
- Xi admin

Firewall Port Requirements

To allow two-way replication between an on-prem Nutanix cluster and and Xi Cloud Services, you must enable certain ports in your external firewall. To know about the required ports, see *Disaster Recovery - Leap* in Port Reference.

Networking Requirements

Requirements for static IP address preservation after failover

You can preserve one IP address of a guest VM (with static IP address) for its failover (DR) to an IPAM network. After the failover, the other IP addresses of the guest VM have to be reconfigured manually. To preserve an IP address of a guest VM (with static IP address), ensure that:

> Caution: By default, you cannot preserve statically assigned DNS IP addresses after failover (DR) of guest VMs. However, you can create custom in-guest scripts to preserve the statically assigned DNS IP addresses. For more information, see Creating a Recovery Plan (Xi Leap) on page 209.

- Both the primary and the recovery Nutanix clusters run AOS 5.11 or newer.
- The protected guest VMs have Nutanix Guest Tools (NGT) version 1.5 or newer installed.

  For information about installing NGT, see *Nutanix Guest Tools* in *Prism Web Console Guide*.
- The protected guest VMs have at least one empty CD-ROM slot.

  The empty CD-ROM is required for mounting NGT at the recovery site.

- The protected guest VMs have NetworkManager command-line tool (nmcli) version 0.9.10.0 or newer installed.

  Also, the NetworkManager must manage the networks on Linux VMs. To enable NetworkManager on a Linux VM, in the interface configuration file, set the value of the `NM_CONTROLLED` field to `yes`. After setting the field, restart the network service on the VM.

  > Tip: In CentOS, the interface configuration file is `/etc/sysconfig/network-scripts/ifcfg-eth0`.

Requirements for static IP address mapping of guest VMs between source and target virtual networks

You can explicitly define IP addresses for protected guest VMs that have static IP addresses at the primary site. On recovery, such guest VMs retain the explicitly defined IP address. To map static IP addresses of guest VMs between source and target virtual networks, ensure that:

- Both the primary and the recovery Nutanix clusters run AOS 5.17 or newer.
- The protected guest VMs have static IP addresses at the primary site.
- The protected guest VMs have Nutanix Guest Tools (NGT) version 1.5 or newer installed.

  For information about installing NGT, see *Nutanix Guest Tools* in *Prism Web Console Guide*.
- The protected guest VMs have at least one empty CD-ROM slot.

  The empty CD-ROM is required for mounting NGT at the recovery site.
- The protected guest VMs can reach the Controller VM from both the sites.
- The recovery plan selected for failover has VM-level IP address mapping configured.

Virtual Network Design Requirements

You can design the virtual subnets that you plan to use for DR to the recovery site so that they can accommodate the guest VMs running in the source virtual network.

- To use a virtual network as a recovery virtual network, ensure that the virtual network meets the following requirements.
  - The network prefix is the same as the network prefix of the source virtual network. For example, if the source network address is 192.0.2.0/24, the network prefix of the recovery virtual network must also be 24.
  - The gateway IP address is the same as the gateway IP address in the source network. For example, if the gateway IP address in the source virtual network 192.0.2.0/24 is 192.0.2.10, the last octet of the gateway IP address in the recovery virtual network must also be 10.
- To use a single Nutanix cluster as a target for DR from multiple primary Nutanix clusters, ensure that the number of virtual networks on the recovery cluster is equal to the sum of the number of virtual networks on the individual primary Nutanix clusters. For example, if there are two primary Nutanix clusters, with one cluster having `m` networks and the other cluster having `n` networks, ensure that the recovery cluster has `m + n` networks. Such a design ensures that all recovered VMs attach to a network.
- After the recovery of guest VMs to Xi Cloud Services, ensure that the router in your primary site stops advertising the subnet that hosted the guest VMs.

- The protected guest VMs and Prism Central VM must be on different networks.

  If protected guest VMs and Prism Central VM are on the same network, the Prism Central VM becomes inaccessible when the route to the network is removed after failover.

- Xi Cloud Services supports the following third-party VPN gateway solutions.

  - CheckPoint
  - Cisco ASA
  - PaloAlto

    > Note: If you are using the Palo Alto VPN gateway solution, set the **MTU** value to **1356** in the Tunnel Interface settings. The replication fails for the default **MTU** value (**1427**).

  - Juniper SRX
  - Fortinet
  - SonicWall
  - VyOS

Additional Requirements

- Both the primary and recovery Nutanix clusters must have an external IP address.
- Both the primary and recovery Prism Centrals and Nutanix clusters must have a data services IP address.
- The Nutanix cluster that hosts the Prism Centrals must meet the following requirements.

  - The Nutanix cluster must be registered to the Prism Central instance.
  - The Nutanix cluster must have an iSCSI data services IP address configured on it.
  - The Nutanix cluster must also have sufficient memory to support a hot add of memory to all Prism Central nodes when you enable Leap. A small Prism Central instance (4 vCPUs, 16 GB memory) requires a hot add of 4 GB, and a large Prism Central instance (8 vCPUs, 32 GB memory) requires a hot add of 8 GB. If you enable Nutanix Flow, each Prism Central instance requires an extra hot-add of 1 GB.

- Each node in a scaled-out Prism Central instance must have a minimum of 4 vCPUs and 16 GB memory.

  For more information about the scaled-out deployments of a Prism Central, see Leap Terminology on page 8.

- The protected guest VMs must have Nutanix VM mobility drivers installed.

  Nutanix VM mobility drivers are required for accessing the guest VMs after failover. Without Nutanix VM mobility drivers, the guest VMs become inaccessible after a failover.

Xi Leap Limitations

Consider the following general limitations before configuring protection and disaster recovery (DR) with Xi Leap. Along with the general limitations, there are specific limitations of protection with the following supported replication schedules.

- For specific limitations of protection with Asynchronous replication schedule (1 hour or greater RPO), see Asynchronous Replication Limitations (Xi Leap) on page 198.
- For specific limitations of protection with NearSync replication schedule (1–15 minutes RPO), see NearSync Replication Limitations (Xi Leap) on page 233.

Virtual Machine Limitations

- You cannot start or replicate the following guest VMs at Xi Cloud Services.
  - VMs configured with a GPU resource.
  - VMs configured with four or more vNUMA sockets.
  - VMs configured with more than 24 vCPUs.
  - VMs configured with more than 128 GB memory.
- You cannot deploy witness VMs.
- You cannot protect multiple guest VMs that use disk sharing (for example, multi-writer sharing, Microsoft Failover Clusters, Oracle RAC).
- You cannot protect VMware fault tolerance enabled guest VMs.
- You cannot recover vGPU console enabled guest VMs efficiently.

  When you perform DR of vGPU console-enabled guest VMs, the VMs recover with the default VGA console (without any alert) instead of vGPU console. The guest VMs fail to recover when you perform cross-hypervisor disaster recovery (CHDR).
- You cannot configure NICs for a guest VM across both the virtual private clouds (VPC).

  You can configure NICs for a guest VM associated with either production or test VPC.

Volume Groups Limitation

You cannot protect volume groups.

Network Segmentation Limitation

You cannot apply network segmentation for DR. If you enable network segmentation for DR, the replications fail.

Self-Service Restore

You cannot perform self-service restore.

Virtual Network Limitations

Although there is no limit to the number of VLANs that you can create, only the first 500 VLANs list in drop-down of **Network Settings** while creating a recovery plan. For more information about VLANs in the recovery plan, see Nutanix Virtual Networks on page 191.

Xi Leap Recommendations

Nutanix recommends the following best practices for configuring protection and disaster recovery (DR) with Xi Leap.

General Recommendations

- Create all entities (protection policies, recovery plans, and VM categories) at the primary availability zone (site).
- Upgrade Prism Central before upgrading the Nutanix clusters (Prism Elements) registered to it.

Recommendation for Migrating Protection Domains to Protection Policies

You can protect a guest VM either with legacy DR solution (protection domain-based) or with Leap. To protect a legacy DR-protected guest VM with Leap, you must migrate the guest VM from protection domain to a protection policy. During the migration, do not delete the guest VM snapshots in the protection domain. Nutanix recommends keeping the guest VM snapshots in the protection domain until the first recovery point for the guest VM is available on Prism Central. For more information, see Migrating Guest VMs from a Protection Domain to a Protection Policy on page 248.

Recommendation for Virtual Networks

- Map the networks while creating a recovery plan in Prism Central.
- Recovery plans do not support overlapping subnets in a network-mapping configuration. Do not create virtual networks that have the same name or overlapping IP address ranges.

Xi Leap Service-Level Agreements (SLAs)

Xi Leap is essentially an extension of Leap to Xi Cloud Services. Xi Leap enables protection of your guest VMs and disaster recovery (DR) to Xi Cloud Services. With reverse synchronization, Xi Leap can protect you guest VMs and enable DR from Xi Cloud Services to a Nutanix cluster at an on-prem availability zone (site). A Nutanix cluster is essentially an AHV or an ESXi cluster running AOS. In addition to performing DR from AHV clusters to Xi Cloud Services (only AHV), you can also perform cross-hypervisor disaster recovery (CHDR)—DR from ESXi clusters to Xi Cloud Services.

You can protect your guest VMs with the following replication schedules.

- Asynchronous (1 hour or greater RPO). For information about protection with Asynchronous replication in Xi Leap, see Protection with Asynchronous Replication and DR (Xi Leap) on page 196.
- NearSync (1–15 minute RPO). For information about protection with NearSync replication in Xi Leap, see Protection with NearSync Replication and DR (Xi Leap) on page 228.

Xi Leap Views

The disaster recovery views enable you to perform CRUD options on the following types of Leap VMs.

- Configured entities (for example, availability zones, protection policies, and recovery plans)
- Created entities (for example, VMs, and recovery points)

Some views available in the Xi Cloud Services differ from the corresponding view in on-prem Prism Central. For example, the option to connect to an availability zone is on the **Availability Zones** page in an on-prem Prism Central, but not on the **Availability Zones** page in Xi Cloud Services. However, the views of both user interfaces are largely the same. This chapter describes the views of Xi Cloud Services.

Availability Zones View in Xi Cloud Services

The **Availability Zones** view lists all of your paired availability zones.

The following figure is a sample view, and the tables describe the fields and the actions that you can perform in this view.

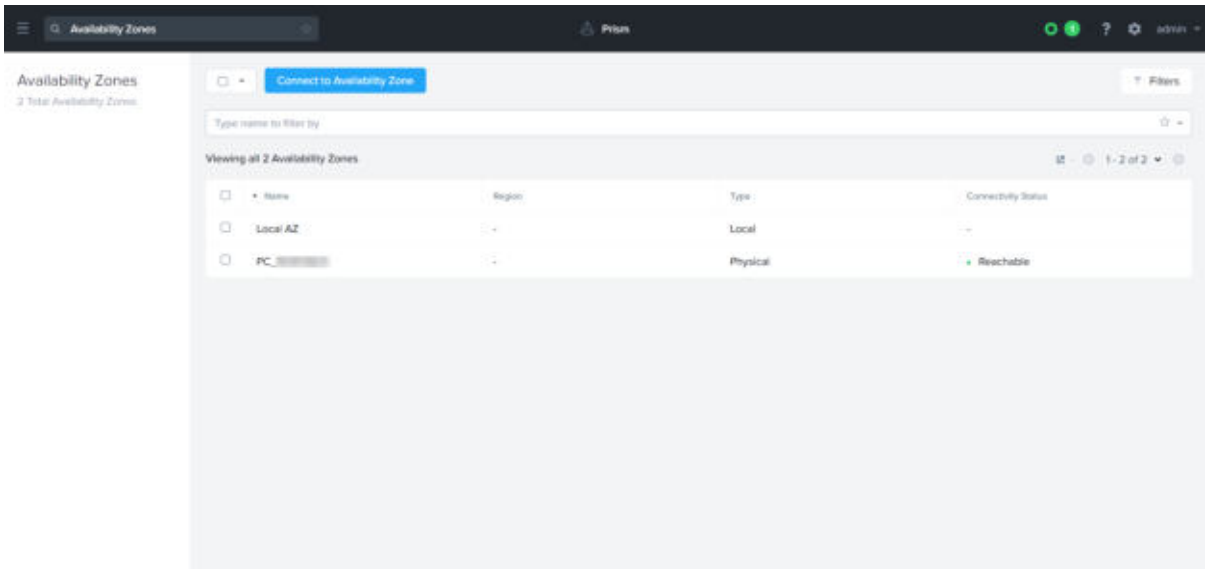


Figure 88: Availability Zones View

Table 29: Availability Zones View Fields

Field	Description
Name	Name of the availability zone.
Region	Region to which the availability zone belongs.
Type	Type of availability zone. Availability zones in Xi Cloud Services are shown as being of type Xi. Availability zones that are backed by on-prem Prism Central instances are shown to be of type physical. The availability zone that you are logged in to is shown as a local availability zone.
Connectivity Status	Status of connectivity between the local availability zone and the paired availability zone.

Table 30: Workflows Available in the Availability Zones View

Workflow	Description
Connect to Availability Zone (on-prem Prism Central only)	Connect to an on-prem Prism Central or to a Xi Cloud Services for data replication.

Table 31: Actions Available in the Actions Menu

Action	Description
Disconnect	Disconnect the remote availability zone. When you disconnect an availability zone, the pairing is removed.

Protection Policies View in Xi Cloud Services

The **Protection Policies** view lists all of configured protection policies from all availability zones.

The following figure is a sample view, and the tables describe the fields and the actions that you can perform in this view.

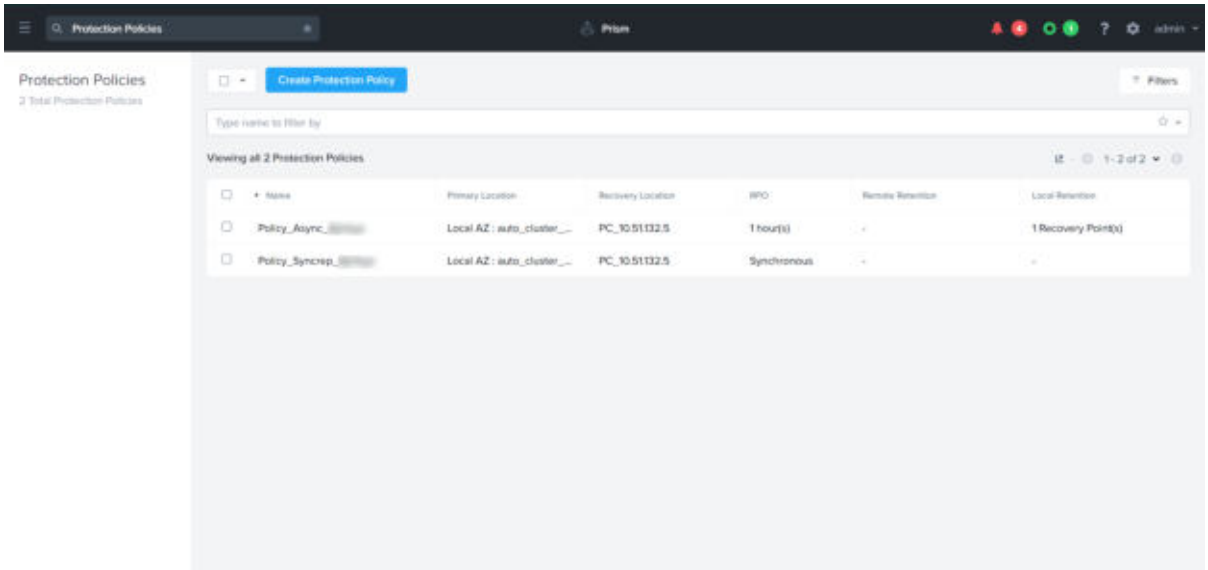


Figure 89: Protection Policies View

Table 32: Protection Policies View Fields

Field	Description
Name	Name of the protection policy.
Primary Location	Replication source site for the protection policy.
Recovery Location	Replication target site for the protection policy.
RPO	Recovery point objective for the protection policy.
Remote Retention	Number of retention points at the remote site.
Local Retention	Number of retention points at the local site.

Table 33: Workflows Available in the Protection Policies View

Workflow	Description
Create protection policy	Create a protection policy.

Table 34: Actions Available in the Actions Menu

Action	Description
Update	Update the protection policy.
Clone	Clone the protection policy.
Delete	Delete the protection policy.

Recovery Plans View in Xi Cloud Services

The **Recovery Plans** view lists all of configured recovery plans from all availability zones.

The following figure is a sample view, and the tables describe the fields and the actions that you can perform in this view.

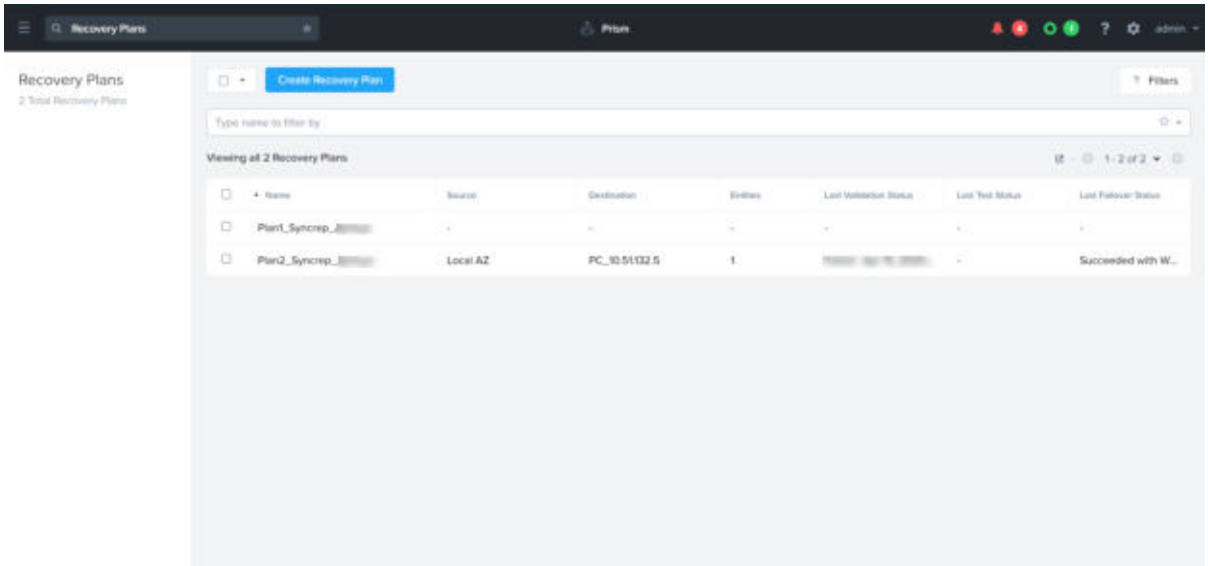


Figure 90: Recovery Plans View

Table 35: Recovery Plans View Fields

Field	Description
Name	Name of the recovery plan.
Source	Replication source site for the recovery plan.
Destination	Replication target site for the recovery plan.

Field	Description
Entities	Sum of the following VMs: • Number of local, live VMs that are specified in the recovery plan. • Number of remote VMs that the recovery plan can recover at this site.
Last Validation Status	Status of the most recent validation of the recovery plan.
Last Test Status	Status of the most recent test performed on the recovery plan.

Table 36: Workflows Available in the Recovery Plans View

Workflow	Description
Create Recovery Plan	Create a recovery plan.

Table 37: Actions Available in the Actions Menu

Action	Description
Validate	Validates the recovery plan to ensure that the VMs in the recovery plan have a valid configuration and can be recovered.
Test	Test the recovery plan.
Update	Update the recovery plan.
Failover	Perform a failover.
Delete	Delete the recovery plan.

Dashboard Widgets in Xi Cloud Services

The Xi Cloud Services dashboard includes widgets that display the statuses of configured protection policies and recovery plans. If you have not configured these VMs, the widgets display a summary of the steps required to get started with Leap.

To view these widgets, click the **Dashboard** tab.

The following figure is a sample view of the dashboard widgets.

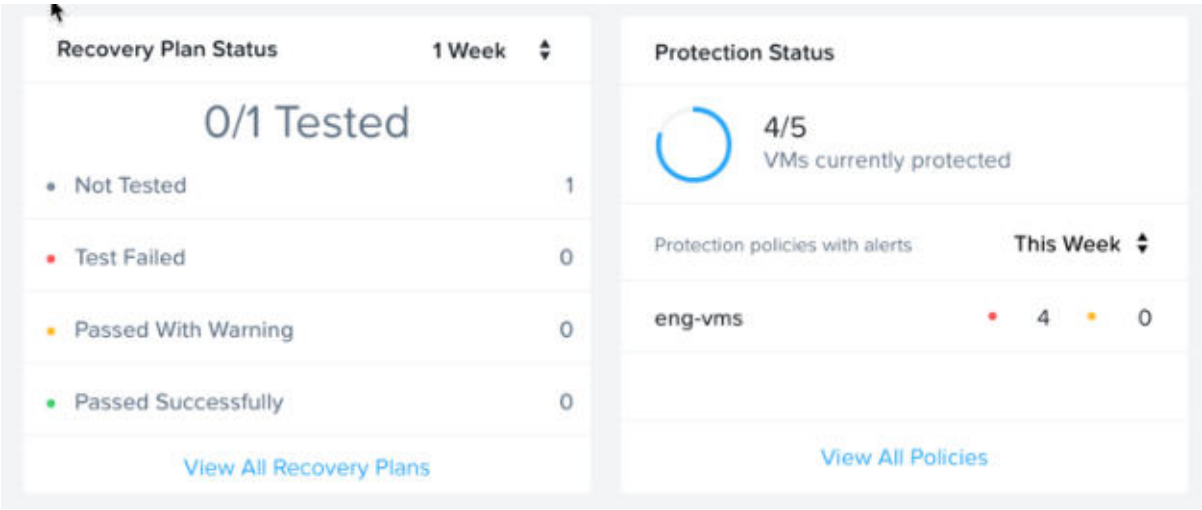


Figure 91: Dashboard Widgets for Xi Leap

Enabling Leap for On-Prem Site

To perform disaster recovery (DR) from Xi Cloud Services to a Nutanix cluster at an on-prem availability zone (site), enable Leap at the on-prem site (Prism Central) only. You need not enable Leap in the Xi Cloud Services portal; Xi Cloud Services does that by default for you. Without enabling Leap, you can configure protection policies and recovery plans that synchronize to the on-prem site but you cannot perform failover and failback operations.

To enable Leap at the on-prem site, see Enabling Leap for On-Prem Site on page 36.

Xi Leap Environment Setup

You can set up a secure environment to enable replication between an on-prem site and Xi Cloud Services with virtual private network (VPN). To configure the required environment, perform the following steps.

1. Pair your on-prem availability zone (site) with Xi Cloud Services. For more information about pairing, see Pairing Availability Zones (Xi Leap) on page 167.
2. Set up an on-prem VPN solution.
3. Enable VPN on the production virtual private cloud by using the Xi Cloud Services portal.
4. Set up a VPN client as a VM in Xi Cloud Services to enable connectivity to the applications that have failed over to the Xi Cloud Services.
5. Configure policy-based routing (PBR) rules for the VPN to successfully work with the Xi Cloud Services. If you have a firewall in the Xi Cloud Services and a floating IP address is assigned to the firewall, create a PBR policy in the Xi Cloud Services to configure the firewall as the gateway to the Internet. For example, specify 10.0.0.2/32 (private IP address of the firewall) in the **Subnet IP**. For more information, see *Policy Configuration* in *Xi Infrastructure Service Administration Guide* .
6. Configure the custom DNS in your virtual private cloud in the Xi Cloud Services. For more information, see *Virtual Private Cloud Management* in *Xi Infrastructure Service Administration Guide* .

> Note: For more information about Xi Cloud Services, see *Xi Infrastructure Service Administration Guide* .

Pairing Availability Zones (Xi Leap)

To perform disaster recovery (DR) from Xi Cloud Services to a Nutanix cluster at an on-prem availability zone (site), pair the on-prem site (Prism Central) only to Xi Cloud Services. For reverse synchronization, you need not pair again from Xi Cloud Services portal; Xi Cloud Services captures the paring configuration from the on-prem site that pairs Xi Cloud Services.

To pair an on-prem site with Xi Cloud Services, see Pairing Availability Zones (Leap) on page 37.

VPN Configuration (On-prem and Xi Cloud Services)

Xi Cloud Services enables you to set up a secure VPN connection between your on-prem sites and Xi Cloud Services to enable end-to-end disaster recovery services of Leap. A VPN solution between your on-prem site and Xi Cloud Services enables secure communication between your on-prem Prism Central instance and the production virtual private cloud (VPC) in Xi Cloud Services. If your workload fails over to Xi Cloud Services, the communication between the on-prem resources and failed over resources in Xi Cloud Services takes place over an IPSec tunnel established by the VPN solution.

> Note: Set up the VPN connection before data replication begins.

You can connect multiple on-prem sites to Xi Cloud Services. If you have multiple remote sites, you can set up secure VPN connectivity between each of your remote sites and Xi Cloud Services. With this configuration, you do not need to force the traffic from your remote site through your main site to Xi Cloud Services.

A VPN solution to connect to Xi Cloud Services includes a VPN gateway appliance in the Xi Cloud and a VPN gateway appliance (remote peer VPN appliance) in your on-prem site. A VPN gateway appliance learns about the local routes, establishes an IPSec tunnel with its remote peer, exchanges routes with its peer, and directs network traffic through the VPN tunnel.

After you complete the VPN configuration in the Xi Cloud Services portal, Nutanix creates a virtual VPN gateway appliance in the Xi Cloud. To set up a remote peer VPN gateway appliance in your on-prem site, you can either use the On Prem - Nutanix VPN solution (provided by Nutanix) or use a third-party VPN solution:

- **On Prem - Nutanix** (recommended): If you select this option, Nutanix creates a VPN gateway VM (remote peer appliance) on your on-prem cluster, connects the appliance to your network, and establishes an IPsec tunnel with the VPN gateway that is running in the Xi Cloud.

  The Nutanix VPN controller runs as a service in the Xi Cloud and on the on-prem Nutanix cluster and is responsible for the creation, setup, and lifecycle maintenance of the VPN gateway appliance (in the Xi Cloud and on-prem). The VPN controller deploys the virtual VPN gateway appliance in the Xi Cloud after you complete the VPN configuration in the Xi Cloud Services portal. The on-prem VPN controller deploys the virtual VPN gateway appliance on the on-prem cluster in the subnet you specify when you configure a VPN gateway in the Xi Cloud Services portal.

  The virtual VPN gateway appliance in the Xi Cloud and VPN gateway VM (peer appliance) in your on-prem cluster each consume 1 physical core, 4 GB RAM, and 10 GB storage.

- **On Prem - Third Party**: If you select this option, you must manually set up a VPN solution as an on-prem VPN gateway (peer appliance) that can establish an IPsec tunnel with the VPN gateway VM in the Xi Cloud. The on-prem VPN gateway (peer appliance) can be a virtual or hardware appliance. See On-Prem - Third-Party VPN Solution on page 176 for a list of supported third-party VPN solutions.

VPN Configuration Entities

To set up a secure VPN connection between your on-prem sites and Xi Cloud Services, configure the following entities in the Xi Cloud Services portal:

- **VPN Gateway**: Represents the gateway of your VPN appliances.

  VPN gateways are of the following types:

  - **Xi Gateway**: Represents the Xi VPN gateway appliance
  - **On Prem - Nutanix Gateway**: Represents the VPN gateway appliance at your on-prem site if you are using the on-prem Nutanix VPN solution.
  - **On Prem - Third Party Gateway**: Represents the VPN gateway appliance at your on-prem site if you are using your own VPN solution (provided by a third-party vendor).
- **VPN Connection**: Represents the VPN IPSec tunnel established between a VPN gateway in the Xi Cloud and VPN gateway in your on-prem site. When you create a VPN connection, you select a Xi gateway and on-prem gateway between which you want to create the VPN connection.

You configure a VPN gateway in the Xi Cloud and at each of the on-prem sites you want to connect to the Xi Cloud. You then configure a VPN connection between a VPN gateway in the Xi Cloud and VPN gateway in your on-prem site.

Single-Site Connection

If you want to connect only one on-prem site to the Xi Cloud, configure the following entities in the Xi Cloud Services portal:

1. One Xi gateway to represent the Xi VPN gateway appliance
2. One on-prem gateway (On-prem - Nutanix Gateway or on-prem - third-party Gateway) to represent the VPN gateway appliance at your on-prem site
3. One VPN connection to connect the two VPN gateways

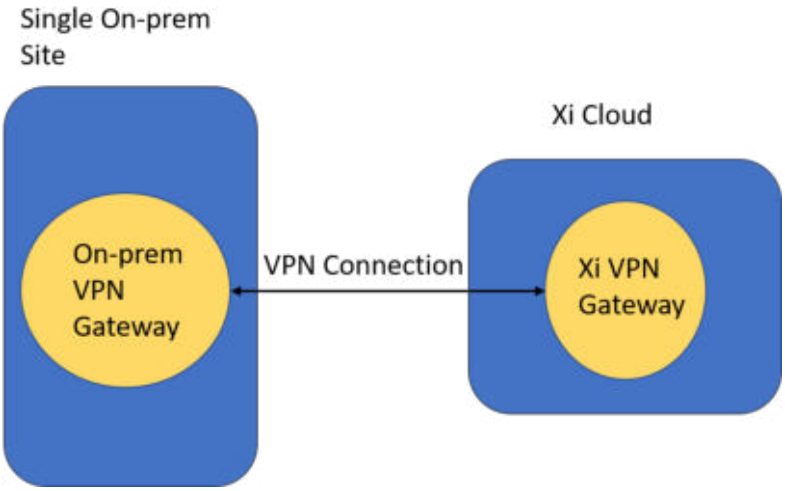


Figure 92: Single-Site Connection

Multi-Site Connection

If you want to connect multiple on-prem sites to the Xi Cloud, configure the following entities in the Xi Cloud Services portal:

1. One Xi gateway to represent the Xi VPN gateway appliance
2. On-prem gateways (On-prem - Nutanix Gateway or on-prem - third-party Gateway) for each on-prem site
3. VPN connections to connect the Xi gateway and the on-prem gateway at each on-prem site

For example, if you want to connect two on-prem sites to the Xi Cloud, configure the following:

1. One Xi gateway
2. Two on-prem gateways for the two on-prem sites
3. Two VPN connections

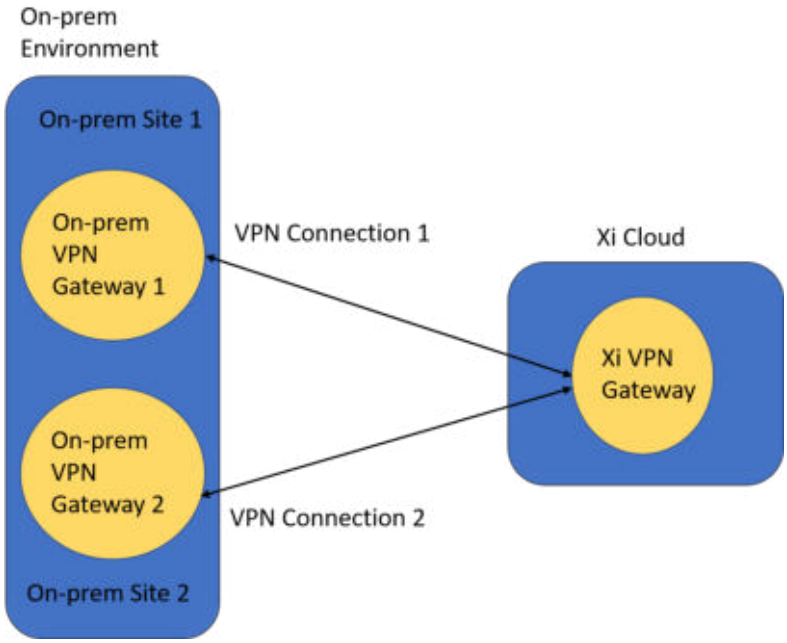


Figure 93: Multi-Site Connection for less the 1 Gbps Bandwidth

One Xi VPN gateway provides 1 Gbps of aggregate bandwidth for IPSec traffic. Therefore, connect only as many on-prem VPN gateways to one Xi VPN gateway to accommodate 1 Gbps of aggregate bandwidth.

If you require an aggregate bandwidth of more than 1 Gbps, configure multiple Xi VPN gateways.

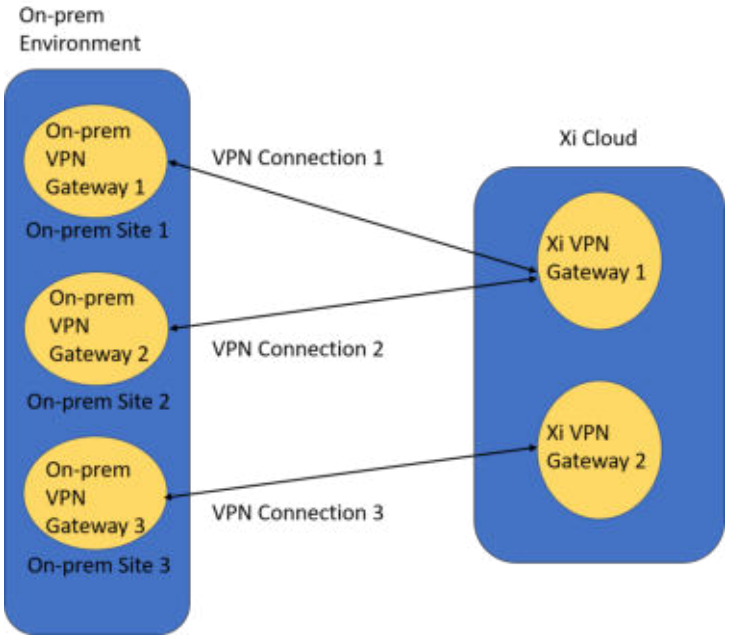


Figure 94: Multi-Site Connection for more the 1 Gbps Bandwidth

On-Prem - Nutanix VPN Solution

You can use the on-prem - Nutanix VPN solution to set up VPN between your on-prem site and Xi Cloud Services. If you select this option, you are using an end-to-end VPN solution provided by Nutanix and you do not need to use your own VPN solution to connect to Xi Cloud Services.

After you complete the VPN configuration in the Xi Cloud Services portal, Nutanix creates a virtual VPN gateway appliance in the Xi Cloud. The On Prem - Nutanix VPN solution creates a VPN gateway VM (remote peer appliance) on your on-prem cluster, connects the appliance to your network, and establishes an IPsec tunnel with the VPN gateway VM that is running in the Xi Cloud.

Following is the workflow if you choose the On Prem - Nutanix VPN solution to set up a VPN connection between your on-prem site and Xi Cloud Services.

1. Create one or more Xi VPN gateways.
2. The VPN controller running in Xi Cloud Services creates a VPN gateway VM in the Xi Cloud. The Xi VPN gateway VM runs in your (tenant) overlay network.
3. Create one or more on-prem VPN gateways.

   Create a VPN gateway for each on-prem site that you want to connect to the Xi Cloud.
4. Create one or more VPN connections.

   Create a VPN connection between each on-prem site (on-prem VPN gateway) and Xi Cloud (Xi gateway).
5. The VPN controller creates a VPN gateway VM on the on-prem cluster in the subnet you specify when you create an on-prem VPN gateway. The VPN gateway VM becomes the peer appliance to the VPN gateway VM in the Xi Cloud.

6. Both the VPN appliances are now configured, and the appliances now proceed to perform the following:
   1. An on-prem router communicates the on-prem routes to the on-prem VPN gateway by using iBGP or OSPF.
   2. The Xi VPN controller communicates the Xi subnets to the Xi VPN gateway VM.
   3. The on-prem VPN gateway VM then establishes a VPN IPsec tunnel with the Xi VPN gateway VM. Both appliances establish an eBGP peering session over the IPsec tunnel and exchange routes.
   4. The on-prem VPN gateway VM publishes the Xi subnet routes to the on-prem router by using iBGP or OSPF.

Nutanix VPN Solution Requirements

In your on-prem site, ensure the following before you configure VPN on Xi Cloud Services:

1. The Prism Central instance and cluster are running AOS 5.11 or newer for AHV and AOS 5.19 or newer for ESXi.
2. A router with iBGP, OSPF, or Static support to communicate the on-prem routes to the on-prem VPN gateway VM.
3. Depending on whether you are using iBGP or OSPF, ensure that you have one of the following:
   - **Peer IP** (for iBGP): The IP address of the on-prem router to exchange routes with the VPN gateway VM.
   - **Area ID** (for OSPF): The OSPF area ID for the VPN gateway in the IP address format.
4. Determine the following details for the deployment of the on-prem VPN gateway VM.
   - **Subnet UUID**: The UUID of the subnet of the on-prem cluster in which you want to install the on-prem VPN gateway VM. Log on to your on-prem Prism Central web console to determine the UUID of the subnet.
   - **Public IP address of the VPN Gateway Device**: A public WAN IP address that you want the on-prem gateway to use to communicate with the Xi VPN gateway appliance.
   - **VPN VM IP Address**: A static IP address that you want to allocate to the on-prem VPN gateway VM.
   - **IP Prefix Length**: The subnet mask in CIDR format of the subnet on which you want to install the on-prem VPN gateway VM.
   - **Default Gateway IP**: The gateway IP address for the on-prem VPN gateway appliance.
   - **On Prem Gateway ASN**: ASN must not be the same as any of your on-prem BGP ASNs. If you already have a BGP environment in your on-prem site, the customer gateway is the ASN for your organization. If you do not have a BGP environment in your on-prem site, you can choose any number. For example, you can choose a number in the 65000 range.

Firewall Port Requirements for On-Prem Site

Configure rules for ports in your on-prem firewall depending on your deployment scenario.

On-Prem Behind a Network Address Translation or Firewall Device

In this scenario, the IPSec tunnel terminates behind a network address translation (NAT) or firewall device. For NAT to work, open UDP ports 500 and 4500 in both directions. Ports 1024–1034 are ephemeral ports used by the CVMs.

Enable the on-prem VPN gateway to allow the traffic according to the rules described in the *Port Rules* table.

IPSec Terminates on the Firewall Device

In this scenario, you do not need to open the ports for NAT (500 and 4500), but enable the on-prem VPN gateway to allow the traffic according to the rules described in the *Port Rules* table.

In the following table, PC subnet refers to the subnet where your on-prem Prism Central is running. The Xi infrastructure load balancer route is where the traffic for Xi CVMs and PC is located. You receive this information when you begin using Xi Cloud Services.

Table 38: Port Rules

Source address	Destination address	Source port	Destination port
PC subnet	Load balancer route advertised	Any	1024–1034
Xi infrastructure load balancer route	PC and CVM subnet	Any	2020 2009 9440
The following port requirements are applicable only if you are using the Nutanix VPN solution.			
Nutanix VPN VM	8.8.8.8 and 8.8.4.4 IP addresses of the DNS server	VPN VM	DNS UDP port 53
Nutanix VPN VM	time.google.com, 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org of the NTP server	VPN VM	NTP UDP port 123
Nutanix VPN VM	ICMP ping to NTP servers	NA	NA
CVM IP address in AHV clusters	HTTPS request to the Internet	AHV hosts	HTTPS port 443
CVM IP address in ESXi clusters	HTTPS and FTP requests to the Internet	ESXi hosts	HTTPS port 443 and FTP 21

Creating a Xi VPN Gateway

Create a VPN gateway to represent the Xi VPN gateway appliance.

About this task

Perform the following to create a Xi VPN gateway.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.

2. Click **Create VPN Gateway**.

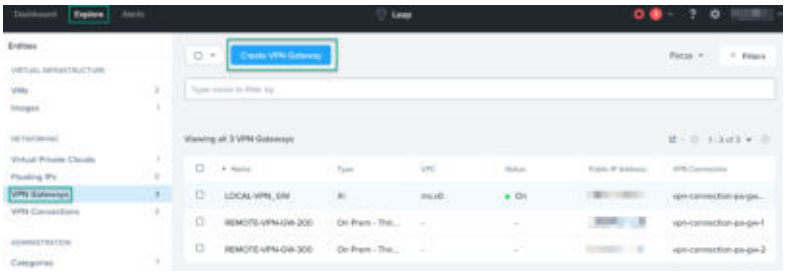


Figure 95: Create a Xi Gateway

The **Create VPN Gateway** window appears.

3. Do the following in the indicated fields.

   a. Name: Enter a name for the VPN gateway.

   b. VPC: Select the production VPC.

   c. Type: Select Xi Gateway.

   d. **Routing Protocol**: Select **eBGP** or **Static** to set up a routing protocol between the Xi and on-prem gateways.

   e. (**eBGP** only): Select this option if you want to set up the eBGP routing protocol between the Xi and on-prem gateways. Do the following in the indicated fields.

      - In the **ASN** field, set an ASN for the Xi gateway. Ensure that the Xi gateway ASN is different from that on-prem gateway ASN.
      - In the **eBGP Password** field, set up a password for the eBGP session that is established between the on-prem VPN gateway and Xi VPN gateway. The eBGP password can be any string, preferably alphanumeric.

   f. (**Static** only) If you select this option, manually set up static routes between the Xi and on-prem gateways.

      For more information, see *Adding a Static Route* in *Xi Infrastructure Service Administration Guide*.

4. Click **Save**.
   The Xi gateway you create is displayed in the **VPN Gateways** page.

Creating an On-Prem VPN Gateway (Nutanix)

Create a VPN gateway to represent the on-prem VPN gateway appliance.

About this task

Perform the following to create an on-prem VPN gateway.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.

2. Click **Create VPN Gateway**.

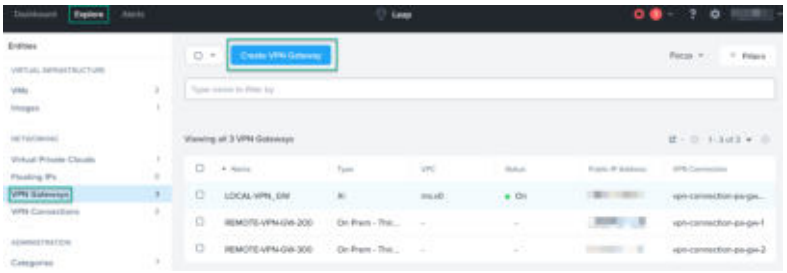


Figure 96: Create a on-prem Gateway

The **Create VPN Gateway** window appears.

3. Do the following in the indicated fields.

   a. **Name**: Enter a name for the VPN gateway.

   b. **Type**: Select **On Prem - Nutanix**.

   c. **Automatically add route in PC and PE CVMs to enable replication**: Select this option to automatically enable traffic between the on-prem CVMs and CVMs in Xi Cloud Services. If you select this option, a route to the CVMs in Xi Cloud Services is added with the on-prem VPN gateway as the next-hop. Therefore, even if you choose to have static routes between your on-prem router and the on-prem gateway, you do not need to manually add those static routes (see step g).

      A route to Xi CVMs is added with the on-prem VPN gateway as the next-hop.

      Note: This option is only for the CVM-to-CVM (on-prem CVM and Xi Cloud CVMs) traffic.

   d. Under **Routing Protocol (between Xi Gateway and On Prem Nutanix Gateway)**, do the following to set up the eBGP routing protocol between the Xi and on-prem gateways:

      - In the **ASN** field, enter the ASN for your on-prem gateway. If you do not have a BGP environment in your on-prem site, you can choose any number. For example, you can choose a number in the 65000 range. Ensure that the Xi gateway ASN and on-prem gateway ASN are not the same.

- In the **eBGP Password** field, enter the same eBGP password as the Xi gateway.

e. **Subnet UUID**: Enter the UUID of the subnet of the on-prem cluster in which you want to install the on-prem VPN gateway VM. Log on to your on-prem Prism Central web console to determine the UUID of the subnet.

f. Under **IP Address Information**, do the following in the indicated fields:

- **Public IP Address of the On Premises VPN Gateway Device**: Enter a public WAN IP address for the on-prem VPN gateway VM.
- **VPN VM IP Address**: Enter a static IP address that you want to allocate to the on-prem VPN gateway VM.
- **IP Prefix Length**: Enter the subnet mask of mask length 24 of the subnet on which you want to install the on-prem VPN gateway VM.
- **Default Gateway IP**: Enter the gateway IP address of the subnet on which you want to install the on-prem VPN gateway VM.

g. Under **Routing Protocol Configuration**, do the following in the indicated fields:

- In the **Routing Protocol** drop-down list, select the dynamic routing protocol (**OSPF**, **iBGP**, or **Static**) to set up the routing protocol between the on-prem router and on-prem gateway.
- (**Static** only) If you select **Static**, manually add these routes in Xi Cloud Services.

  For more information, see *Adding a Static Route* in *Xi Infrastructure Service Administration Guide*.

  > Note: You do not need to add static routes for CVM-to-CVM traffic (see step c).

- (**OSPF** only) If you select **OSPF**, in the **Area ID** field, type the OSPF area ID for the VPN gateway in the IP address format. In the **Password Type** field, select **MD5** and type a password for the OSPF session.
- (**iBGP** only) If you select **iBGP**, in the **Peer IP** field, type the IP address of the on-prem router to exchange routes with the VPN gateway VM. In the **Password** field, type the password for the iBGP session.

4. Click **Save**.
   The on-prem gateway you create is displayed in the **VPN Gateways** page.

Creating a VPN Connection

Create a VPN connection to establish a VPN IPSec tunnel between a VPN gateway in the Xi Cloud and VPN gateway in your on-prem site. Select the Xi gateway and on-prem gateway between whom you want to create the VPN connection.

About this task

Perform the following to create a VPN connection.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Connections**.

2. Click **Create VPN Connection**.

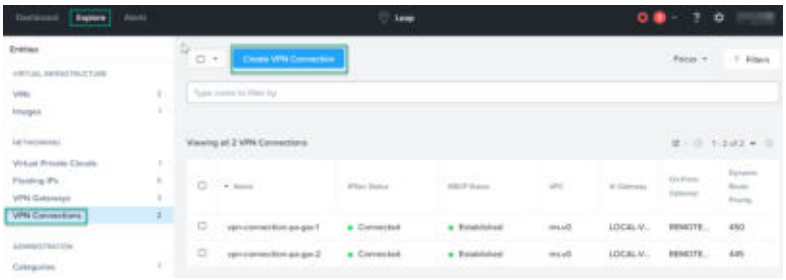


Figure 97: Create a VPN Connection

The **Create VPN Connection** window appears.

3. Do the following in the indicated fields:

   a. **Name**: Enter a name for the VPN connection.

   b. **Description**: Enter a description for the VPN connection.

   c. **IPSec Secret**. Enter an alphanumeric string as the IPSec string for the VPN connection.

   d. **Xi Gateway**: Select the Xi gateway for which you want to create this VPN connection.

   e. **On Premises Gateway**: Select the on-prem gateway for which you want to create this VPN connection.

   f. **Dynamic Route Priority**: This is not a mandatory field. Set this field if you have multiple routes to the same destination. For example, consider you have VPN connection 1 and VPN connection 2 and you want VPN connection 1 to take precedence over VPN connection 2, set the priority for VPN connection 1 higher than VPN connection 2. Higher the priority number, higher is the precedence of that connection. You can set a priority number from 10 through 1000.

      See the **Routes Precedence** section in *Routes Management* in *Xi Infrastructure Service Administration Guide* for more information.

4. Click **Save**.

   The VPN connection you create is displayed in the **VPN Connections** page.

On-Prem - Third-Party VPN Solution

You can use your own VPN solution to connect your on-prem site to Xi Cloud Services. If you select this option, you must manually set up a VPN solution by using a supported third-party VPN solution as an on-prem VPN gateway (peer appliance) that can establish an IPsec tunnel with the VPN gateway VM in the Xi Cloud.

Following is the workflow if you want to use a third-party VPN solution to set up a VPN connection between your on-prem site and Xi Cloud Services.

1. Create one or more Xi VPN gateways.
2. The VPN controller running in Xi Cloud Services creates a VPN gateway VM in the Xi Cloud. The Xi VPN gateway VM runs in your (tenant) overlay network.
3. Create one or more on-prem VPN gateways.

   Create a VPN gateway for each on-prem site that you want to connect to the Xi Cloud.

4. Create one or more VPN connections.

   Create a VPN connection to create an IPSec tunnel between each on-prem site (on-prem VPN gateway) and Xi Cloud (Xi gateway).
5. Configure a peer VPN gateway appliance (hardware or virtual) in your on-prem site. Depending upon your VPN solution, you can download detailed instructions about how to configure your on-prem VPN gateway appliance. For more information, see Downloading the On-Prem VPN Appliance Configuration on page 182.

   Xi Cloud Services supports the following third-party VPN gateway solutions.

   - CheckPoint
   - Cisco ASA
   - PaloAlto

     > Note: If you are using the Palo Alto VPN gateway solution, set the MTU value to 1356 in the Tunnel Interface settings. The replication fails for the default MTU value (1427).

   - Juniper SRX
   - Fortinet
   - SonicWall
   - VyOS

Third-Party VPN Solution Requirements

Ensure the following in your on-prem site before you configure VPN in Xi Cloud Services.

1. A third-party VPN solution in your on-prem site that functions as an on-prem VPN gateway (peer appliance).
2. The on-prem VPN gateway appliance supports the following.

   - IPSec IKEv2
   - Tunnel interfaces
   - External Border Gateway Protocol (eBGP)
3. Note the following details of the on-prem VPN gateway appliance.

   - **On Prem Gateway ASN**: Assign an ASN for your on-prem gateway. If you already have a BGP environment in your on-prem site, the customer gateway is the ASN for your organization. If you do not have a BGP environment in your on-prem site, you can choose any number. For example, you can choose a number in the 65000 range.
   - **Xi Gateway ASN**: Assign an ASN for the Xi gateway. The Xi gateway ASN must not be the same as the on-prem gateway ASN.
   - **eBGP Password**: The eBGP password is the shared password between the Xi gateway and on-prem gateway. Set the same password for both the gateways.
   - **Public IP address of the VPN Gateway Device**: Ensure that the public IP address of the on-prem VPN gateway appliance can reach the public IP address of Xi Cloud Services.
4. The on-prem VPN gateway appliance can route the traffic from the on-prem CVM subnets to the established VPN tunnel.

5. Ensure that the following ports are open in your on-prem VPN gateway appliance.
   - IKEv2: Port number 500 of the payload type UDP.
   - IPSec: Port number 4500 of the payload type UDP.
   - BGP: Port number 179 of the payload type TCP.

Firewall Port Requirements for On-Prem Site

Configure rules for ports in your on-prem firewall depending on your deployment scenario.

On-Prem Behind a Network Address Translation or Firewall Device

In this scenario, the IPSec tunnel terminates behind a network address translation (NAT) or firewall device. For NAT to work, open UDP ports 500 and 4500 in both directions. Ports 1024–1034 are ephemeral ports used by the CVMs.

Enable the on-prem VPN gateway to allow the traffic according to the rules described in the *Port Rules* table.

IPSec Terminates on the Firewall Device

In this scenario, you do not need to open the ports for NAT (500 and 4500), but enable the on-prem VPN gateway to allow the traffic according to the rules described in the *Port Rules* table.

In the following table, PC subnet refers to the subnet where your on-prem Prism Central is running. The Xi infrastructure load balancer route is where the traffic for Xi CVMs and PC is located. You receive this information when you begin using Xi Cloud Services.

Table 39: Port Rules

Source address	Destination address	Source port	Destination port
PC subnet	Load balancer route advertised	Any	1024–1034
Xi infrastructure load balancer route	PC and CVM subnet	Any	2020 2009 9440
The following port requirements are applicable only if you are using the Nutanix VPN solution.			
Nutanix VPN VM	8.8.8.8 and 8.8.4.4 IP addresses of the DNS server	VPN VM	DNS UDP port 53
Nutanix VPN VM	time.google.com, 0.pool.ntp.org, 1.pool.ntp.org, 2.pool.ntp.org of the NTP server	VPN VM	NTP UDP port 123
Nutanix VPN VM	ICMP ping to NTP servers	NA	NA

Source address	Destination address	Source port	Destination port
CVM IP address in AHV clusters	HTTPS request to the Internet	AHV hosts	HTTPS port 443
CVM IP address in ESXi clusters	HTTPS and FTP requests to the Internet	ESXi hosts	HTTPS port 443 and FTP 21

Creating a Xi VPN Gateway

Create a VPN gateway to represent the Xi VPN gateway appliance.

About this task

Perform the following to create a Xi VPN gateway.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.
2. Click **Create VPN Gateway**.

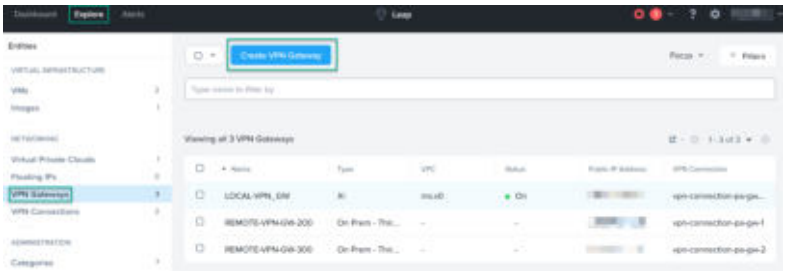


Figure 98: Create a Xi Gateway

The **Create VPN Gateway** window appears.

3. Do the following in the indicated fields.

   a. Name: Enter a name for the VPN gateway.

   b. VPC: Select the production VPC.

   c. Type: Select Xi Gateway.

   d. **Routing Protocol**: Select **eBGP** or **Static** to set up a routing protocol between the Xi and on-prem gateways.

   e. (**eBGP** only): Select this option if you want to set up the eBGP routing protocol between the Xi and on-prem gateways. Do the following in the indicated fields.

      - In the **ASN** field, set an ASN for the Xi gateway. Ensure that the Xi gateway ASN is different from that on-prem gateway ASN.

      - In the **eBGP Password** field, set up a password for the eBGP session that is established between the on-prem VPN gateway and Xi VPN gateway. The eBGP password can be any string, preferably alphanumeric.

   f. (**Static** only) If you select this option, manually set up static routes between the Xi and on-prem gateways.

      For more information, see *Adding a Static Route* in *Xi Infrastructure Service Administration Guide*.

4. Click **Save**.

   The Xi gateway you create is displayed in the **VPN Gateways** page.

Creating an On-Prem VPN Gateway (Third-Party)

Create a VPN gateway to represent the on-prem VPN gateway appliance.

Before you begin

Ensure that you have all the details about your on-prem VPN appliance as described in Third-Party VPN Solution Requirements on page 177.

About this task

Perform the following to create an on-prem VPN gateway.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.

2. Click **Create VPN Gateway**.

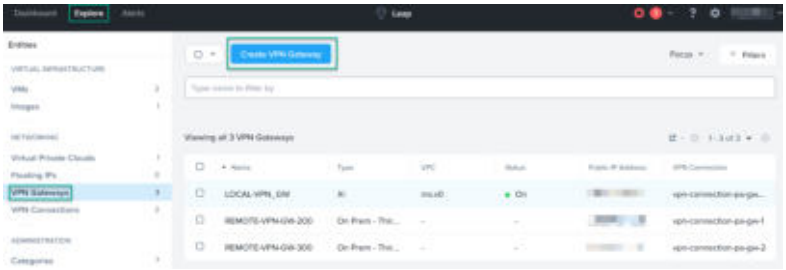


Figure 99: Create an On Prem Gateway

3. Do the following in the indicated fields.

   a. **Name**: Enter a name for the VPN gateway.

   b. **Type**: Select **On Prem - Third Party**.

   c. **IP Address of your Firewall or Router Device performing VPN**: Enter the IP address of the on-prem VPN appliance.

   d. **Routing Protocol**: Select eBGP or Static to set up a routing protocol between the Xi and on-prem gateways.

   e. (**eBGP** only) If you select eBGP, do the following:

      - In the **ASN** field, enter the ASN for your on-prem gateway. If you do not have a BGP environment in your on-prem site, you can choose any number. For example, you can choose a number in the 65000 range. Ensure that the Xi gateway ASN and on-prem gateway ASN are not the same.

      - In the **eBGP Password** field, enter the same eBGP password as the Xi gateway.

   f. (**Static** only) If you select **Static**, manually set up static routes between the Xi and on-prem gateways.

      For more information, see *Adding a Static Route* in *Xi Infrastructure Service Administration Guide*.

Creating a VPN Connection

Create a VPN connection to establish a VPN IPSec tunnel between a VPN gateway in the Xi Cloud and VPN gateway in your on-prem site. Select the Xi gateway and on-prem gateway between whom you want to create the VPN connection.

About this task

Perform the following to create a VPN connection.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Connections**.

2. Click **Create VPN Connection**.

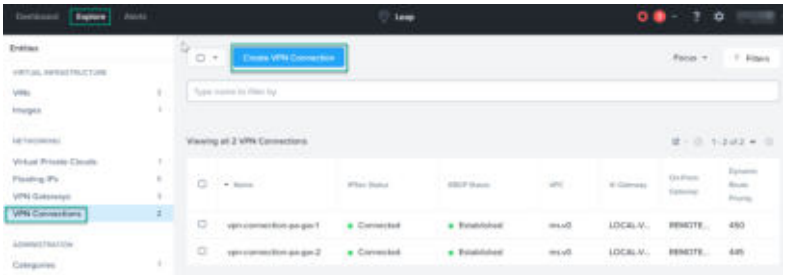


Figure 100: Create a VPN Connection

The **Create VPN Connection** window appears.

3. Do the following in the indicated fields:
   a. **Name**: Enter a name for the VPN connection.
   b. **Description**: Enter a description for the VPN connection.
   c. **IPSec Secret**. Enter an alphanumeric string as the IPSec string for the VPN connection.
   d. **Xi Gateway**: Select the Xi gateway for which you want to create this VPN connection.
   e. **On Premises Gateway**: Select the on-prem gateway for which you want to create this VPN connection.
   f. **Dynamic Route Priority**: This is not a mandatory field. Set this field if you have multiple routes to the same destination. For example, consider you have VPN connection 1 and VPN connection 2 and you want VPN connection 1 to take precedence over VPN connection 2, set the priority for VPN connection 1 higher than VPN connection 2. Higher the priority number, higher is the precedence of that connection. You can set a priority number from 10 through 1000.

      See the **Routes Precedence** section in *Routes Management* in *Xi Infrastructure Service Administration Guide* for more information.
4. Click **Save**.

   The VPN connection you create is displayed in the **VPN Connections** page.

Downloading the On-Prem VPN Appliance Configuration

Depending upon your VPN solution, you can download detailed instructions about how to configure your on-prem VPN gateway appliance.

About this task

Perform the following to download the instructions to configure your on-prem VPN gateway appliance.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.
2. Click an on-prem VPN gateway.
3. In the details page, click **On Prem Gateway Configuration**.

Figure 101: On-prem VPN Gateway Appliance Configuration

4. Select the type and version of your on-prem VPN gateway appliance and click **Download**.
5. Follow the instructions in the downloaded file to configure the on-prem VPN gateway appliance.

VPN Gateway Management

You can see the details of each VPN gateway, update the gateway, or delete the gateway.

All your VPN gateways are displayed in the VPN Gateways page.

Displaying the Details of a VPN Gateway

You can display the details such as the type of gateway, VPC, IP addresses, protocols, and connections associated with the gateways.

About this task

Perform the following to display the details of a VPN gateway.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.
   A list of all your VPN gateways is displayed. The VPN gateways table displays details such as the name, type, VPC, status, public IP address, and VPN connections associated with each VPN gateway.

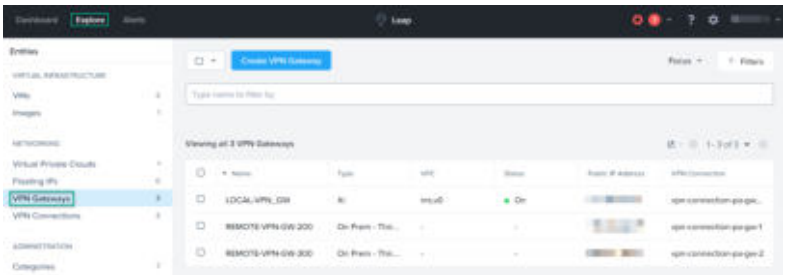


Figure 102: VPN Gateways List

2. Click the name of a VPN gateway to display additional details of that VPN gateway.
3. In the details page, click the name of a VPN connection to display the details of that VPN connection associated with the gateway.

Updating a VPN Gateway

The details that you can update in a VPN gateway depend on the type of gateway (Xi gateway or On Prem gateway).

About this task

Perform the following to update a VPN gateway.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.

2. Do one of the following:

   - Select the checkbox next to the name of the VPN gateway and, in the **Actions** drop-down list, click **Update**.

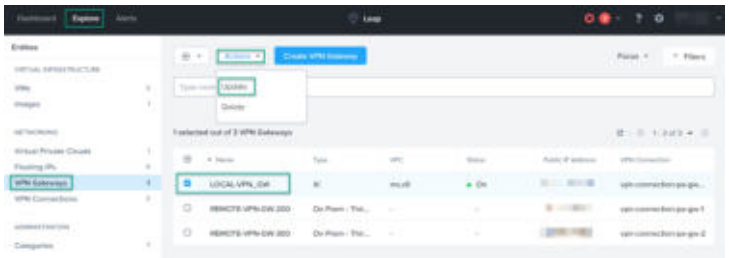


Figure 103: Use the Actions drop-down list

   - Click the name of the VPN gateway and, in the details page that appears, click **Update**.

   The **Update VPN Gateway** dialog box appears.

3. Update the details as required.

   The fields are similar to the **Create VPN Gateway** dialog box. For more information, see Creating a Xi VPN Gateway on page 172, Creating an On-Prem VPN Gateway (Nutanix) on page 173, or Creating an On-Prem VPN Gateway (Third-Party) on page 180 depending on the type of gateway you are updating.

4. Click **Save**.

Deleting a VPN Gateway

If you want to delete a VPN gateway, you must first delete all the VPN connections associated with the gateway and only then you can delete the VPN gateway.

About this task

Perform the following to delete a VPN gateway.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.

2. Do one of the following:

   » Select the checkbox next to the name of the VPN gateway and, in the **Actions** drop-down list, click **Delete**.

   » Click the name of the VPN gateway and, in the details page that appears, click **Delete**.

3. Click **OK** in the confirmation message that appears to delete the VPN gateway.

VPN Gateway Upgrades

Nutanix deployment can detect and install upgrades for the on-prem Nutanix Gateways and the Xi VPN Gateways in the XI Cloud Services based disaster recovery (DR) solutions (Xi Leap configurations).

For onprem Nutanix VPN Gateways (deployments that do not use Xi Cloud Services), the upgrades need to be detected and installed on the respective PC on which each Nutanix VPN Gateway is installed.

For Xi Cloud Services based Xi VPN Gateways, the upgrades are automatically detected on the Xi PC. If not installed within a certain amount of time, the Xi PC automatically installs the upgrades to protect Xi Cloud Services. For onprem Nutanix Gateways connected to Xi PC, the onprem PC detects the upgrades but the Xi PC installs the upgrades.

For more information, see Detecting Upgrades for VPN Gateways on page 187.

When PC (onprem PC or Xi PC) detects the upgrades, it displays a banner on the VPN Gateways tab of the VPN page. The banner notifies you that a VPN upgrade is available after you have run LCM inventory. The table on the VPN Gateways tab also displays an alert (exclamation mark) icon for the VPN gateways that the upgrade applies to. The hover message for the icon informs you that an upgrade is available for that VPN Gateway.

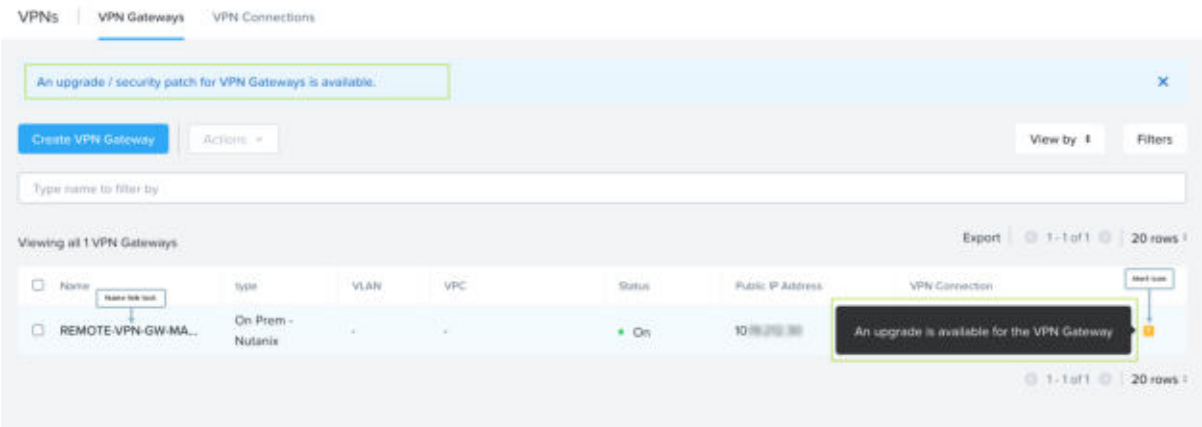


Figure 104: Upgrade Banner

For more information, see Upgrading the Xi-based VPN Gateways on page 187 and Upgrading the PC-managed Onprem Nutanix VPN Gateways on page 188.

For information about identifying the VPN Gateway, see Identifying the VPN Gateway Version on page 185.

Identifying the VPN Gateway Version

About this task

To identify the current VPN Gateway version, do the following:

Procedure

- On the VPN Gateway tab, click the VPN Gateway name link text to open the VPN Gateway details page.

  In the VPN Gateway table, the VPN Gateway name is a clickable link text.

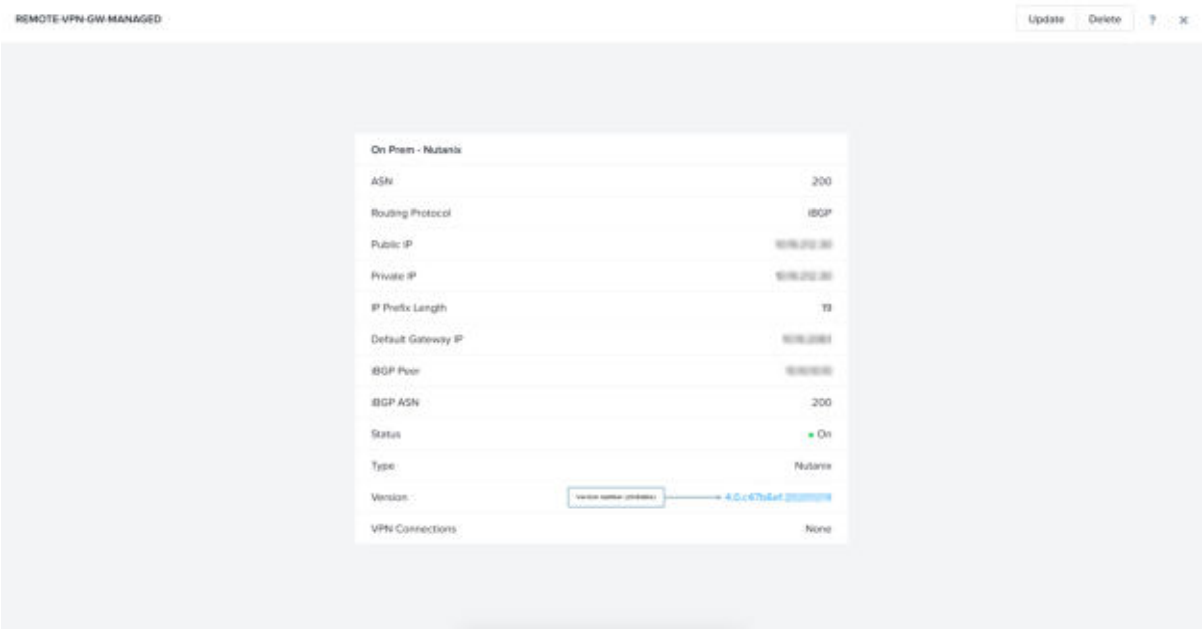


Figure 105: VPN Gateway Details

- Click the version number link text to open the VPN Version dialog box.

  On the VPN Gateway details page, the version identifier or number is clickable.

  The VPN Version dialog box provides information about the current version. It informs you if the current version is the latest. If the current version is the latest, the **Update** button in the dialog box is unavailable (not clickable). If an upgrade is available, the **Update** button is available (clickable).

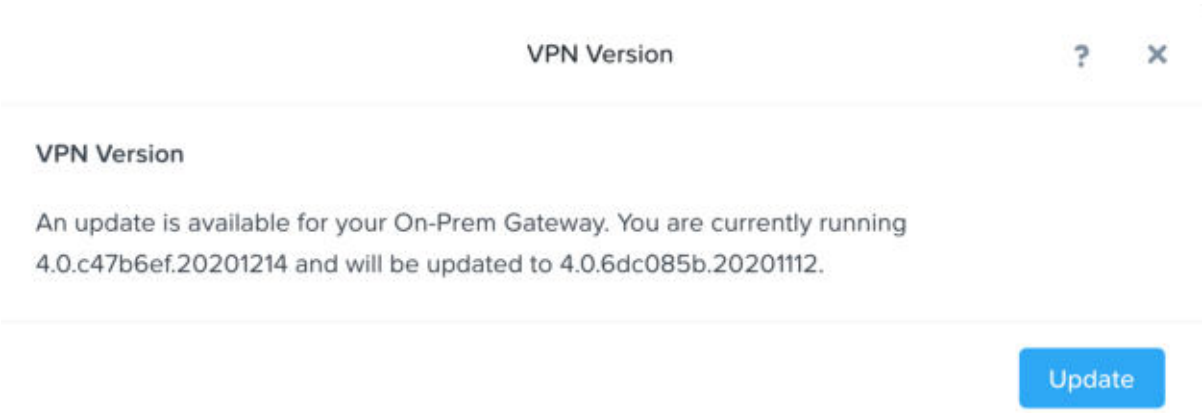


Figure 106: VPN Version Dialog Box

Detecting Upgrades for VPN Gateways

About this task

Prism Central can detect whether new VPN Gateway upgrades are available, or not, for Nutanix VPN Gateways. You can then install the upgrade.

> Note:
>
> Xi PC detects Xi VPN Gateway upgrades automatically. You can directly upgrade the Xi VPN Gateway when the notification banner appears on the VPN Gateway page.
>
> If you do not upgrade the Xi VPN Gateway version when a new version is available, the Xi PC automatically installs the upgrades to protect Xi Cloud Services.

Procedure

- Click **Perform Inventory** on on-prem PC LCM page, for the Xi-managed on-prem Nutanix VPN Gateways.

  In the on-prem deployment of Leap where the primary sites are on-prem and recovery site is Xi Cloud Services, you need to **Perform Inventory** in LCM on the on-prem Prism Central to detect new versions or upgrades of the Onprem Nutanix VPN Gateway.

  > Note:
  >
  > Nutanix recommends that you select **Enable LCM Auto Inventory** in the LCM page in on-prem Prism Central to continuously detect new VPN upgrades as soon as they are available.

  The upgrade notification banner is displayed on the VPN Gateways page.

  The Xi PC user interface displays a banner on the VPN Gateways page notifying you that a VPN upgrade is available after you have run LCM inventory. The table on the VPN Gateways page also displays an icon for the VPN gateways that the upgrade applies to.

  The VPN Gateways page in Xi displays the same banner and icon (as in case of Xi PC) as soon as the upgrades are available in LCM inventory.

Upgrading the Xi-based VPN Gateways

About this task

Each VPN gateway is upgraded independently of one another on both sides of the VPN connection. In a Xi Leap configuration, you must perform the VPN Gateway upgrades on both sides of the VPN connection using the Xi user interface.

Perform upgrades of Xi-managed Onprem Nutanix VPN Gateways from the Xi PC.

> Note:
>
> You cannot upgrade Xi-managed Nutanix VPN Gateways from the Onprem PC.
>
> If you do not upgrade the Xi VPN Gateway version when a new version is available, the Xi PC automatically installs the upgrades to protect Xi Cloud Services.

To upgrade the Xi VPN Gateways and the Xi-managed on-prem Nutanix VPN Gateways, do the following:

Procedure

1. Do one of the following to open the VPN Version dialog box:

   » On the VPN Gateway details page, the version identifier or number is clickable.

   Click the version number link text.

   » Click the alert icon for the VPN Gateway in the VPN Gateway table.

   The VPN Version dialog box opens.

   See Identifying the VPN Gateway Version on page 185 for a sample view of the VPN Version dialog box.

2. Click the **Update** button to upgrade the VPN Gateway version.

Upgrading the PC-managed Onprem Nutanix VPN Gateways

About this task

Perform upgrades of PC-managed Nutanix Gateways using the respective PC on which the Gateway is created.

To upgrade the on-prem Nutanix Gateways, do the following:

Procedure

1. Log on to the Prism Central as the admin user and click the gear icon.
2. Go to Administration > LCM > Inventory.
3. Click **Perform Inventory**.

   When you click **Perform Inventory**, the system scans the registered Prism Central cluster for software versions that are running currently. Then it checks for any available upgrades and displays the information on the LCM page under Software.

   > Note:
   >
   > Skip this step if you have enabled auto-inventory in the LCM page in Prism Central.

4. Go to Updates > Software. Select the Gateway version you want to upgrade to and click Update.

   LCM upgrades the Gateway version. This process takes sometime.

VPN Connection Management

You can see the details of each VPN connection, update the connection, or delete the connection.

All your VPN connections are displayed in the VPN Connections page.

Displaying the Details of a VPN Connection

You can display details such as the gateways associated with the connection, protocol details, Xi gateway routes, throughput of the connection, and logs of the IPSec and eBGP sessions for troubleshooting purposes.

About this task

Perform the following to display the details of a VPN connection.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Connections**.

   A list of all your VPN connections is displayed. The VPN connections table displays details such as the name, IPSec and eBGP status, dynamic route priority, and the VPC and gateways associated with each VPN connection.

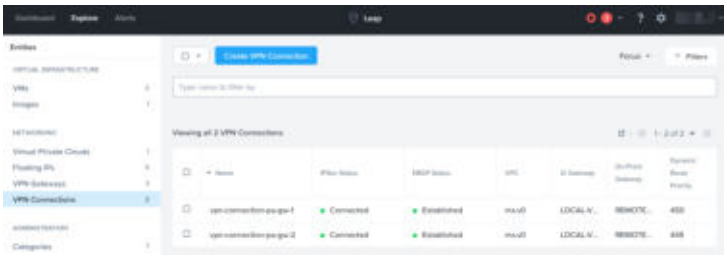


Figure 107: VPN Connections List

2. Click the name of a VPN connection to display more details of that VPN connection.

   The details page displays the following tabs:

   - **Summary**: Displays details of each gateway, protocol, and Xi gateway routes associated with the connection.
   - **Throughput**: Displays a graph for throughput of the VPN connection.
   - **IPSec Logging**: Displays logs of the IPSec sessions of the VPN connection. You can see these logs to troubleshoot any issues with the VPN connection.
   - **EBGP Logging**: Displays logs of the eBGP sessions of the VPN connection. You can see these logs to troubleshoot any issues with the VPN connection.

   Click the name of the tab to display the details in that tab. For example, click the **Summary** tab to display the details.

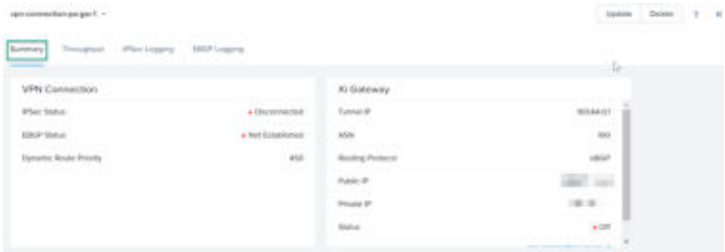


Figure 108: VPN Connection Summary Tab

Updating a VPN Connection

You can update the name, description, IPSec secret, and dynamic route priority of the VPN connection.

About this task

Perform the following to update a VPN connection.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Connections**.
2. Do one of the following:
   - Select the checkbox next to the name of the VPN connection and, in the **Actions** drop-down list, click **Update**.

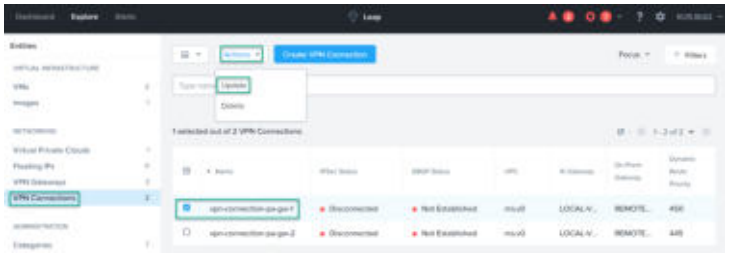


Figure 109: Use the Actions drop-down list

   - Click the name of the VPN connection and, in the details page that appears, click **Update**.

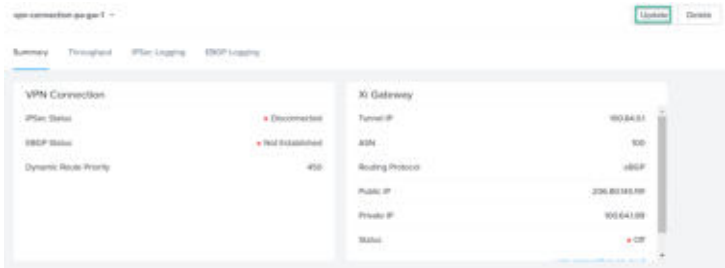


Figure 110: Click the name of the VPN connection

   The **Update VPN Connection** dialog box appears.

3. Update the details as required.
   The fields are similar to the **Create VPN Connection** dialog box. See Creating a VPN Connection on page 175 for more information.
4. Click **Save**.

Deleting a VPN Connection

About this task

Perform the following to delete a VPN connection.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Connections**.
2. Do one of the following:
   - » Select the checkbox next to the name of the VPN connection and, in the **Actions** drop-down list, click **Delete**.
   - » Click the name of the VPN connection and, in the details page that appears, click **Delete**.

3. Click **OK** in the confirmation message that appears to delete the VPN connection.

Upgrading the VPN Gateway Appliances

You can upgrade the VPN gateway VM in the Xi Cloud and on-prem VPN gateway VM in your on-prem site if you are using the On Prem - Nutanix VPN solution by using the Xi Cloud Services portal. If you are using a third-party VPN solution, you can upgrade only the VPN gateway VM running in the Xi Cloud by using the Xi Cloud Services portal. To upgrade the on-prem VPN gateway appliance provided by a third-party vendor, see the documentation of that vendor for instructions about how to upgrade the VPN appliance.

About this task

> Note: The VPN gateway VM restarts after the upgrade is complete. Therefore, perform the upgrade during a scheduled maintenance window.

Perform the following to upgrade your VPN gateway appliances.

Procedure

1. In the Xi Cloud Services portal, go to **Explore** -> **Networking** -> **VPN Gateways**.
2. Click the name of the VPN gateway.

   To upgrade the VPN gateway VM running in the Xi Cloud, select a Xi gateway.

   To upgrade the VPN gateway VM running in your on-prem site, select the on-prem gateway associated with that on-prem VPN gateway VM.
3. In the details page of the gateway, click the link in the **Version** row.

   The **VPN Version** dialog box appears.

   If you are using the latest version of the VPN gateway VM, the **VPN Version** dialog box displays a message that your VPN gateway VM is up to date.

   If your VPN gateway VM is not up to date, the **VPN Version** dialog box displays the **Upgrade** option.
4. In the **VPN Version** dialog box, click **Upgrade** to upgrade your VPN gateway VM to the latest version.

   The VPN gateway VM restarts after the upgrade is complete and starts with the latest version.

Nutanix Virtual Networks

A planned or an unplanned failover for production workloads requires production virtual networks in both the primary and the recovery site. To ensure that a failover operation, whenever necessary, goes as expected, you also need test virtual network in both the sites for testing your recovery configuration in both directions (failover and failback). To isolate production and test workflows, a recovery plan in Leap uses four separate virtual networks, which are as follows.

Two Production Networks

A production virtual network in the primary site is mapped to a production network in the recovery site. Production failover and failback are confined to these virtual networks.

Two Test Networks

The production virtual network in each site is mapped to a test virtual network in the paired site. Test failover and failback are confined to these virtual networks.

The following figures show the source and target networks for planned, unplanned, and test failovers.

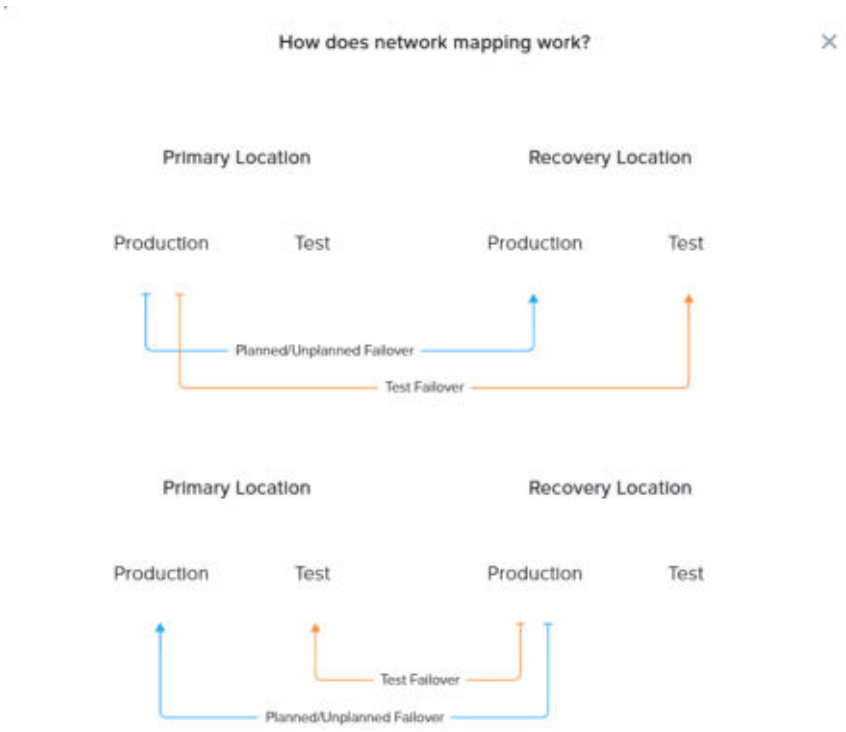


Figure 111: Virtual Network Mapping

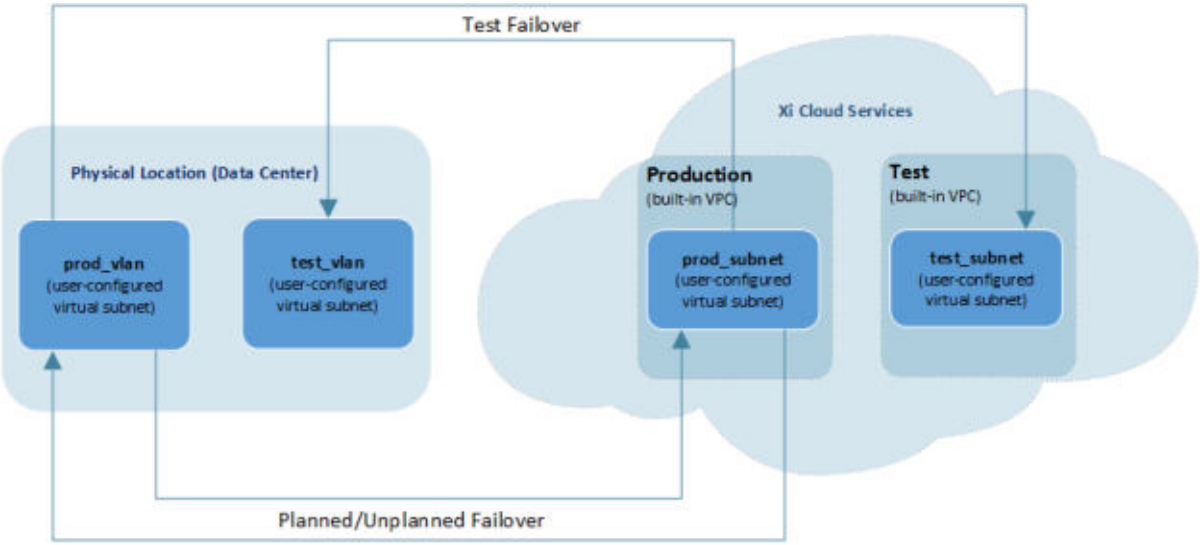


Figure 112: Virtual Network Mapping (On-Prem to Xi Cloud Services)

Virtual networks on on-prem Nutanix clusters are virtual subnets bound to a single VLAN. At on-prem sites (including the recovery site), you must manually create the production and test virtual networks before you create your first recovery plan.

The virtual networks required in Xi Cloud Services are contained within virtual private clouds (VPCs). Virtual networks required for production workloads are contained within a virtual

private cloud named production. Virtual networks required for testing failover from on-prem sites are contained within a virtual private cloud named Test. The task of creating virtual networks in the VPCs in Xi Cloud Services is an optional one. If you do not create a virtual network in a VPC, Leap dynamically creates the virtual networks for you when a failover operation is in progress. Leap cleans up dynamically created virtual networks when they are no longer required (after failback).

> Note: You cannot create more VPCs in Xi Cloud Services. However, you can update the VPCs to specify settings such as DNS and DHCP, and you can configure policies to secure the virtual networks.

Virtual Subnet Configuration in On-Prem Site

You can use your on-prem Prism Central instance to create, modify, and remove virtual networks. For information about how to perform these procedures by using Prism Central, see the *Prism Central Guide*.

Virtual Subnet Configuration in Xi Cloud Services

You can create virtual subnets in the production and test virtual networks. This is an optional task. You must perform these procedures in Xi Cloud Services. For more information, see the *Xi Infrastructure Services Guide*.

Xi Leap RPO Sizer

Nutanix offers standard service level agreements (SLAs) for data replication from your on-prem AHV clusters to Xi Cloud Services based on RPO and RTO. The replication to Xi Cloud Services occurs over public Internet (VPN or DirectConnect) and therefore Nutanix cannot control the network bandwidth available for replication to Xi Cloud Services. The unstable network bandwidth and the lack of network information affects the amount of data that can be replicated in a given timeframe. You can test your RPO objectives by setting up a real protection policy or use Xi Leap RPO sizer utility to simulate the protection plan (without replicating data to Xi Cloud Services). Xi Leap RPO Sizer provides you with information required to determine if the RPO SLAs are achievable. The utility provides insights on your network bandwidth, estimates performance, calculates actual change rate, and calculates the feasible RPO for your data protection plan.

About this task

See Xi Leap Service-Level Agreements (SLAs) on page 161 for more information about Nutanix SLAs for data replication to Xi Cloud Services. To use the Xi Leap RPO Sizer utility, perform the following steps.

Procedure

1. Log on to the My Nutanix portal with your account credentials.
2. Click **Launch** in Xi Leap RPO Sizer widget.
3. (optional) Download the bundle (rpo_sizer.tar) using the hyperlink given in the instructions.

> Tip: You can also download the bundle directly (using `wget` command in CLI) into the directory after step 4.a.

4. Log on to the on-prem Prism Central VM through an SSH session (as the "nutanix" user) and do the following.

   a. Create a separate directory to ensure that all the downloaded and extracted files inside the downloaded bundle remain in one place.

   ```
   nutanix@cvm$ mkdir dir_name
   ```

   Replace *dir_name* with an identifiable name. For example, rpo_sizer.

   b. (optional) Copy the downloaded bundle into the directory created in the previous step.

   ```
   nutanix@cvm$ cp download_bundle_path/rpo_sizer.tar ./dir_name/
   ```

   Replace *download_bundle_path* with the path to the downloaded bundle.

   Replace *dir_name* with the directory name created in the previous step.

   > Tip: If you download the bundle directly (using **wget** command in CLI) from the directory, you can skip this step.

   c. Go to the directory where the bundle is stored and extract the bundle.

   ```
   nutanix@cvm$ cd ./dir_name
   ```

   Replace *dir_name* with the directory name created in the step 4.a.

   ```
   nutanix@cvm$ tar -xvf rpo_sizer.tar
   ```

   This command generates rpo_sizer.sh and rposizer.tar in the same directory.

   d. Change the permissions to make the extracted shell file executable.

   ```
   nutanix@cvm$ chmod +x rpo_sizer.sh
   ```

   e. Run the shell script in the bundle.

   ```
   nutanix@cvm$ ./rpo_sizer.sh
   ```

   > Note: If you ran the Xi Leap RPO Sizer previously on the Prism Central VM, ensure that you clean up the script before you run the shell script again. Run the command **./rpo_sizer.sh delete** to clean up the script. If you do not clean up the script, you get an error similar to
   >
   > ```
   > The container name "/rpo_sizer" is already in use by container "xxxx"(where xxxx
   > is the container name. You have to remove (or rename) that container to be able
   > to reuse that name
   > ```
   >
   > .

5. Open a web browser and go to **http://*Prism_Central_IP_address*:8001/** to run the RPO test.

   Replace *Prism_Central_IP_address* with the virtual IP address of your Prism Central deployment.

   > Note: If you have set up a firewall on Prism Central, ensure that the port 8001 is open. If the port 8001 is closed, open it.
   >
   > ```
   > nutanix@cvm$ modify_firewall -p 8001 -o open -i eth0 -a
   > ```
   >
   > Close the port after running the RPO test.
   >
   > ```
   > nutanix@cvm$ modify_firewall -p 8001 -o close -i eth0 -a
   > ```

6. Click **Configure and execute test** and specify the following information in the Configuration Wizard.

   > Note: If you are launching the Xi Leap RPO Sizer utility for the first time, generate an API key pair. To generate API key pair, see *Creating an API Key* in the *Nutanix Licensing Guide* guide.

   a. In the **API Key and PC Credentials** tab, specify the following information.

      - 1. **API Key**: Enter the API key you generated.
        2. **Key ID**: Enter the Key ID you generated.
        3. **PC IP**: Enter the IP address of Prism Central VM.
        4. **Username**: Enter the username of your Prism Central deployment.
        5. **Password**: Enter the password of your Prism Central deployment.
        6. Click **Next**.

   b. In the **Select Desired RPO and Entities** tab, select the desired RPO from the drop-down list, select the VM Categories or individual VMs, and click **+**. If you want add more RPO and entities to the test, enter the information again and click **Next**.

      > Note: Only when you select the desired RPO, you can select the VM Categories or individual VMs on which you can test the selected RPO.

      The system discovers Prism Element automatically based on the VM Categories and the individual VMs you choose.

   c. In the **Enter PE credentials** tab, enter the SSH password or SSH key for Prism Element ("nutanix" user) running on AHV cluster and click **Next**.

   d. In the **Network Configuration** tab, specify the following information.

      - 1. **Select region**: Select a region closest to Xi Leap datacenter from the drop-down list where the workloads should be copied.
        2. **Select availability zone**: Select an availability zone (site) from the drop-down list.
        3. **NAT Gateway IPs**: Enter the public facing IP address of Prism Element running on your AHV cluster.

           > Note: To find the NAT gateway IP address of Prism Element running on your AHV cluster, log on to Prism Element through an SSH session (as the "nutanix" user) and run the `curl ifconfig.me` command.

           > Note: Do not turn on the **Configure Advanced Options** toggle unless advised by the Nutanix Support.

        4. Click **Next**.

   e. In the **View Configuration** tab, review the RPO, entity, and network configuration, the estimated test duration, and click **Submit**.

   The new window shows the ongoing test status in a progress bar.

7. When the RPO test completes, click **Upload result** to upload the test result. using upload result button. To view the detailed and intuitive report of the test, click **View Report**. To abort the test, click **X**.

   > Note: If a test is in progress, a new test cannot be triggered.

Protection and Automated DR (Xi Leap)

Automated data recovery (DR) configurations use protection policies to protect the guest VMs, and recovery plans to orchestrate the recovery of those guest VMs to Xi Cloud Services. With reverse synchronization, you can protect guest VMs and enable DR from Xi Cloud Services to a Nutanix cluster at an on-prem availability zone (site). You can automate protection of your guest VMs with the following supported replication schedules in Xi Leap.

- Asynchronous replication schedule (1 hour or greater RPO). For information about protection with Asynchronous replication schedule, see Protection with Asynchronous Replication and DR (Xi Leap) on page 196.
- NearSync replication schedule (1–15 minute RPO). For information about protection with NearSync replication schedule, see Protection with NearSync Replication and DR (Xi Leap) on page 228.

Protection with Asynchronous Replication and DR (Xi Leap)

Asynchronous replication schedules enable you to protect your guest VMs with an RPO of 1 hour or beyond. A protection policy with an Asynchronous replication schedule creates a recovery point in an hourly time interval, and replicates it to Xi Cloud Services for High Availability. For guest VMs protected with Asynchronous replication schedule, you can perform disaster recovery (DR) to Xi Cloud Services. With reverse synchronization, you can perform DR from Xi Cloud Services to a Nutanix cluster at an on-prem site. In addition to performing DR from AHV clusters to Xi Cloud Services (only AHV), you can also perform cross-hypervisor disaster recovery (CHDR)—DR from ESXi clusters to Xi Cloud Services.

> Note: Nutanix provides multiple disaster recovery (DR) solutions to secure your environment. See Nutanix Disaster Recovery Solutions on page 11 for the detailed representation of the DR offerings of Nutanix.

Asynchronous Replication Requirements (Xi Leap)

The following are the specific requirements for protecting your guest VMs with Asynchronous replication schedule. Ensure that you meet the following requirements in addition to the general requirements of Xi Leap.

For information about the general requirements of Xi Leap, see Xi Leap Requirements on page 156.

For information about the on-prem node, disk and Foundation configurations required to support Asynchronous replication schedules, see On-Prem Hardware Resource Requirements on page 15.

Hypervisor Requirements

AHV or ESXi

- The AHV clusters must be running on AHV versions that come bundled with the latest version of AOS.
- The ESXi clusters must be running on version ESXi 6.5 GA or newer.

Nutanix Software Requirements

The on-prem Prism Central and their registered clusters (Prism Elements) must be running the following versions of AOS.

- AOS 5.10 or newer with AHV.

- AOS 5.11 or newer with ESXi.

Xi Cloud Services runs the latest versions of AOS.

Cross Hypervisor Disaster Recovery (CHDR) Requirements

Guest VMs protected with Asynchronous replication schedule support cross-hypervisor disaster recovery. You can perform failover (DR) to recover guest VMs from ESXi clusters to AHV clusters (Xi Cloud Services) by considering the following requirements.

- The on-prem Nutanix clusters must be running AOS 5.11 or newer.
- Install and configure Nutanix Guest Tools (NGT) on all the guest VMs. For more information, see *Enabling and Mounting Nutanix Guest Tools* in *Prism Web Console Guide*.

  NGT configures the guest VMs with all the required drivers for VM portability. For more information about general NGT requirements, see *Nutanix Guest Tools Requirements and Limitations* in *Prism Web Console Guide*.
- CHDR supports guest VMs with flat files only.
- CHDR supports IDE/SCSI and SATA disks only.
- For all the non-boot SCSI disks of Windows guest VMs, set the SAN policy to OnlineAll so that they come online automatically.
- In vSphere 6.7, guest VMs are configured with UEFI secure boot by default. Upon CHDR to an AHV cluster, these guest VMs do not start if the host does not support the UEFI secure boot feature. For more information about supportability of UEFI secure boot on Nutanix clusters, see https://portal.nutanix.com/page/documents/compatibility-interoperability-matrix/guestos.

  For operating systems that support UEFI and Secure Boot, see UEFI and Secure Boot Support for CHDR on page 227.
- Nutanix does not support vSphere inventory mapping (for example, VM folder and resource pools) when protecting workloads between VMware clusters.
- Nutanix does not support vSphere snapshots or delta disk files.

  If you have delta disks attached to a guest VM and you proceed with failover, you get a validation warning and the guest VM does not recover. Contact Nutanix Support for assistance.

Table 40: Operating Systems Supported for CHDR

Operating System	Version	Requirements and limitations
Windows	• Windows 2008 R2 or newer versions • Windows 7 or newer versions	• Only 64-bit operating systems are supported.

Operating System	Version	Requirements and limitations
Linux	• CentOS 6.5 and 7.0 • RHEL 6.5 or newer and RHEL 7.0 or newer. • Oracle Linux 6.5 and 7.0 • Ubuntu 14.04	• SLES operating system is not supported.

Additional Requirement

The storage container name of the protected guest VMs must be the same on both the primary and recovery clusters. Therefore, a storage container must exist on the recovery cluster with the same name as the one on the primary cluster. For example, if the protected VMs are in the *SelfServiceContainer* storage container on the primary cluster, there must also be a *SelfServiceContainer* storage container on the recovery cluster.

Asynchronous Replication Limitations (Xi Leap)

Consider the following specific limitations before protecting your guest VMs with Asynchronous replication schedule. These limitations are in addition to the general limitations of Xi Leap.

For information about the general limitations of Leap, see Xi Leap Limitations on page 159.

- You cannot restore guest VMs with incompatible GPUs at the recovery site.
- You cannot protect guest VMs configured as part of a network function chain.
- You cannot retain hypervisor-specific properties after cross hypervisor disaster recovery (CHDR).

  Cross hypervisor disaster recovery (CHDR) does not preserve hypervisor-specific properties (for example, multi-writer flags, independent persistent and non-persistent disks, changed block tracking (CBT), PVSCSI disk configurations).

Creating a Protection Policy with Asynchronous Replication Schedule (Xi Leap)

To protect the guest VMs in an hourly replication schedule, configure an Asynchronous replication schedule while creating the protection policy. The policy takes recovery points of those guest VMs in the specified time intervals (hourly) and replicates them to Xi Cloud Services for High Availability. With reverse synchronization, you can create policy at Xi Cloud Services and replicate to an on-prem availability zone (site). For protection from Xi Cloud Services to an on-prem site, the protection policy allows you to add only one Asynchronous replication schedule.

Before you begin

See Asynchronous Replication Requirements (Xi Leap) on page 196 and Asynchronous Replication Limitations (Xi Leap) on page 198 before you start.

About this task

To create a protection policy with an Asynchronous replication schedule, perform the following procedure at Xi Cloud Services. You can also create a protection policy at the on-prem site. Protection policies you create or update at the on-prem site synchronize back to Xi Cloud Service.

Procedure

1. Log on to Xi Cloud Services.
2. Click the hamburger icon at the top-left corner of the window. Go to **Data Protection & Recovery** > **Protection Policies** in the left pane.
3. Click **Create Protection Policy**.
   Specify the following information in the Create Protection Policy window.

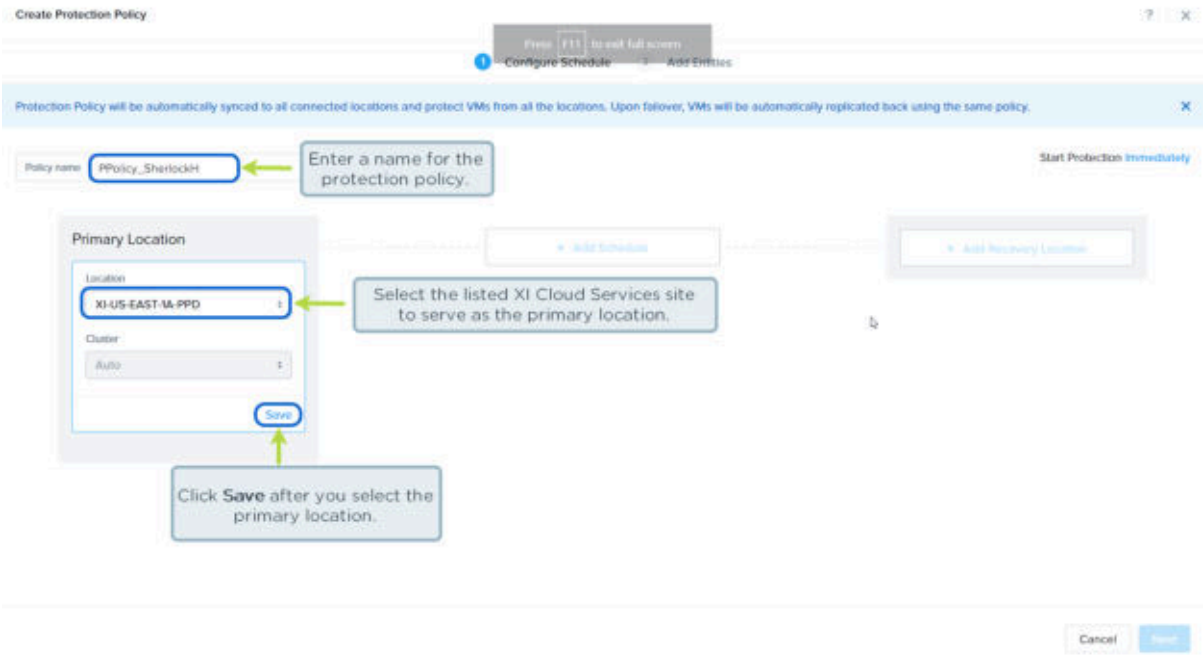


Figure 113: Protection Policy Configuration: Asynchronous

a. **Policy name**: Enter a name for the protection policy.

> Caution: The name can be of only alphanumeric, dot, dash, and underscore characters.

b. In the **Primary Location** pane, specify the following information.

- 1. **Location**: From the drop-down list, check the Xi Cloud Services availability zone (site) that hosts the guests VMs to protect.

  The drop-down lists all the sites paired with the local site. Local AZ represents the local site (Prism Central). For your primary site, you can check either the local site or a non-local site.
  2. **Cluster**: Xi Cloud Services automatically selects the cluster for you. Therefore the only option available is Auto.

  The drop-down lists all the Nutanix clusters registered to Prism Central representing the selected site. If you want to protect the guest VMs from multiple Nutanix clusters in the same protection policy, check the clusters that host those guest

VMs. All Clusters protects the guest VMs of all Nutanix clusters registered to Prism Central.

3. Click **Save**.

   Clicking **Save** activates the **Recovery Location** pane. After saving the primary site configuration, you can optionally add a local schedule (step iv) to retain the recovery points at the primary site.

4. Click **+ Add Local Schedule** if you want to retain recovery points locally in addition to retaining recovery points in a replication schedule (step d.iv). For example, you can create a local schedule to retain 15 minute recovery points locally and also

an hourly replication schedule to retain recovery points and replicate them to a recovery site every 2 hours. The two schedules apply differently on the guest VMs.

Specify the following information in the Add Schedule window.

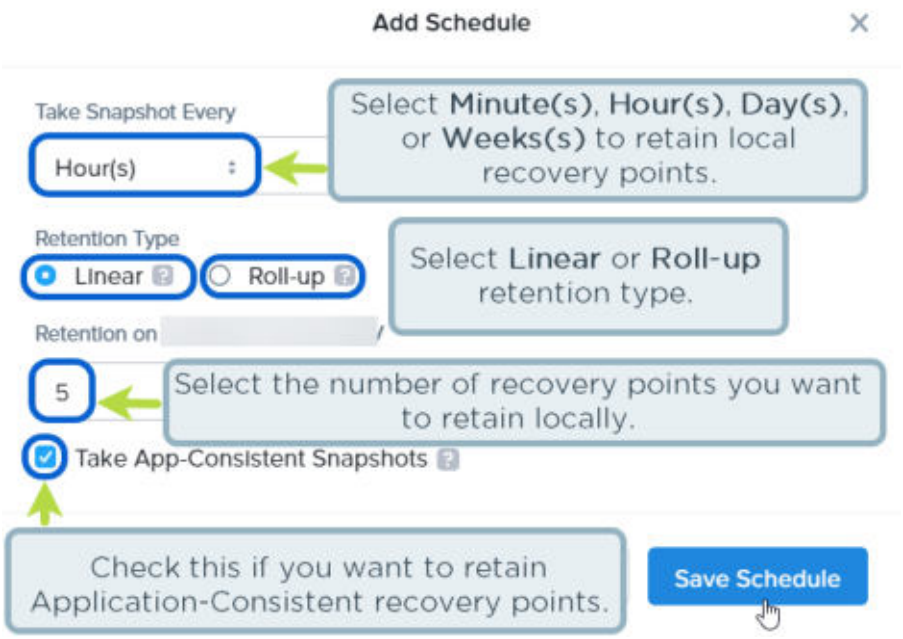


Figure 114: Protection Policy Configuration: Add Local Schedule

1. **Take Snapshot Every**: Specify the frequency in minutes, hours, days, or weeks at which you want the recovery points to be taken locally.
2. **Retention Type**: Specify one of the following two types of retention policy.

   - **Linear**: Implements a simple retention scheme at the local site. If you set the retention number to n, the local site retains the n recent recovery points.

     When you enter the frequency in minutes, the system selects the **Roll-up** retention type by default because minutely recovery points do not support **Linear** retention types.

   - **Roll-up**: Rolls up the recovery points into a single recovery point at the local site.

     For more information about the roll-up recovery points, see step d.iii.
3. **Retention on XI-US-EAST-1A-PPD : Auto**: Specify the retention number for the local site.
4. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Irrespective of the local or replication schedules, the recovery points are of the specified type. If you check **Take App-Consistent Recovery Point**, the recovery points generated are application-consistent and if you do not check **Take App-Consistent Recovery Point**, the recovery points generated are crash-consistent.

If the time in the local schedule and the replication schedule match, the single recovery point generated is application-consistent.

> Note: See Application-consistent Recovery Point Conditions and Limitations on page 55 before you take application-consistent snapshot.

5. Click **Save Schedule**.

c. In the **Recovery Location** pane, specify the following information.

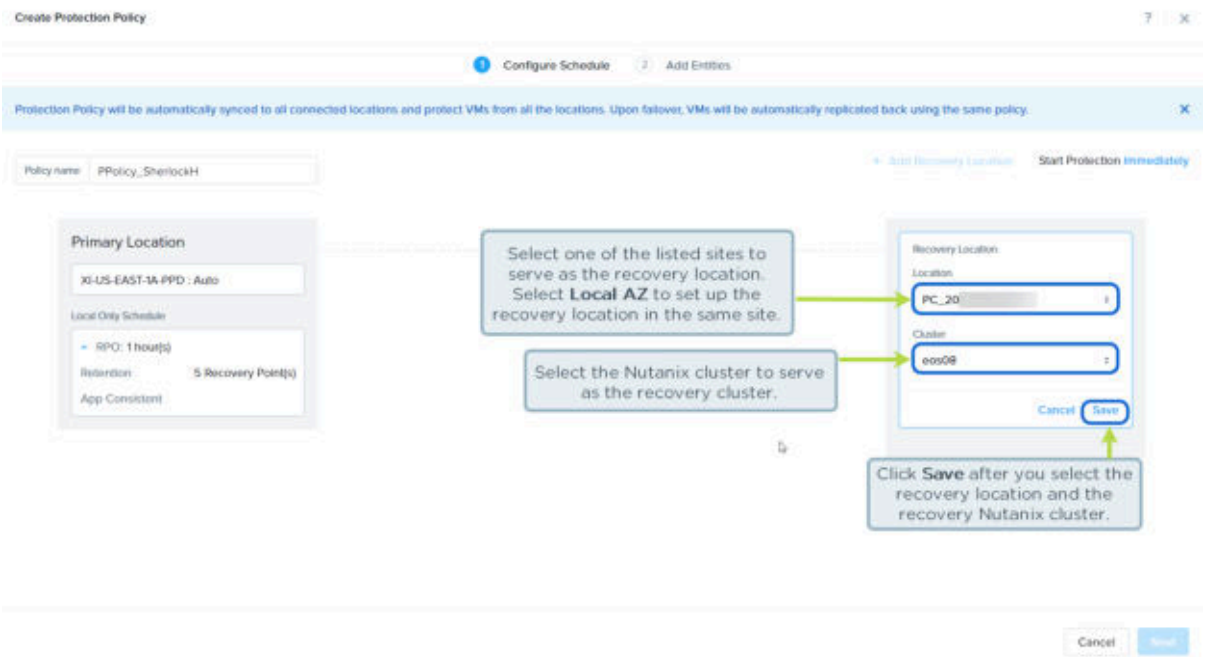


Figure 115: Protection Policy Configuration: Select Recovery Location

- 1. **Location**: From the drop-down list, select the availability zone (site) where you want to replicate the recovery points.

  The drop-down lists all the sites paired with the Xi Cloud Services. XI-US-EAST-1A-PPD : Auto represents the local site (Prism Central). Do not select XI-US-EAST-1A-PPD : Auto because a duplicate location is not supported in Xi Cloud Services.

  If you do not select a site, local recovery points that are created by the protection policy do not replicate automatically. You can, however, replicate the recovery points manually and use recovery plans to recover the guest VMs. For more information, see Protection and Manual DR (Leap) on page 143.

  2. **Cluster**: Xi Cloud Services automatically selects the cluster for you. Therefore the only option available is Auto.

  The drop-down lists all the Nutanix clusters registered to Prism Central representing the selected site. If you want to protect the guest VMs from multiple Nutanix clusters in the same protection policy, check the clusters that host those guest

VMs. All Clusters protects the guest VMs of all Nutanix clusters registered to Prism Central.

3. Click **Save**.

   Clicking **Save** activates the **+ Add Schedule** button between the primary and the recovery site. After saving the recovery site configuration, you can optionally add a local schedule to retain the recovery points at the recovery site.

4. Click **+ Add Local Schedule** if you want to retain recovery points locally in addition to retaining recovery points in a replication schedule (step d.iv). For example, you can create a local schedule to retain one hourly recovery points locally to supplement the hourly replication schedule. The two schedules apply differently on the guest VMs after failover, when the recovery points replicate back to the primary site.

   Specify the following information in the Add Schedule window.

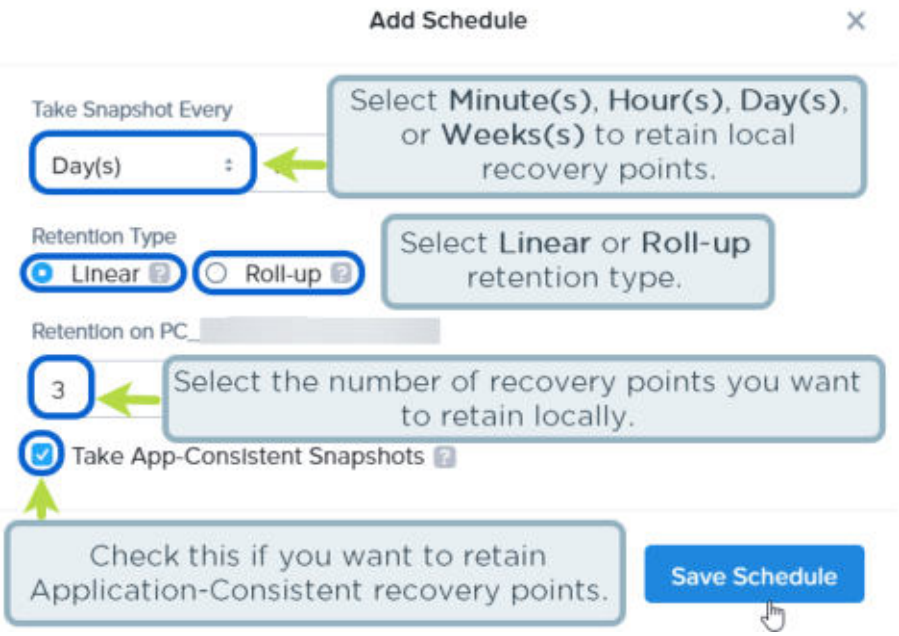


Figure 116: Protection Policy Configuration: Add Local Schedule

1. **Take Snapshot Every**: Specify the frequency in minutes, hours, days, or weeks at which you want the recovery points to be taken locally.

2. **Retention Type**: Specify one of the following two types of retention policy.

   - **Linear**: Implements a simple retention scheme at the local site. If you set the retention number to n, the local site retains the n recent recovery points.

     When you enter the frequency in minutes, the system selects the **Roll-up** retention type by default because minutely recovery points do not support **Linear** retention types.

   - **Roll-up**: Rolls up the recovery points into a single recovery point at the local site.

     For more information about the roll-up recovery points, see step d.iii.

3. **Retention on PC_xx.xx.xxx:PE_yyy**: Specify the retention number for the local site.
4. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Irrespective of the local or replication schedules, the recovery points are of the specified type. If you check **Take App-Consistent Recovery Point**, the recovery points generated are application-consistent and if you do not check **Take App-Consistent Recovery Point**, the recovery points generated are crash-consistent.

If the time in the local schedule and the replication schedule match, the single recovery point generated is application-consistent.

> Note: See Application-consistent Recovery Point Conditions and Limitations on page 55 before you take application-consistent snapshot.

5. Click **Save Schedule**.

d. Click **+ Add Schedule** to add a replication schedule between the primary and the recovery site.

Specify the following information in the Add Schedule window. The window auto-populates the Primary Location and Recovery Location that you have selected in step b and step c.

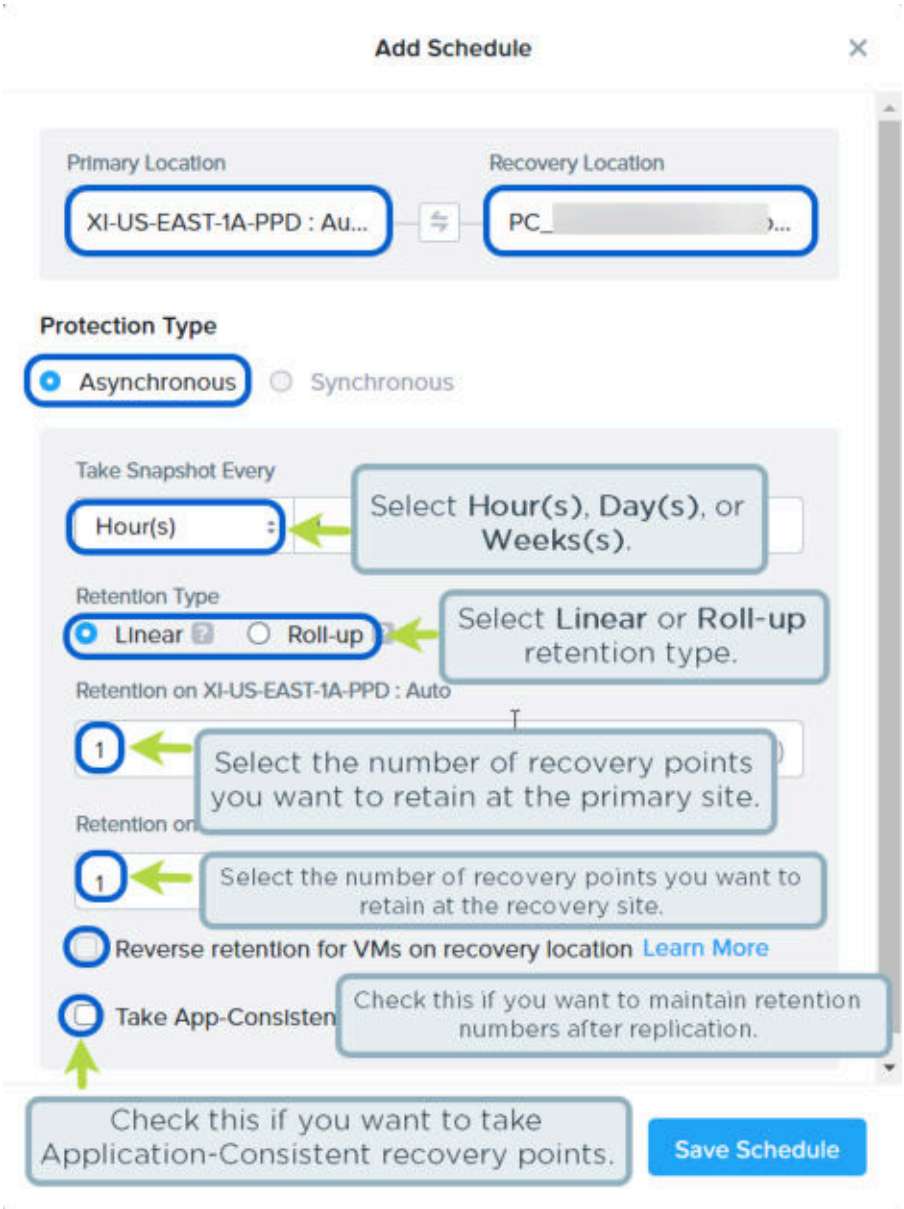


Figure 117: Protection Policy Configuration: Add Schedule (Asynchronous)

- 1. **Protection Type**: Click **Asynchronous**.

2. **Take Snapshot Every**: Specify the frequency in hours, days, or weeks at which you want the recovery points to be taken.

   The specified frequency is the RPO. For more information about RPO, see Leap Terminology on page 8.
3. **Retention Type**: Specify one of the following two types of retention policy.

   - **Linear**: Implements a simple retention scheme at both the primary (local) and the recovery (remote) site. If you set the retention number for a given site to n, that site retains the n recent recovery points. For example, if the RPO is 1 hour, and the retention number for the local site is 48, the local site retains 48 hours (48 X 1 hour) of recovery points at any given time.

     > Tip: Use linear retention policies for small RPO windows with shorter retention periods or in cases where you always want to recover to a specific RPO window.

   - **Roll-up**: Rolls up the recovery points as per the RPO and retention period into a single recovery point at a site. For example, if you set the RPO to 1 hour, and the retention time to 5 days, the 24 oldest hourly recovery points roll up into a single daily recovery point (one recovery point = 24 hourly recovery points) after every 24 hours. The system keeps one day (of rolled-up hourly recovery points) and 4 days of daily recovery points.

     > Note:
     >
     > - If the retention period is n days, the system keeps 1 day of RPO (rolled-up hourly recovery points) and n-1 days of daily recovery points.
     > - If the retention period is n weeks, the system keeps 1 day of RPO, 1 week of daily and n-1 weeks of weekly recovery points.
     > - If the retention period is n months, the system keeps 1 day of RPO, 1 week of daily, 1 month of weekly, and n-1 months of monthly recovery points.
     > - If the retention period is n years, the system keeps 1 day of RPO, 1 week of daily, 1 month of weekly, and n-1 months of monthly recovery points.

     > Note: The recovery points that are used to create a rolled-up recovery point are discarded.

     > Tip: Use roll-up retention policies for anything with a longer retention period. Roll-up policies are more flexible and automatically handle recovery point aging/pruning while still providing granular RPOs for the first day.

4. To specify the retention number for the primary and recovery sites, do the following.

   - **Retention on XI-US-EAST-1A-PPD : Auto**: Specify the retention number for the primary site.

     This field is unavailable if you do not specify a recovery location.

   - **Retention on PC_xx.xx.xx.xxx:PE_yyy**: Specify the retention number for the recovery site.

     If you select linear retention, the remote and local retention count represents the number of recovery points to retain at any given time. If you select roll-up retention, these numbers specify the retention period.

5. If you want to enable reverse retention of the recovery points, check **Reverse retention for VMs on recovery location**.

   > Note: **Reverse retention for VMs on recovery location** is available only when the retention numbers on the primary and recovery sites are different.

   Reverse retention maintains the retention numbers of recovery points even after failover to a recovery site in the same or different availability zones. For example, if you retain two recovery points at the primary site and three recovery points at the recovery site, and you enable reverse retention, a failover event does not change the initial retention numbers when the recovery points replicate back to the primary site. The recovery site still retains two recovery points while the primary site retains three recovery points. If you do not enable reverse retention, a failover event changes the initial retention numbers when the recovery points replicate back to the primary site. The recovery site retains three recovery points while the primary site retains two recovery points.

   Maintaining the same retention numbers at a recovery site is required if you want to retain a particular number of recovery points, irrespective of where the guest VM is after its failover.

6. If you want to take application-consistent recovery points, check **Take App-Consistent Recovery Point**.

   Application-consistent recovery points ensure that application consistency is maintained in the replicated recovery points. For application-consistent recovery points, install NGT on the guest VMs running on AHV clusters. For guest VMs running on ESXi clusters, you can take application-consistent recovery points without installing NGT, but the recovery points are hypervisor-based, and leads to VM stuns (temporary unresponsive VMs) after failover to the recovery sites.

   > Note: See Application-consistent Recovery Point Conditions and Limitations on page 55 before you take application-consistent snapshot.

   > Caution: Application-consistent recovery points fail for EFI-boot enabled Windows 2019 VM running on ESXi when NGT is not installed. Nutanix recommends installing NGT on guest VMs running on ESXi also.

7. Click **Save Schedule**.

e. Click **Next**.

   Clicking **Next** shows a list of VM categories where you can optionally check one or more VM categories to protect in the protection policy. DR configurations using Leap allows you to protect a guest VM by using only one protection policy. Therefore, VM categories specified in another protection policy are not in the list. If you protect a guest VM in another protection policy by specifying the VM category of the guest VM (category-based inclusion), and if you protect the guest VM from the **VMs** page in this policy

(individual inclusion), the individual inclusion supersedes the category-based inclusion. Effectively, only the protection policy that protected the individual guest VM protects the guest VM.

For example, the guest VM *VM_SherlockH* is in the category *Department:Admin*, and you add this category to the protection policy named *PP_AdminVMs*. Now, if you add *VM_SherlockH* from the **VMs** page to another protection policy named *PP_VMs_UK*, *VM_SherlockH* is protected in *PP_VMs_UK* and unprotected from *PP_AdminVMs*.

f. If you want to protect the guest VMs category wise, check the VM categories that you want to protect from the list and click **Add**.

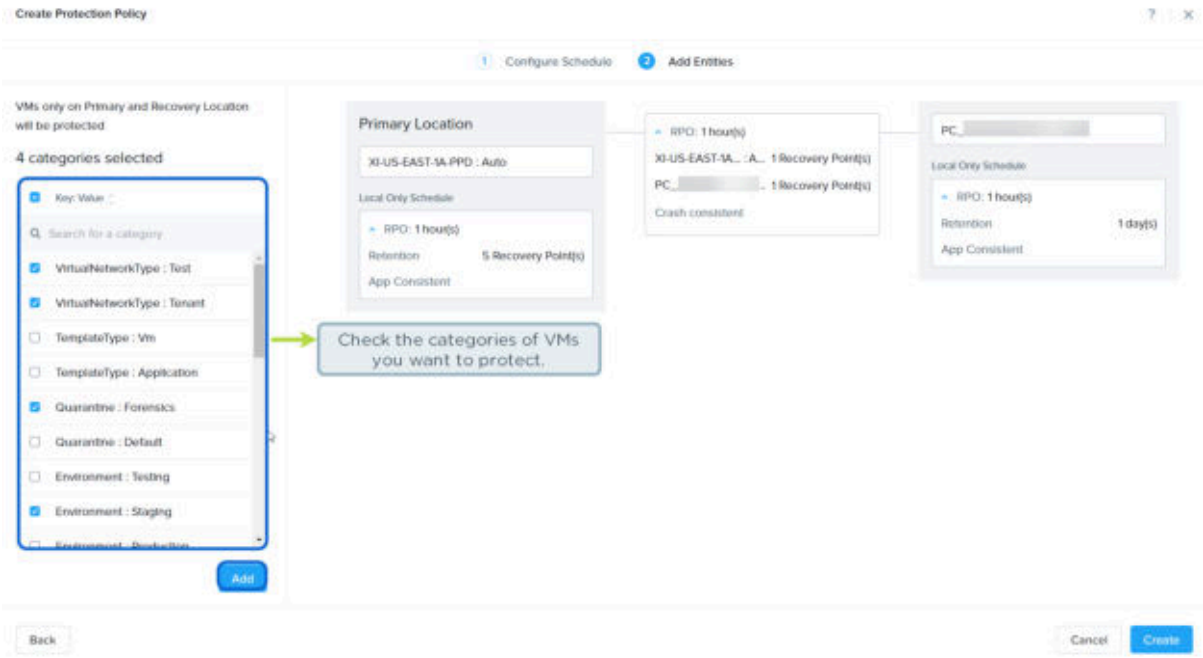


Figure 118: Protection Policy Configuration: Add VM Categories

Prism Central includes built-in VM categories for frequently encountered applications (for example, MS Exchange and Oracle). If the VM category or value you want is not available, first create the category with the required values, or update an existing category so that it has the values you require. Doing so ensures that the VM categories and values are available for selection. You can add VMs to the category either before or after you configure the protection policy. If the guest VMs have a common characteristic, such as belonging to a specific application or location, create a VM category and add the guest VMs into the category.

If you do not want to protect the guest VMs category wise, proceed to the next step without checking VM categories. You can add the guest VMs individually to the protection policy later from the **VMs** page (see Adding Guest VMs Individually to a Protection Policy on page 136).

g. Click **Create**.
The protection policy with an Asynchronous replication schedule is created. To verify the protection policy, see the **Protection Policies** page. You can add VMs individually (without VM categories) to the protection policy or remove VMs from the protection policy. For information about the operations that you can perform on a protection policy, see Protection Policy Management on page 237.

Creating a Recovery Plan (Xi Leap)

To orchestrate the failover (disaster recovery) of the protected guest VMs to the recovery site, create a recovery plan. After a failover, a recovery plan recovers the protected guest VMs to the recovery availability zone (site). To create a recovery plan, perform the following procedure at Xi Cloud Services. You can also create a recovery plan at the on-prem site. The recovery plan you create or update at the on-prem site synchronizes back to Xi Cloud Service.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Recovery Plans** in the left pane.
3. Click **Create Recovery Plan**.

   Specify the following information in the Create Protection Policy window.

   a. **Primary Location**: Select the primary site that hosts the guest VMs to protect. This list displays the **Local AZ** by default and is unavailable for editing.

   b. **Recovery Location**: Select the on-prem site where you want to replicate the recovery points.

   c. Click **Proceed**.

   > Tip: After you create the recovery plan, you cannot change the **Recovery Location** from the **Recovery Plans** page. To change the recovery location on an existing recovery plan, do the following.
   >
   > - Update the protection policy to point to the new recovery location. For more information, see Updating a Protection Policy on page 240.
   > - Configure the network mapping. For more information, see Nutanix Virtual Networks on page 191.

   > Caution: If all the VMs in the recovery plan do not point to the new recovery location, you get an availability zone conflict alert.

4. In the **General** tab, enter **Recovery Plan Name**, **Recovery Plan Description**. Click **Next**.

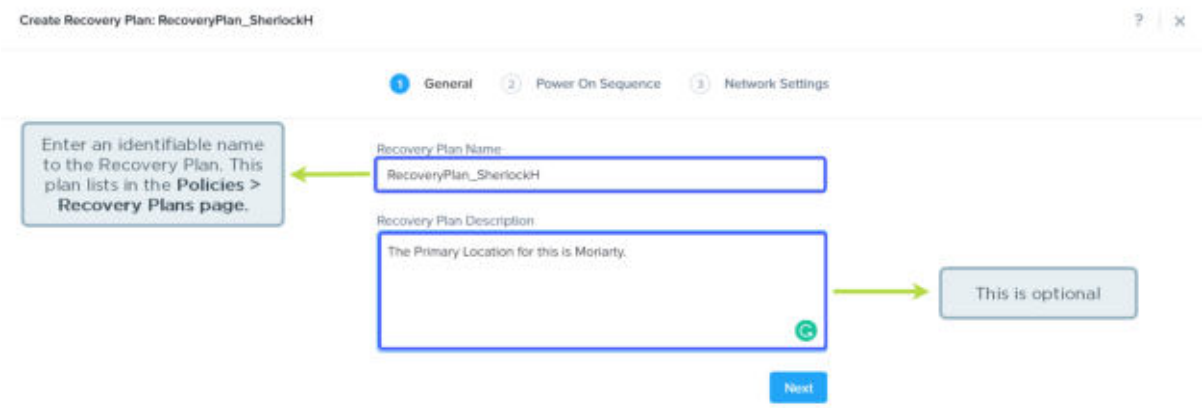


Figure 119: Recovery Plan Configuration: General

5. In the **Power On Sequence** tab, click **+ Add Entities** to add VMs to the sequence and do the following.

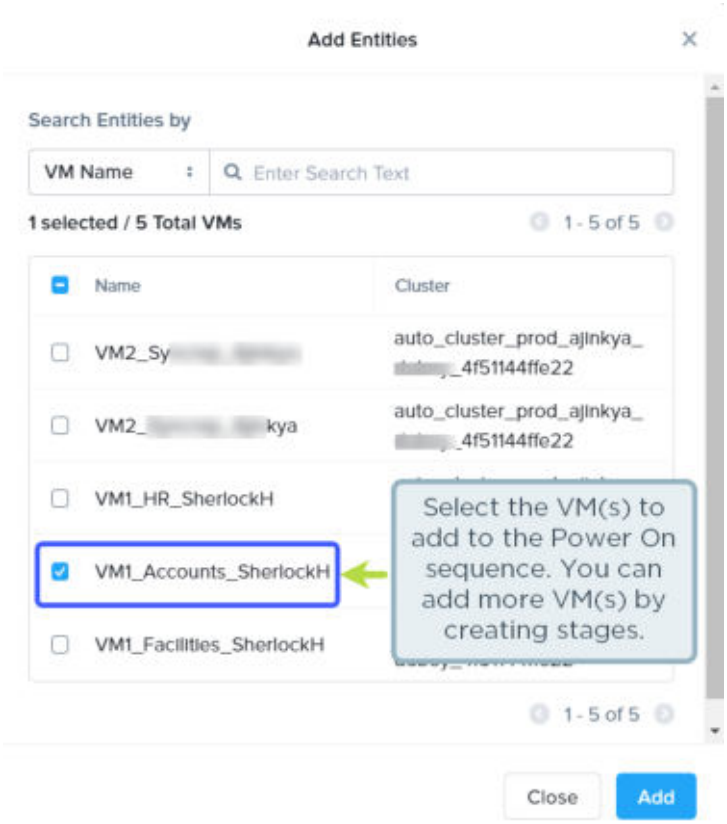


Figure 120: Recovery Plan Configuration: Add Entities

a. In the **Search Entities by**, select **VM Name** from the drop-down list to specify VMs by name.

b. In the **Search Entities by**, select **Category** from the drop-down list to specify VMs by category.

c. To add the VMs or VM categories to the stage, select the VMs or VM categories from the list.

> Note: The VMs listed in the search result are in the active state of replication.

d. Click **Add**.

The selected VMs are added to the sequence. You can also create multiple stages and add VMs to those stages to define their power-on sequence. For more information about stages, see Stage Management on page 68.

> Caution: Do not include the guest VMs protected with Asynchronous, NearSync, and Synchronous replication schedules in the same recovery plan. You can include guest VMs protected with Asynchronous or NearSync replication schedules in the same recovery plan. However, if you combine these guest VMs with the guest VMs protected by Synchronous replication schedules in a recovery plan, the recovery fails.

6. To manage in-guest script execution on guest VMs during recovery, select the individual VMs or VM categories in the stage. Click **Manage Scripts** and then do the following.

Note: In-guest scripts allow you to automate various task executions upon recovery of the VMs. For example, in-guest scripts can help automate the tasks in the following scenarios.

- After recovery, the VMs must use new DNS IP addresses and also connect to a new database server that is already running at the recovery site.

  Traditionally, to achieve this new configuration, you would manually log on to the recovered VM and modify the relevant files. With in-guest scripts, you have to write a script to automate the required steps and enable the script when you configure a recovery plan. The recovery plan execution automatically invokes the script and performs the reassigning of DNS IP address and reconnection to the database server at the recovery site.

- If VMs are part of domain controller *siteA.com* at the primary site *AZ1*, and after the VMs recover on the site *AZ2*, you want to add the recovered VMs to the domain controller *siteB.com*.

  Traditionally, to reconfigure, you would manually log on to the VM, remove the VM from an existing domain controller, and then add the VM to a new domain

controller. With in-guest scripts, you can automate the task of changing the domain controller.

Note: In-guest script execution requires NGT version 1.9 or newer installed on the VM. The in-guest scripts run as a part of the recovery plan only if they have executable permissions for the following.

- Administrator user (Windows)
- Root user (Linux)

Note: You can have only two in-guest batch or shell scripts—one for production (planned and unplanned failover) while the other for test failover. One script, however, can invoke other scripts. Place the scripts at the following locations in the VMs.

- In Windows VMs,
  - Batch script file path for production failover:

    `C:\Program Files\Nutanix\Scripts\production\`*`vm_recovery`*

    Replace *`vm_recovery`* with the script name (without extension).
  - Batch script file path for test failover:

    `C:\Program Files\Nutanix\Scripts\test\`*`vm_recovery`*

    Replace *`vm_recovery`* with the script name (without extension).
- In Linux VMs,
  - Shell script file path for production failover:

    `/usr/local/sbin/`*`production_vm_recovery`*

    Replace *`production_vm_recovery`* with the script name (without extension).
  - Shell script file path for test failover:

    `/usr/local/sbin/`*`test_vm_recovery`*

    Replace *`test_vm_recovery`* with the script name (without extension).

Note: When an in-guest script runs successfully, it returns code **`0`**. Error code **`1`** signifies that the execution of the in-guest script was unsuccessful.

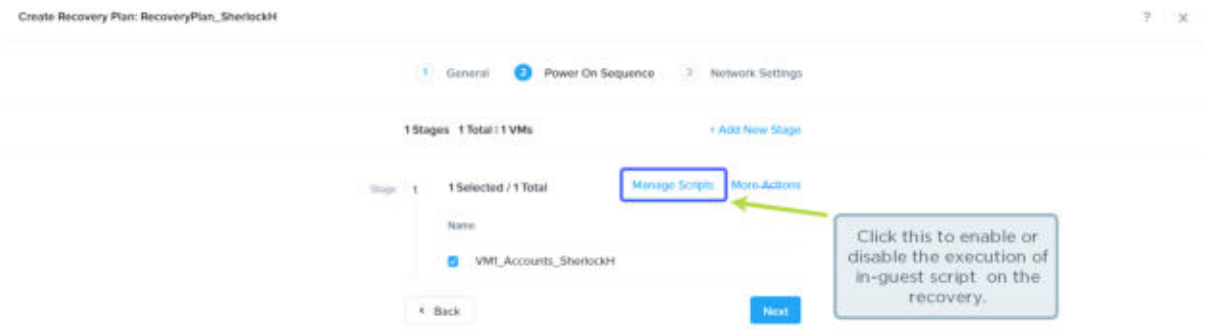


Figure 121: Recovery Plan Configuration: In-guest Script Execution

a. To enable script execution, click **Enable**.

   A command prompt icon appears against the VMs or VM categories to indicate that in-guest script execution is enabled on those VMs or VM categories.

b. To disable script execution, click **Disable**.

7. In the **Network Settings** tab, map networks in the primary cluster to networks at the recovery cluster.

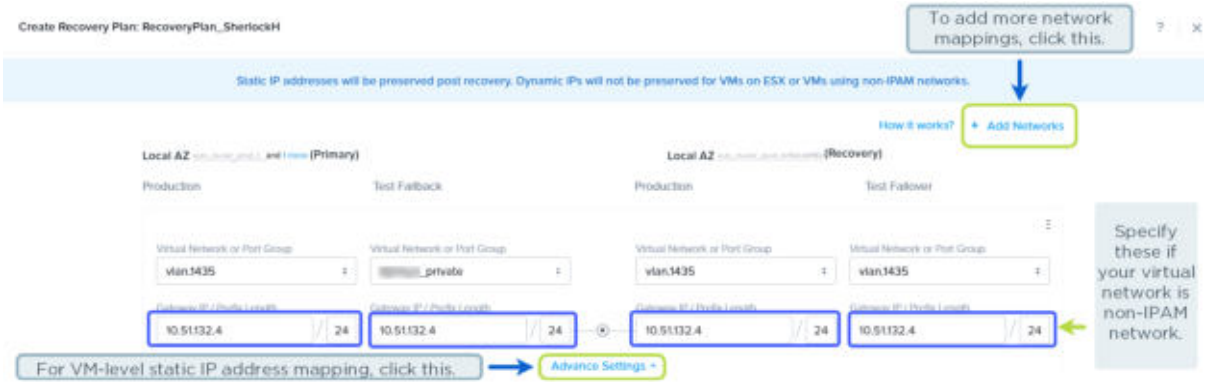


Figure 122: Recovery Plan Configuration: Network Settings

Network mapping enables replicating the network configurations of the primary clusters to the recovery clusters, and recover VMs into the same subnet at the recovery cluster. For example, if a VM is in the *vlan0* subnet at the primary cluster, you can configure the network mapping to recover that VM in the same *vlan0* subnet at the recovery cluster. To specify the source and destination network information for a network mapping, do the following in **Local AZ (Primary)** and **PC 10.51.1xx.xxx (Recovery)**.

a. Under **Production** in **Virtual Network or Port Group**, select the production subnet that contains the protected VMs for which you are configuring a recovery plan. (Optional)

If the virtual network is a non-IPAM network, specify the gateway IP address and prefix length in the **Gateway IP/Prefix Length** field.

b. Under **Test Failback** in **Virtual Network or Port Group**, select the test subnet that you want to use for testing failback from the recovery cluster. (Optional) If the virtual network is a non-IPAM network, specify the gateway IP address and prefix length in the **Gateway IP/Prefix Length** field.

c. To add a network mapping, click **Add Networks** at the top-right corner of the page, and then repeat the steps 7.a-7.b.

> Note: The primary and recovery Nutanix clusters must have identical gateway IP addresses and prefix length. Therefore you cannot use a test failover network for two or more network mappings in the same recovery plan.

d. Click **Done**.

> Note: For ESXi, you can configure network mapping for both standard and distributed (DVS) port groups. For more information about DVS, see VMware documentation.

> Caution: Leap does not support VMware NSX-T datacenters. For more information about NSX-T datacenters, see VMware documentation.

8. If you want to enable the VMs in the production VPC to access the Internet, enable **Outbound Internet Access**.

9. To assign floating IP addresses to the VMs when they are running in Xi Cloud Services, click **+ Floating IPs** in **Floating IPs** section and do the following.

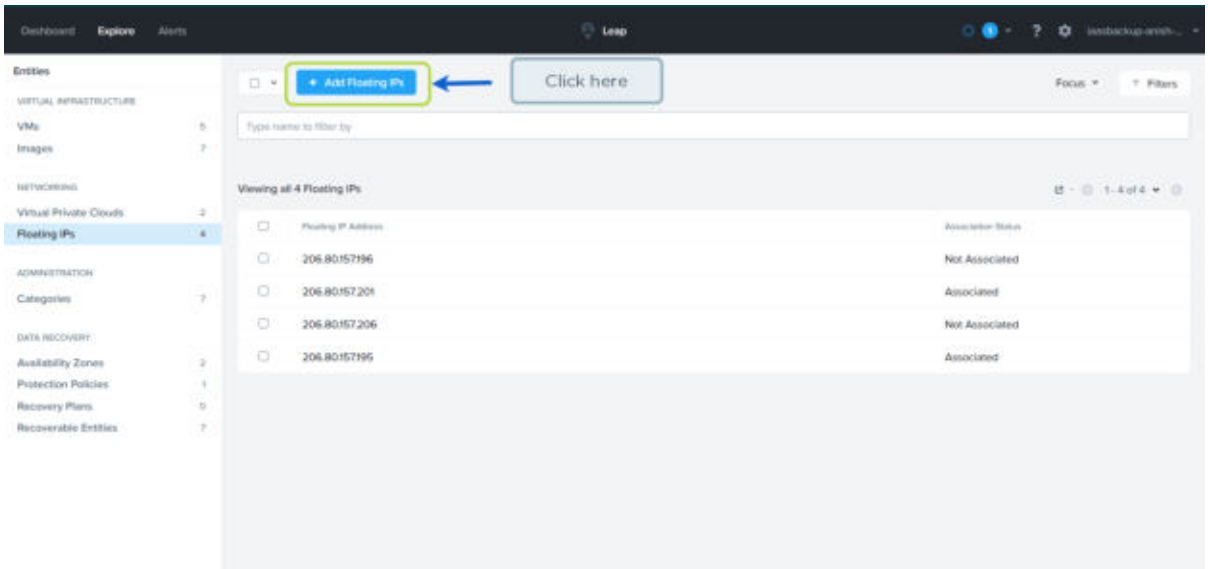


Figure 123: Recovery Plan Configuration: Assign Floating IP Address

a. In the **NUMBER OF FLOATING IPS**, enter the number of floating IP addresses you need for assigning to VMs.

b. In the **ASSIGN FLOATING IPS TO VMS (OPTIONAL)**, enter the name of the VMs and select the IP address for it.

c. In **Actions**, click **Save**.

d. To assign a floating IP address to another VM, click **+ Assign Floating IP**, and then repeat the steps for assigning a floating IP address.

10. Click **Done**.

    The recovery plan is created. To verify the recovery plan, see the **Recovery Plans** page. You can modify the recovery plan to change the recovery location, add, or remove the protected guest VMs. For information about various operations that you can perform on a recovery plan, see Recovery Plan Management on page 141.

Failover and Failback Operations (Xi Leap)

You perform failover of the protected guest VMs when unplanned failure events (for example, natural disasters) or planned events (for example, scheduled maintenance) happen at the primary availability zone (site) or the primary cluster. The protected guest VMs migrate to the recovery site where you perform the failover operations. On recovery, the protected guest VMs start in the Xi Cloud Services region you specify in the recovery plan that orchestrates the failover.

The following are the types of failover operations in Xi Leap.

Test Failover

To ensure that the protected guest VMs failover efficiently to the recovery site, you perform a test failover. When you perform a test failover, the guest VMs recover in the virtual network designated for testing purposes at the recovery site (a manually created

virtual subnet in the test VPC in Xi Cloud Services). However, the guest VMs at the primary site are not affected. Test failovers rely on the presence of VM recovery points at the recovery sites.

Planned Failover

To ensure VM availability when you foresee service disruption at the primary site, you perform a planned failover to the recovery site. For a planned failover to succeed, the guest VMs must be available at the primary site. When you perform a planned failover, the recovery plan first creates a recovery point of the protected guest VM, replicates the recovery point to the recovery site, and then starts the guest VM at the recovery site. The recovery point used for migration is retained indefinitely. After a planned failover, the guest VMs no longer run at the primary site. After a planned failover, the VMs no longer run at the primary site.

Unplanned Failover

To ensure VM availability when a disaster causing service disruption occurs at the primary site, you perform an unplanned failover to the recovery site. In an unplanned failover, you can expect some data loss to occur. The maximum data loss possible is equal to the least RPO you specify in the protection policy, or the data that was generated after the last manual recovery point for a given guest VM. In an unplanned failover, by default, the protected guest VMs recover from the most recent recovery point. However, you can recover from an earlier recovery point by selecting a date and time of the recovery point.

After the failover, replication begins in the reverse direction. You can perform an unplanned failover operation only if recovery points have replicated to the recovery cluster. At the recovery site, failover operations cannot use recovery points that were created locally in the past. For example, if you perform an unplanned failover from the primary site *AZ1* to recovery site *AZ2* in Xi Cloud Services and then attempt an unplanned failover (failback) from *AZ2* to *AZ1*, the recovery succeeds at *AZ1* only if the recovery points are replicated from *AZ2* to *AZ1* after the unplanned failover operation. The unplanned failover operation cannot perform recovery based on the recovery points that were created locally when the VMs were running in *AZ1*.

The procedure for performing a planned failover is the same as the procedure for performing an unplanned failover. You can perform a failover even in different scenarios of network failure. For more information about network failure scenarios, see Leap and Xi Leap Failover Scenarios on page 71.

Performing a Test Failover (Xi Leap)

After you create a recovery plan, you can run a test failover periodically to ensure that the failover occurs smoothly when required. You can perform the test failover from Xi Cloud Services.

About this task

To perform a test failover to Xi Cloud Services, do the following.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Recovery Plans** in the left pane.
3. Select the recovery plan that you want to test.
4. Click **Test** from the **Actions** drop-down menu.

5. In the **Test Recovery Plan** dialog box, do the following.
   a. In **Primary Location**, select the primary availability zone (site).
   b. In **Recovery Location**, select the recovery availability zone.
   c. Click **Test**.
   If you get errors or warnings, see the failure report that is displayed. Click the report to review the errors and warnings. Resolve the error conditions and then restart the test procedure.
6. Click **Close**.

Cleaning up Test VMs (Xi Leap)

After testing a recovery plan, you can remove the test VMs that the recovery plan created in the recovery test network on Xi Cloud Services. To clean up the test VMs created when you test a recovery plan, do the following.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Recovery Plans** in the left pane.
3. Click the recovery plans whose VMs you want to remove.
4. Click **Clean Up Test VMs** from the **Actions** drop-down menu.
5. In the **Clean Up Test VMs** dialog box, click **Clean**.
   Test VMs are deleted. If you get errors or warnings, see the failure report that is displayed. Click the report to review the errors and warnings. Resolve the error conditions and then restart the test procedure.

Performing a Planned Failover (Xi Leap)

Perform a planned failover at the recovery site. To perform a planned failover to Xi Cloud Services, do the following procedure.

Procedure

1. Log on to Xi Cloud Services.

2. Click **Explore** and go to **Recovery Plans** in the left pane.

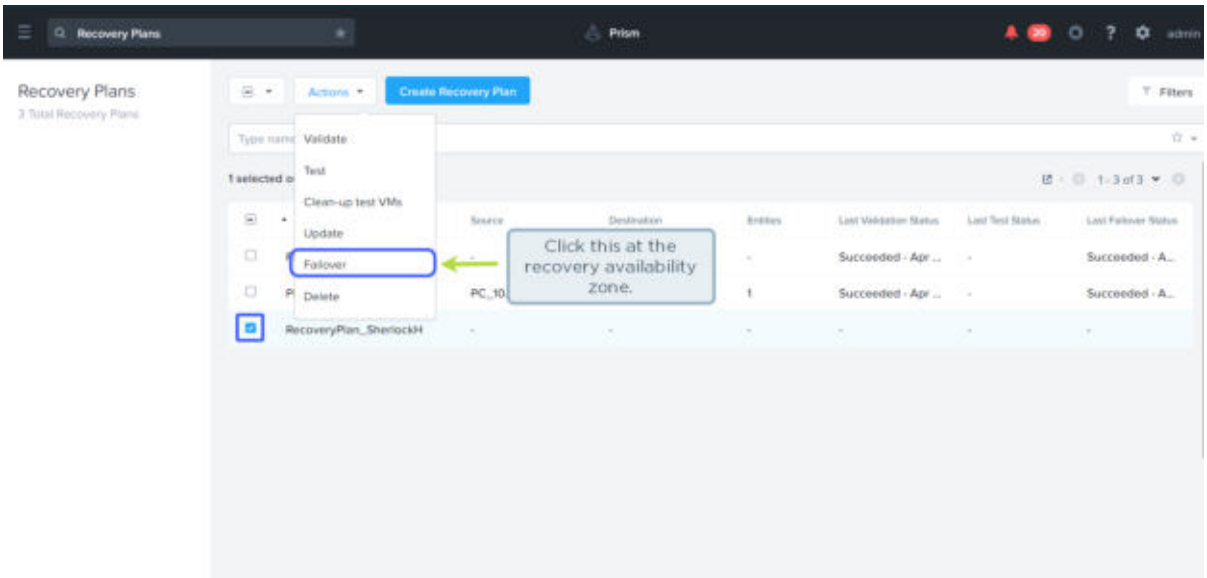


Figure 124: Planned Failover

3. Select a recovery plan for the failover operation.

4. Click **Failover** from the **Actions** drop-down menu. Specify the following information in the **Failover from Recovery Plan** dialog box.

> Note: The **Failover** action is available only when all the selected recovery plans have the same primary and recovery locations.

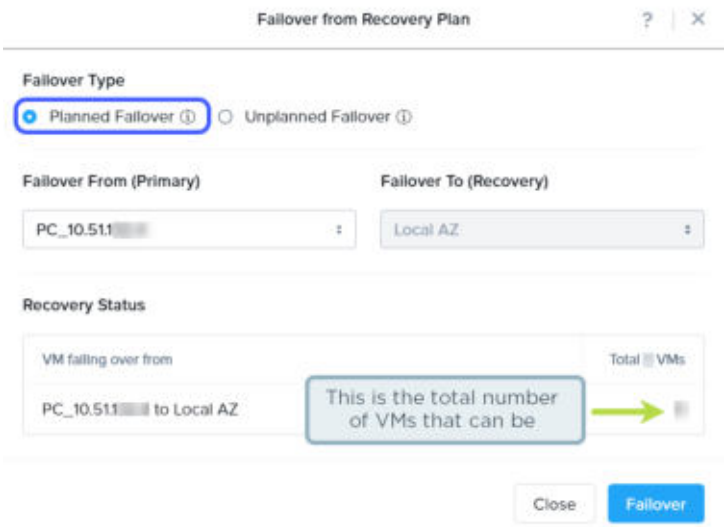


Figure 125: Planned Failover

a. **Failover Type**: Click **Planned Failover**.

b. **Failover From (Primary)**: Select the protected primary cluster.

c. **Failover To (Recovery)**: Select the recovery cluster where you want the VMs to failover. This list displays **Local AZ** by default and is unavailable for editing.

> Note: Click **+** to add more combinations of primary and recovery clusters. You can add as many primary clusters as there are in the selected recovery plan.

5. Click **Failover**.

   The **Failover from Recovery Plan** dialog box lists the errors and warnings, if any, and allows you to stop or continue the failover operation.

6. If you see errors, do the following.
   a. To review errors or warnings, click **View Details** in the description.
   b. Click **Cancel** to return to the **Failover from Recovery Plan** dialog box.
   c. Select one of the following.
      » To stop the failover operation, click **Abort**.
      » To continue the failover operation despite the warnings, click **Execute Anyway**.

      You cannot continue the failover operation when the validation fails with errors.

      Note:

      The entities of AHV/ESXi clusters recover at a different path on the ESXi clusters if their files conflict with the existing files on the recovery ESXi cluster. For example, there is a file name conflict if a VM (VM1) migrates to a recovery cluster that already has a VM (VM1) in the same container.

      However, the entities recover at a different path with `VmRecoveredAtAlternatePath` alert only if the following conditions are met.

      - Both the primary and the recovery clusters (Prism Elements) are of version 5.17 or newer.
      - A path for the entity recovery is not defined while initiating the failover operation.
      - The protected entities do not have shared disk/s.

      If these conditions are not satisfied, the failover operation fails.

Performing an Unplanned Failover (Xi Leap)

Perform an unplanned failover at the recovery site. To perform an unplanned failover to Xi Cloud Services, do the following procedure.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Recovery Plans** in the left pane.
3. Select a recovery plan for the failover operation.

4. Click **Failover** from the **Actions** drop-down menu. Specify the following information in the **Failover from Recovery Plan** dialog box.

> Note: The **Failover** action is available only when all the selected recovery plans have the same primary and recovery locations.

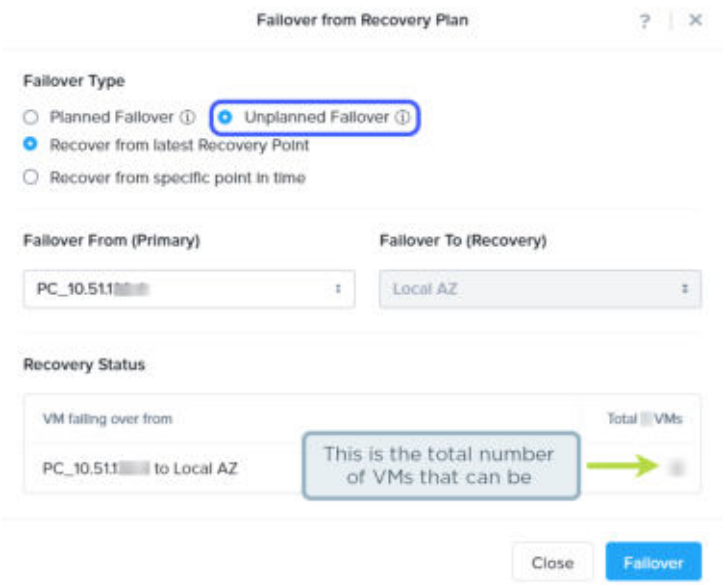


Figure 126: Unplanned Failover

a. **Failover Type**: Click **Unplanned Failover** and do one of the following.

   » Click **Recover from latest Recovery Point** to use the latest recovery point for recovery.

   » Click **Recover from specific point in time** to use a recovery point taken at a specific point in time for recovery.

b. **Failover From (Primary)**: Select the protected primary cluster.

c. **Failover To (Recovery)**: Select the recovery cluster where you want the VMs to failover. This list displays **Local AZ** by default and is unavailable for editing.

> Note: Click **+** to add more combinations of primary and recovery clusters. You can add as many primary clusters as there are in the selected recovery plan.

> Note: If recovery plans contain VM categories, the VMs from those categories recover in the same category after an unplanned failover to the recovery site. Also, the recovery points keep generating at the recovery site for those recovered VMs. Since the VM count represents the number of recoverable VMs (calculated from recovery points), the recovered VMs and their newly created recovery points sum up. Their sum gives double the count of the originally recovered VMs on the recovery plans page. Now, if some VMs belonging to the given category at the primary or recovery site are deleted, the VM count at both sites still stay the same until the recovery points of deleted VMs expire. For example, when two VMs have failed over, the recovery plans page at the recovery site shows four VMs (two replicated recovery points from source and two newly generated recovery points). The page shows four VMs even if the VMs are deleted from the primary or recovery site. The VM count synchronizes and becomes consistent in the subsequent RPO cycle conforming to the retention policy set in the protection policy (due to the expiration of recovery points).

5. Click **Failover**.
   The **Failover from Recovery Plan** dialog box lists the errors and warnings, if any, and allows you to stop or continue the failover operation.
6. If you see errors, do the following.
   a. To review errors or warnings, click **View Details** in the description.
   b. Click **Cancel** to return to the **Failover from Recovery Plan** dialog box.
   c. Select one of the following.
      » To stop the failover operation, click **Abort**.
      » To continue the failover operation despite the warnings, click **Execute Anyway**.

> Note: You cannot continue the failover operation when the validation fails with errors.

> Note:
>
> The entities of AHV/ESXi clusters recover at a different path on the ESXi clusters if their files conflict with the existing files on the recovery ESXi cluster. For example, there is a file name conflict if a VM (VM1) migrates to a recovery cluster that already has a VM (VM1) in the same container.
>
> However, the entities recover at a different path with `VmRecoveredAtAlternatePath` alert only if the following conditions are met.
>
> - Both the primary and the recovery clusters (Prism Elements) are of version 5.17 or newer.
> - A path for the entity recovery is not defined while initiating the failover operation.
> - The protected entities do not have shared disk/s.
>
> If these conditions are not satisfied, the failover operation fails.

> Note: To avoid conflicts when the primary site becomes active after the failover, shut down the guest VMs associated with this recovery plan. Manually power off the guest VMs on either primary or recovery site after the failover is complete. You can also block the guest VMs associated with this recovery plan through the firewall.

Performing Failback (Xi Leap)

A failback is similar to a failover but in the reverse. The same recovery plan applies to both the failover and the failback operations. Therefore, how you perform a failback is identical to how you perform a failover. Log on to the site where you want the VMs to failback, and then perform a failover. For example, if you failed over VMs from an on-prem site to Xi Cloud Services, to failback to the on-prem site, perform the failover from the on-prem site.

About this task

To perform a failback, do the following procedure at the primary site.

Procedure

1. Log on to the Prism Central web console.
2. Click the hamburger icon at the top-left corner of the window. Go to **Policies** > **Recovery Plans** in the left pane.
3. Select a recovery plan for the failover operation.

4. Click **Failover** from the **Actions** drop-down menu.

> Note: If you select more than one recovery plan in step 3, the **Failover** action is available only when the selected recovery plans have the same primary and recovery locations.

Specify the following information in the **Failover from Recovery Plan** window. The window auto-populates the **Failover From** and **Failover To** locations from the recovery plan you select in step 3.

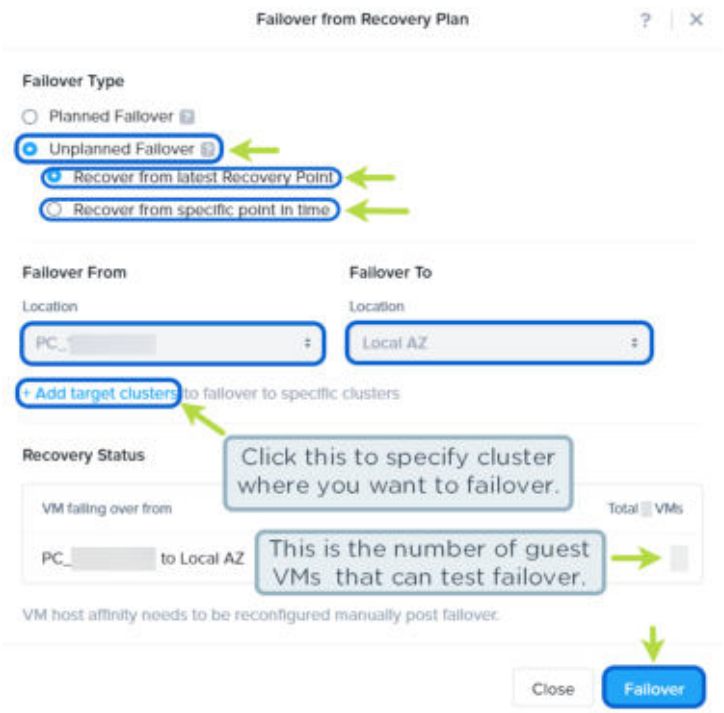


Figure 127: Unplanned Failover

a. **Failover Type**: Click **Unplanned Failover** and do one of the following.

> Tip: You can also click **Planned Failover** to perform planned failover procedure for a failback.

» Click **Recover from latest Recovery Point** to use the latest recovery point for recovery.

» Click **Recover from specific point in time** to use a recovery point taken at a specific point in time for recovery.

b. Click **+ Add target clusters** if you want to failover to specific Nutanix clusters at the primary site.

If you do not add target clusters, the recovery plan recovers the guest VMs to any eligible cluster at the primary site.

> Note: If recovery plans contain VM categories, the VMs from those categories recover in the same category after an unplanned failover to the recovery site. Also, the recovery points keep generating at the recovery site for those recovered VMs. Since the VM count represents the number of recoverable VMs (calculated from recovery points), the recovered VMs and their newly created recovery points sum up. Their sum gives double the count of the originally recovered VMs on the recovery plans page. Now, if some VMs belonging to the given

category at the primary or recovery site are deleted, the VM count at both sites still stay the same until the recovery points of deleted VMs expire. For example, when two VMs have failed over, the recovery plans page at the recovery site shows four VMs (two replicated recovery points from source and two newly generated recovery points). The page shows four VMs even if the VMs are deleted from the primary or recovery site. The VM count synchronizes and becomes consistent in the subsequent RPO cycle conforming to the retention policy set in the protection policy (due to the expiration of recovery points).

5. Click **Failover**.
   The **Failover from Recovery Plan** dialog box lists the errors and warnings, if any, and allows you to stop or continue the failover operation. If there are no errors or you resolve the errors in step 6, the guest VMs failover to the recovery cluster.

6. If you see errors, do the following.

   - To review errors or warnings, click **View Details** in the description.

     Resolve the error conditions and then restart the failover procedure.

   - Select one of the following.

     - To stop the failover operation, click **Abort**.
     - To continue the failover operation despite the warnings, click **Execute Anyway**.

       Note: You cannot continue the failover operation when the validation fails with errors.

   Note:

   The entities of AHV/ESXi clusters recover at a different path on the ESXi clusters if their files conflict with the existing files on the recovery ESXi cluster. For example, there is a file name conflict if a VM (VM1) migrates to a recovery cluster that already has a VM (VM1) in the same container.

   However, the entities recover at a different path with `VmRecoveredAtAlternatePath` alert only if the following conditions are met.

   - Prism Element running on both the primary and the recovery Nutanix clusters are of version 5.17 or newer.
   - A path for the entity recovery is not defined while initiating the failover operation.
   - The protected entities do not have shared disk/s.

   If these conditions are not satisfied, the failover operation fails.

   Note: To avoid conflicts when the primary site becomes active after the failover, shut down the guest VMs associated with this recovery plan. Manually power off the guest VMs on either primary or recovery site after the failover is complete. You can also block the guest VMs associated with this recovery plan through the firewall.

Monitoring a Failover Operation (Xi Leap)

After you trigger a failover operation, you can monitor failover-related tasks. To monitor a failover, do the following.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Recovery Plans** in the left pane.
3. Click the name of the recovery plan for which you triggered failover.
4. Click the **Tasks** tab.
   The left pane displays the overall status. The table in the details pane lists all the running tasks and their individual statuses.

UEFI and Secure Boot Support for CHDR

Nutanix supports CHDR migrations of guest VMs having UEFI and Secure Boot.

Table 41: Nutanix Software - Minimum Requirements

Nutanix Software	Minimum Supported Version
Minimum AOS	5.19.1
Minimum PC	pc.2021.1
Minimum NGT	2.1.1

Table 42: Applications and Operating Systems Requirements - UEFI

Operating Systems	Versions
Microsoft Windows	• Microsoft Windows 10 • Microsoft Windows Server 2016 • Microsoft Windows Server 2019
Linux	• CentOS Linux 7.3 • Ubuntu 18.04 • Red Hat Enterprise Linux Server versions 7.1 and 7.7

Table 43: Applications and Operating Systems Requirements - Secure Boot

Operating Systems	Versions
Microsoft Windows	• Microsoft Windows Server 2016 • Microsoft Windows Server 2019

Operating Systems	Versions
Linux	• CentOS Linux 7.3 • Red Hat Enterprise Linux Server versions 7.7

Table 44: Recovery Limitations

System Configuration	Limitation
Microsoft Windows Defender Credential Guard	VMs which have Credential Guard enabled cannot be recovered with CHDR recovery solution.
IDE + Secure Boot	VMs on ESXi which have IDE Disks or CD-ROM and Secure Boot enabled cannot be recovered on AHV.
UEFI VMs on CentOS 7.4, CentOS 7.5 and Ubuntu 16.04 may fail to boot after CHDR migration.	CentOS 7.4, CentOS 7.5 and Ubuntu 16.04 UEFI VMs do not boot after cross-hypervisor disaster recovery migrations. See KB-10633 for more information about this limitation. Contact Nutanix Support for assistance with this limitation.
UEFI VM may fail to boot after failback.	When a UEFI VM is booted on AHV for the first time, UEFI firmware settings of the VM are initialized. The next step is to perform a guest reboot or guest shutdown to fully flush the settings into persistent storage in the NVRAM. If this UEFI VM is failed over to an ESXi host without performing the guest reboot/ shutdown, the UEFI settings of the VM remain partial. Although the VM boots on ESXi, it fails to boot on AHV when a failback is performed. See KB-10631 for more information about this limitation. Contact Nutanix Support for assistance with this limitation.

Protection with NearSync Replication and DR (Xi Leap)

NearSync replication enables you to protect your data with an RPO of as low as 1 minute. You can configure a protection policy with NearSync replication by defining the VMs or VM categories. The policy creates a recovery point of the VMs in minutes (1–15 minutes) and replicates it to Xi Cloud Services. You can configure disaster recovery with Asynchronous replication between an on-prem AHV or ESXi clusters and Xi Cloud Services. You can also perform cross-hypervisor disaster recovery (CHDR)—disaster recovery of VMs from AHV clusters to ESXi clusters or of VMs from ESXi clusters to AHV clusters.

> Note: Nutanix provides multiple disaster recovery (DR) solutions to secure your environment. See Nutanix Disaster Recovery Solutions on page 11 for the detailed representation of the DR offerings of Nutanix.

The following are the advantages of NearSync replication.

- Protection for the mission-critical applications. Securing your data with minimal data loss if there is a disaster, and providing you with more granular control during the recovery process.
- No minimum network latency or distance requirements.

  > Note: However, a maximum of 75 ms network latency is allowed for replication between an AHV cluster and Xi Cloud Services.

- Low stun time for VMs with heavy I/O applications.

  Stun time is the time of application freeze when the recovery point is taken.

- Allows resolution to a disaster event in minutes.

To implement the NearSync feature, Nutanix has introduced a technology called lightweight snapshots (LWSs). LWS recovery points are created at the metadata level only, and they continuously replicate incoming data generated by workloads running on the active clusters. LWS recovery points are stored in the LWS store, which is allocated on the SSD tier. When you configure a protection policy with NearSync replication, the system allocates the LWS store automatically.

> Note: The maximum LWS store allocation for each node is 360 GB. For the hybrid systems, it is 7% of the SSD capacity on that node.

Transitioning in and out of NearSync

When you configure a protection policy with NearSync replication, the policy remains in an hourly schedule until its transition into NearSync is complete.

To transition into NearSync, initial seeding of the recovery site with the data is performed, the recovery points are taken on an hourly basis, and replicated to the recovery site. After the system determines that the recovery points containing the seeding data have replicated within a specified amount of time (default is an hour), the system automatically transitions the protection policy into NearSync depending on the bandwidth and the change rate. After you transition into NearSync, you can see the configured NearSync recovery points in the web interface.

The following are the characteristics of the process.

- Until you are transitioned into NearSync, you can see only the hourly recovery points in Prism Central.
- If for any reason, a VM transitions out of NearSync, the system raises alerts in the **Alerts** dashboard, and the protection policy transitions out to the hourly schedule. The system continuously tries to get to the NearSync schedule that you have configured. If the transition is successful, the protection policy automatically transitions back into NearSync, and alerts specific to this condition are raised in the **Alerts** dashboard.

To transition out of NearSync, you can do one of the following.

- Delete the protection policy with NearSync replication that you have configured.
- Update the protection policy with NearSync replication to use an hourly RPO.

- Unprotect the VMs.

> Note: There is no transitioning out of the protection policy with NearSync replication on the addition or deletion of a VM.

Repeated transitioning in and out of NearSync can occur because of the following reasons.

- LWS store usage is high.
- The change rate of data is high for the available bandwidth between the primary and the recovery sites.
- Internal processing of LWS recovery points is taking more time because the system is overloaded.

Retention Policy

Depending on the RPO (1–15 minutes), the system retains the recovery points for a specific amount of time. For protection policy with NearSync replication, you can configure the retention policy for days, weeks, or months on both the primary and recovery sites instead of defining the number of recovery points you want to retain. For example, if you desire an RPO of 1 minute and want to retain the recovery points for 5 days, the following retention policy is applied.

- For every 1 minute, a recovery point is created and retained for a maximum of 15 minutes.

> Note: The recent 15 recovery points are only visible in Prism Central and are available for the restore operation.

- For every hour, a recovery point is created and retained for 6 hours.
- One daily recovery point is created and retained for 5 days.

You can also define recovery point retention in weeks or months. For example, if you configure a 3-month schedule, the following retention policy is applied.

- For every 1 minute, a recovery point is created and retained for 15 minutes.
- For every hour, a recovery point is created and retained for 6 hours.
- One daily recovery point is created and retained for 7 days.
- One weekly recovery point is created and retained for 4 weeks.
- One monthly recovery point is created and retained for 3 months.

> Note:
>
> - You can define different retention policies on the primary and recovery sites.
> - The system retains subhourly and hourly recovery points for 15 minutes and 6 hours respectively. Maximum retention time for days, weeks, and months is 7 days, 4 weeks, and 12 months respectively.
> - If you change the protection policy configuration from hourly schedule to minutely schedule (Asynchronous to NearSync), the first recovery point is not created according to the new schedule. The recovery points are created according to the start time of the old hourly schedule (Asynchronous). If you want to get the maximum retention for the first recovery point after modifying the schedule, update the start time accordingly for NearSync.

NearSync Replication Requirements (Xi Leap)

The following are the specific requirements of configuring protection policies with NearSync replication schedule in Xi Leap. Ensure that you meet the following requirements in addition to the general requirements of Xi Leap.

For more information about the general requirements of Xi Leap, see Xi Leap Requirements on page 156.

For information about the on-prem node, disk and Foundation configurations required to support NearSync replication schedules, see On-Prem Hardware Resource Requirements on page 15.

Hypervisor Requirements

AHV or ESXi clusters running AOS 5.17 or newer, each registered to a different Prism Central

- The on-prem AHV clusters must be running on version 20190916.189 or newer.
- The on-prem ESXi clusters must be running on version ESXi 6.5 GA or newer.

Nutanix Software Requirements

The on-prem Prism Central and its registered clusters (Prism Elements) must be running the following versions of AOS.

- AOS 5.17 or newer with AHV.
- AOS 5.17 or newer with ESXi.

Cross Hypervisor Disaster Recovery (CHDR) Requirements

Data Protection with NearSync replication supports cross-hypervisor disaster recovery. You can configure disaster recovery to recover VMs from AHV clusters to ESXi clusters or VMs from ESXi clusters to AHV clusters by considering the following requirement of CHDR.

- The on-prem clusters are running AOS 5.18 or newer.
- Install and configure Nutanix Guest Tools (NGT) on all the guest VMs. For more information, see *Enabling and Mounting Nutanix Guest Tools* in *Prism Web Console Guide*.

  NGT configures the guest VMs with all the required drivers for VM portability. For more information about general NGT requirements, see *Nutanix Guest Tools Requirements and Limitations* in *Prism Web Console Guide*.
- CHDR supports guest VMs with flat files only.
- CHDR supports IDE/SCSI disks only.

  > Tip: From AOS 5.19.1, CHDR supports SATA disks also.

- For all the non-boot SCSI disks of Windows guest VMs, set the SAN policy to OnlineAll so that they come online automatically.
- In vSphere 6.7, guest VMs are configured with UEFI secure boot by default. Upon CHDR to an AHV cluster, these guest VMs do not start if the host does not support the UEFI secure boot feature. For more information about supportability of UEFI secure boot on Nutanix clusters, see https://portal.nutanix.com/page/documents/compatibility-interoperability-matrix/guestos.
- For information about operating systems that support UEFI and Secure Boot, see UEFI and Secure Boot Support for CHDR on page 227.

- Nutanix does not support vSphere inventory mapping (for example, VM folder and resource pools) when protecting workloads between VMware clusters.

- Nutanix does not support vSphere snapshots or delta disk files. If you have delta disks attached to a VM and you proceed with failover, you get a validation warning and the VM does not recover. Contact Nutanix Support for assistance.

  Note: CHDR does not preserve hypervisor-specific properties (for example, multi-writer flags, independent persistent and non-persistent disks, changed block tracking (CBT), PVSCSI disk configurations).

  Note: In vSphere 6.7, guest VMs are configured with EFI secure boot by default. Upon CHDR to AHV, these guest VMs will not start if the host does not support the UEFI secure boot feature. For more information about the supportability of UEFI secure boot on Nutanix clusters, see https://portal.nutanix.com/page/documents/compatibility-interoperability-matrix/guestos.

Table 45: Operating System Supported for CHDR

Operating System	Version	Requirements and Limitations
Windows	• Windows 2008 R2 or newer versions • Windows 7 or newer versions	• Only 64-bit operating systems are supported.
Linux	• CentOS 6.5 and 7.0 • RHEL 6.5 or newer and RHEL 7.0 or newer. • Oracle Linux 6.5 and 7.0 • Ubuntu 14.04	• SLES operating system is not supported.

Additional Requirements

- Both the primary and the recovery clusters must be of minimum three-nodes.
- See On-Prem Hardware Resource Requirements on page 15 for the on-prem hardware and Foundation configurations required to support NearSync replication schedules.
- Set the virtual IP address and the data services IP address in the primary and the recovery clusters.
- The recovery site container must have as much space as the protected VMs working size set of the primary site. For example, if you are protecting a VM that is using 30 GB of space on the container of the primary site, the same amount of space is required on the recovery site container.
- The bandwidth between the two sites must be approximately equal to or higher than the change rate of the protected VMs (maximum change rate is 20 MBps).

NearSync Replication Limitations (Xi Leap)

The following are the specific limitations of data protection with NearSync replication in Xi Leap. These limitations are in addition to the general limitations of Leap.

For information about the general limitations of Leap, see Xi Leap Limitations on page 159.

- Deduplication enabled on storage containers having VMs protected with NearSync lowers the replication speed.
- All files associated with the VMs running on ESXi must be located in the same folder as the VMX configuration file. The files not located in the same folder as the VMX configuration file might not recover on a recovery cluster. On recovery, the VM with such files fails to start with the following error message.`Operation failed: InternalTaskCreationFailure: Error creating host specific VM change power state task. Error: NoCompatibleHost: No host is compatible with the virtual machine`
- In CHDR, NearSync replication does not support retrieving recovery points from the recovery sites.

  For example, if you have 1 day retention at the primary site and 5 days retention at the recovery site, and you want to go back to a recovery point from 5 days ago. NearSync does not support replicating 5 days retention back from the recovery site to the primary site.

Creating a Protection Policy with NearSync Replication Schedule (Xi Leap)

Create a NearSync protection policy in the primary site Prism Central. The policy schedules recovery points of the protected VMs as per the set RPO and replicates them to Xi Cloud Services for availability. When creating a protection policy, you can specify only VM categories. If you want to include VMs individually, you must first create the protection policy—which can also include VM categories and then include the VMs individually in the protection policy from the **VMs** page.

Before you begin

Ensure that the AHV or ESXi clusters on both the primary and recovery site are NearSync capable. A cluster is NearSync capable if the capacity of each SSD in the cluster is at least 1.2 TB.

See NearSync Replication Requirements (Xi Leap) on page 231 and NearSync Replication Limitations (Xi Leap) on page 233 before you start.

About this task

To create a protection policy with NearSync replication in Xi Cloud Services, perform the following procedure.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Protection Policies** in the left pane.

3. Click **Create Protection Policy**.
   Specify the following information in the Create Protection Policy window.

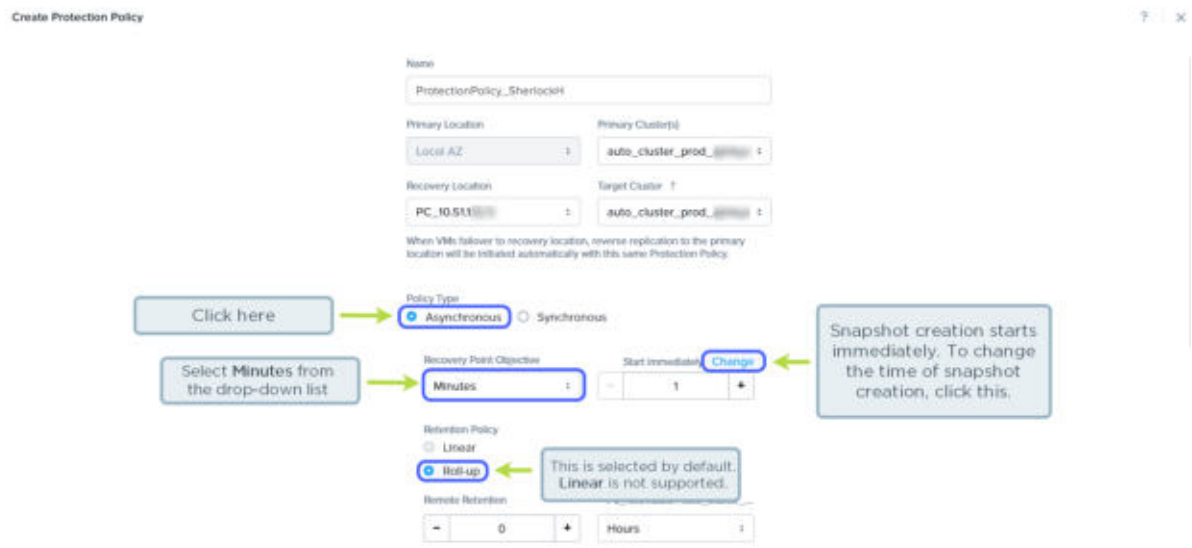


Figure 128: Protection Policy Configuration: NearSync

a. **Name**: Enter a name for the policy.

   Caution: The name can be of only alphanumeric, dot, dash, and underscore characters.

b. **Primary Location**: Select the primary availability zone that hosts the VMs to protect. This list displays the **Local AZ** by default and is unavailable for editing.

c. **Primary Cluster(s)**: Select the cluster that hosts the VMs to protect.

d. **Recovery Location**: Select the recovery availability zone where you want to replicate the recovery points.

   If you do not select a recovery location, the local recovery points that are created by this protection policy do not replicate automatically. You can, however, replicate recovery points manually and use recovery plans to recover the VMs. For more information, see Protection and Manual DR (Xi Leap) on page 244.

e. **Target Cluster**: Select the NearSync capable cluster where you want to replicate the recovery points.

   This field becomes available only if the recovery location is a physical remote site. If the specified recovery location is an availability zone in Xi Cloud Services, the **Target Cluster** field becomes unavailable because Xi Cloud Services selects a cluster for you. If the specified recovery location is a physical location, you can select a cluster of your choice.

   Caution: If the primary cluster contains an IBM Power Systems server, you cannot replicate recovery points to Xi Cloud Services. However, you can replicate recovery points to the

on-prem target cluster if the target on-prem cluster also contains an IBM Power Systems server.

Caution: Select **auto-select** from the drop-down list only if all the clusters at the recovery site are NearSync capable.

f. **Policy Type**: Click **Asynchronous**.

g. **Recovery Point Objective**: Specify the frequency in minutes (anywhere between 1-15 minutes) at which you want recovery points to be taken.

   By default, recovery point creation begins immediately after you create the protection policy. If you want to specify when recovery point creation must begin, click **Change**, and then, in the **Start Time** dialog box, do the following.

   Click **Start from specific point in time.**

   In the time picker, specify the time at which you want to start taking recovery points.

   Click **Save**.

   Tip: NearSync also allows you to recover the data of the minute just before the unplanned failover. For example, on a protection policy with 10 minute RPO, you can use the internal lightweight snapshots (LWS) to recover the data of the 9th minute when there is an unplanned failover.

h. **Retention Policy**: Specify the type of retention policy.

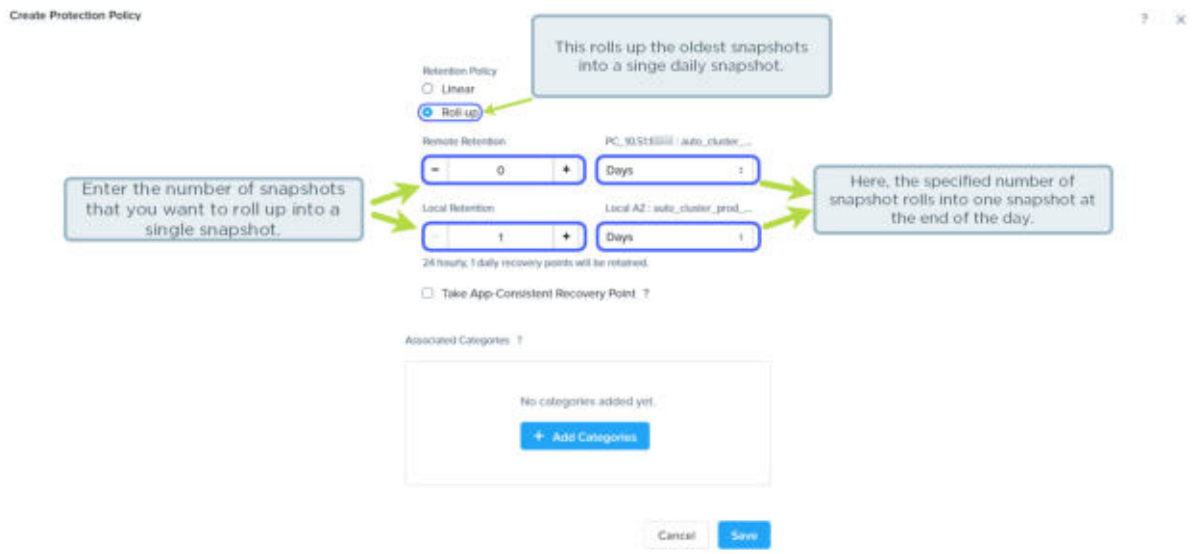


Figure 129: Roll-up Retention Policy

» **Roll-up**: Rolls up the recovery points as per the RPO and retention period into a single recovery point at a site. For example, if you set the RPO to 1 hour, and the retention time to 5 days, the 24 oldest hourly recovery points roll up into a single daily recovery point (one recovery point = 24 hourly recovery points) after every 24 hours.

The system keeps one day (of rolled-up hourly recovery points) and 4 days of daily recovery points.

> Note:
>
> - If the retention period is n days, the system keeps 1 day of RPO (rolled-up hourly recovery points) and n-1 days of daily recovery points.
> - If the retention period is n weeks, the system keeps 1 day of RPO, 1 week of daily and n-1 weeks of weekly recovery points.
> - If the retention period is n months, the system keeps 1 day of RPO, 1 week of daily, 1 month of weekly, and n-1 months of monthly recovery points.
> - If the retention period is n years, the system keeps 1 day of RPO, 1 week of daily, 1 month of weekly, and n-1 months of monthly recovery points.

> Note: The recovery points that are used to create a rolled-up recovery point are discarded.

> Tip: Use roll-up retention policies for anything with a longer retention period. Roll-up policies are more flexible and automatically handle recovery point aging/pruning while still providing granular RPOs for the first day.

> Note: NearSync does not support Linear retention policies. When you enter a minutely time unit in the **Recovery Point Objective**, the **Roll-up** retention policy is automatically selected.

4. To specify the retention number for the sites, do the following.

   a. **Remote Retention**: Specify the retention number for the remote site.

      This field is unavailable if you do not specify a recovery location.

   b. **Local Retention**: Specify the retention number for the local site.

      If you select linear retention, the remote and local retention count represents the number of recovery points to retain at any given time. If you select roll-up retention, these numbers specify the retention period.

5. If you want to take application consistent recovery points, select **Take App-Consistent Recovery Point**.

   Application consistent recovery points ensure that application consistency is maintained in the replicated recovery points. For application-consistent recovery points, install NGT on the VMs running on AHV. For VMs running on ESXi, you can take application-consistent recovery points without installing NGT, but the recovery points are hypervisor-based and lead to VM stuns (temporary unresponsive VMs).

   > Caution: Application-consistent recovery points fail for EFI-boot enabled Windows 2019 VM running on ESXi when NGT is not installed. Nutanix recommends installing NGT on VMs running on ESXi also.

6. **Associated Categories**: To protect categories of VMs, perform the following.

   > Tip: Before associating VM categories to a protection policy, determine how you want to identify the VMs you want to protect. If they have a common characteristic (for example, the VMs belong to a specific application or location), check the **Categories** page to ensure that both the category and the required value are available. Prism Central includes built-in categories for frequently encountered applications such as MS Exchange and Oracle. You can

also create your custom categories. If the category or value you want is not available, first create the category with the required values, or update an existing category so that it has the values that you require. Doing so ensures that the categories and values are available for selection when creating the protection policy. You can add VMs to the category either before or after you configure the protection policy. For more information about VM categories, see *Category Management* in the *Prism Central Guide*.

a. Click **Add Categories**.

b. Select the VM categories from the list to add to the protection policy.

> Note:
>
> You cannot protect a VM by using two or more protection policies. Therefore, VM categories specified in another protection policy are not listed here. Also, if you included a VM in another protection policy by specifying the category to which it belongs (category-based inclusion), and if you add the VM to this policy by using its name (individual inclusion), the individual inclusion supersedes the category-based inclusion. Effectively, the VM is protected only by this protection policy and not by the protection policy in which its category is specified.
>
> For example, the guest VM *VM_SherlockH* is in the category *Department:Admin*, and you add this category to the protection policy named *PP_AdminVMs*. Now, if you add *VM_SherlockH* from the **VMs** page to another protection policy named *PP_VMs_UK*, *VM_SherlockH* is protected in *PP_VMs_UK* and unprotected from *PP_AdminVMs*.

c. Click **Save**.

> Tip: To add or remove categories from the existing protection policy, click **Update**.

7. Click **Save**.
   You have successfully created a protection policy with NearSync replication in Xi Leap. You can add VMs individually (without VM categories) to the protection policy or remove VMs from the protection policy. For information about the operations that you can perform on a protection policy, see Protection Policy Management on page 237.

Creating a Recovery Plan (Xi Leap)

Create a recovery plan in the primary Prism Central. The procedure for creating a recovery plan is the same for all the data protection strategies in Xi Leap.

For more information about creating a recovery plan in Xi Leap, see Creating a Recovery Plan (Xi Leap) on page 209.

Protection Policy Management

A protection policy automates the creation and replication of recovery points. When configuring a protection policy for creating local recovery points, you specify the RPO, retention policy, and the VMs that you want to protect. You also specify the recovery location if you want to automate recovery point replication to Xi Cloud Services.

When you create, update, or delete a protection policy, it synchronizes to the paired Xi Cloud Services. The recovery points automatically start replicating in the reverse direction after you perform a failover at the recovery Xi Cloud Services. For information about how Xi Leap determines the list of availability zones for synchronization, see Entity Synchronization Between Paired Availability Zones on page 245.

> Note: A VM cannot be simultaneously protected by a protection domain and a protection policy. If you want to use a protection policy to protect a VM that is part of a protection domain, first

remove the VM from the protection domain, and then include it in the protection policy. For information, see Migrating Guest VMs from a Protection Domain to a Protection Policy on page 248

Adding Guest VMs Individually to a Protection Policy

You can also add VMs directly to a protection policy from the **VMs** page, without the use of a VM category. To add VMs directly to a protection policy in Xi Cloud Services, perform the following procedure.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **VMs** in the left pane.
3. Click **Protect** from the **Actions** drop-down menu.

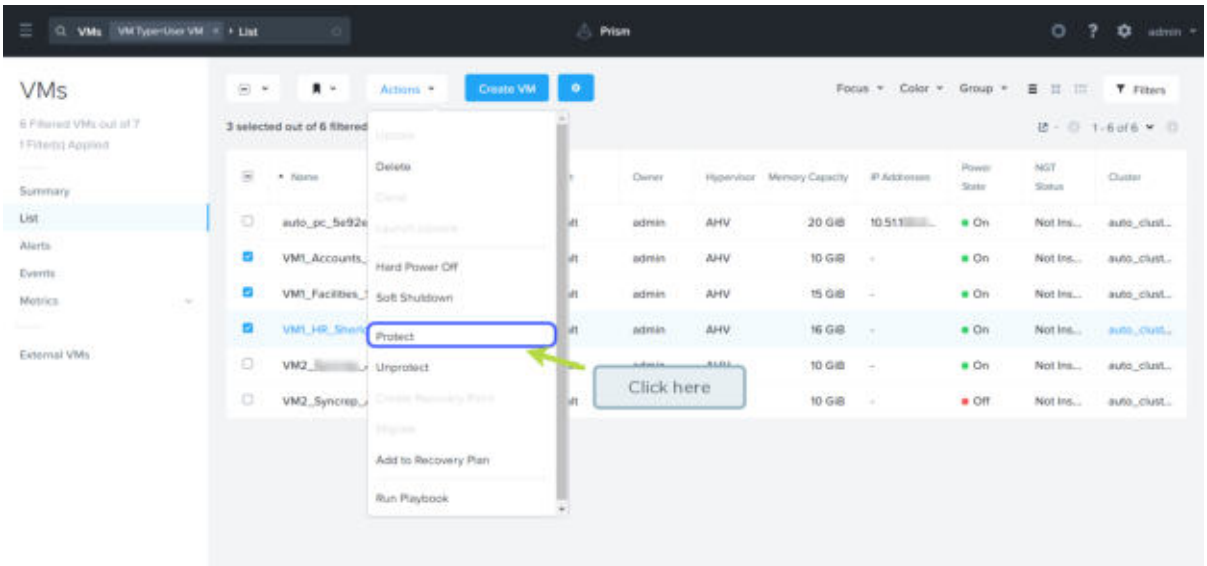


Figure 130: Protect VMs Individually

4. Select the protection policy in the table to include the VMs in a protection policy.

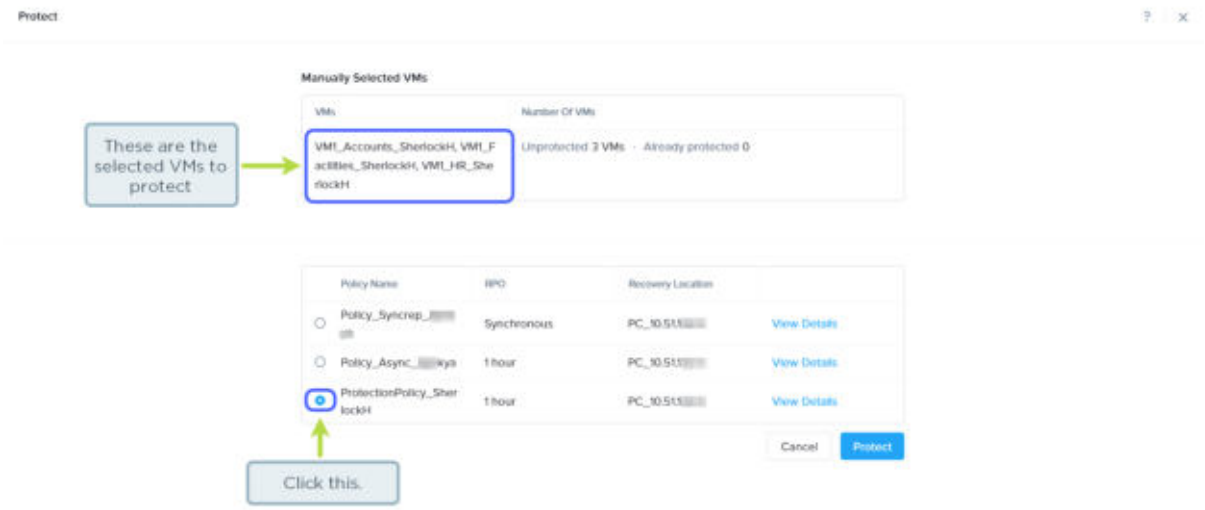


Figure 131: Protection Policy Selection

5. Click **Protect**.
   The VMs are added to the selected protection policy. The updated protection policy starts synchronizing to the recovery Prism Central.

Removing Guest VMs Individually from a Protection Policy

You can directly remove VMs from a protection policy from the **VMs** page. To remove VMs from a protection policy in Xi Cloud Services, perform the following procedure.

About this task

> Note: If a VM is protected individually (not through VM categories), you can remove it from the protection policy only by using this individual removal procedure.

> Note: If a VM has been protected under a VM category, you cannot remove the VM from the protection policy by using this procedure. You can remove the VM from the protection policy only by dissociating the VM from the category.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **VMs** in the left pane.
3. Select the VMs that you want to remove from a protection policy.
4. Click **UnProtect** from the **Actions** drop-down menu.
   The selected VMs are removed from the protection policy. The updated protection policy starts synchronizing to the recovery Prism Central.

Cloning a Protection Policy

If the requirements of the protection policy that you want to create are similar to an existing protection policy in Xi Cloud Services, you can clone the existing protection policy and update the clone.

About this task

To clone a protection policy from Xi Cloud Services, perform the following procedure.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Protection Policies** in the left pane.
3. Select the protection policy that you want to clone.
4. Click **Clone** from the **Actions** drop-down menu.
5. Make the required changes on the **Clone Protection Policy** page. For information about the fields on the page, see:
   - Creating a Protection Policy with Asynchronous Replication Schedule (Xi Leap) on page 198
   - Creating a Protection Policy with NearSync Replication Schedule (Xi Leap) on page 233
6. Click **Save**.
   The selected protection policy is cloned. The updated protection policy starts synchronizing to the recovery Prism Central.

Updating a Protection Policy

You can modify an existing protection policy in the Xi Cloud Services. To update an existing protection policy in Xi Cloud Services, perform the following procedure.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Protection Policies** in the left pane.
3. Select the protection policy that you want to update.
4. Click **Update** from the **Actions** drop-down menu.
5. Make the required changes on the **Update Protection Policy** page. For information about the fields on the page, see:
   - Creating a Protection Policy with Asynchronous Replication Schedule (Xi Leap) on page 198
   - Creating a Protection Policy with NearSync Replication Schedule (Xi Leap) on page 233
6. Click **Save**.
   The selected protection policy is updated. The updated protection policy starts synchronizing to the recovery Prism Central.

Finding the Protection Policy of a Guest VM

You can use the data protection focus on the VMs page to determine the protection policies to which a VM belongs in Xi Cloud Services. To determine the protection policy in Xi Cloud Services to which a VM belongs, do the following.

Procedure

1. Log on to Xi Cloud Services.

2. Click **Explore** and go to **VMs** in the left pane.

3. Click **Data Protection** from the **Focus** menu at the top-right corner.
   The **Protection Policy** column that is displayed shows the protection policy to which the VMs belong.

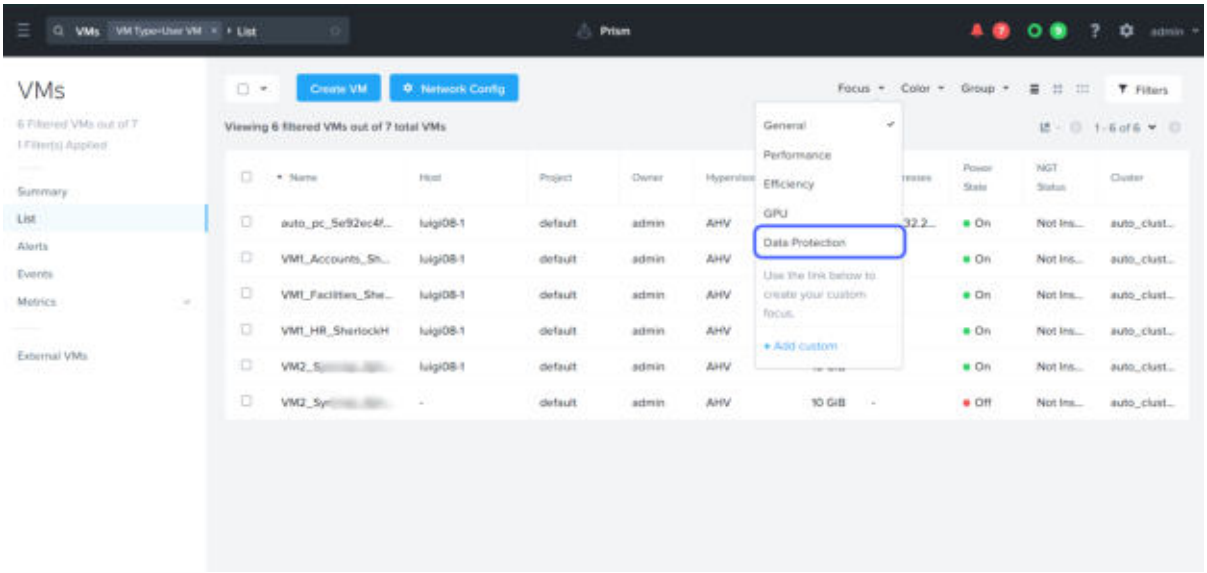


Figure 132: Focus

4. After you review the information, to return the VM page to the previous view, remove the **Focus Data Protection** filter from the filter text box.

Recovery Plan Management

A recovery plan orchestrates the recovery of protected VMs at a recovery site. Recovery plans are predefined procedures (runbooks) that use stages to enforce VM power-on sequence. You can also configure the inter-stage delays to recover applications gracefully. Recovery plans that recover applications in Xi Cloud Services are also capable of creating the required networks during failover and can assign public-facing IP addresses to VMs.

A recovery plan created in one availability zone (site) replicates to the paired availability zone and works bidirectionally. After a failover from the primary site to a recovery site, you can failback to the primary site by using the same recovery plan.

After you create a recovery plan, you can validate or test it to ensure that recovery goes through smoothly when failover becomes necessary. Xi Cloud Services includes a built-in VPC for validating or testing failover.

Recovery plans are independent of protection policies and do not reference protection policies in their configuration information. Also, they do not create recovery points. While the process of planned failover includes the creation of a recovery point so that the latest data can be used for recovery, unplanned and test failovers rely on the availability of the required recovery points at the designated recovery site. A recovery plan therefore requires the VMs in the recovery plan to also be associated with a protection policy.

Recovery plans are synchronized to one or more paired sites when they are created, updated, or deleted. For information about how Leap determines the list of availability zones (sites) for synchronization, see Entity Synchronization Between Paired Availability Zones on page 245.

Adding Guest VMs Individually to a Recovery Plan

You can also add VMs directly to a recovery plan in the **VMs** page, without the use of a category. To add VMs directly to a recovery plan in Xi Cloud Services, perform the following procedure.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **VMs** in the left pane.
3. Select the VMs that you want to add to a recovery plan.
4. Click **Add to Recovery Plan** from the **Actions** drop-down menu.
   The **Update Recovery Plan** page is displayed.
5. Select the recovery plan where you want to add the VMs in the **Add to Recovery Plan** dialog box.
6. Click **Add**.
   The **Update Recovery Plan** dialog box appears.
7. In the **General** tab, check **Recovery Plan Name**, **Recovery Plan Description**. Click **Next**.
8. In the **Power On Sequence** tab, add VMs to the stage. For more information, see Stage Management on page 68
9. Click **Next**.
10. In the **Network Settings** tab, update the network settings as required for the newly added VMs. For more information, see Creating a Recovery Plan (Xi Leap) on page 209.
11. Click **Done**.
    The VMs are added to the recovery plan.

Removing Guest VMs Individually from a Recovery Plan

You can also remove VMs directly from a recovery plan in Xi Cloud Services. To remove VMs directly from a protection policy, perform the following procedure.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **Recovery Plans** in the left pane.
3. Select the recovery plan from which you want to remove VM.
4. Click **Update** from the **Actions** drop-down menu.
   The **Update Recovery Plan** dialog box appears.
5. In the **General** tab, check **Recovery Plan Name**, **Recovery Plan Description**. Click **Next**.
6. In the **Power On Sequence** tab, select the VMs and click **More Actions** > **Remove**.

   > Note: You see **More Actions** in a stage only when one or more VMs in the stage are selected. When none of the VMs in the stage are selected, you see **Actions**.

7. Click **Next**.

8. In the **Network Settings** tab, update the network settings as required for the newly added VMs. For more information, see Stage Management on page 68.

9. Click **Done**.
   The VMs are removed from the selected recovery plan.

Updating a Recovery Plan

You can update an existing recovery plan in Xi Cloud Services. To update a recovery plan, perform the following procedure.

Procedure

1. Log on to Xi Cloud Services.

2. Click **Explore** and go to **Recovery Plans** in the left pane.

3. Select the recovery plan that you want to update.

4. Click **Update** from the **Actions** drop-down menu.
   The **Update Recovery Plan** dialog box appears.

5. Make the required changes to the recovery plan. For information about the various fields and options, see Creating a Recovery Plan (Xi Leap) on page 209.

6. Click **Done**.
   The selected recovery plan is updated.

Validating a Recovery Plan

You can validate a recovery plan from the recovery site. For example, if you perform the validation in the Xi Cloud Services (primary site being an on-prem site), Leap validates failover from the on-prem site to Xi Cloud Services. Recovery plan validation only reports warnings and errors. Failover is not performed. In this procedure, you need to specify which of the two paired sites you want to treat as the primary, and then select the other site as the secondary.

About this task

To validate a recovery plan, do the following.

Procedure

1. Log on to Xi Cloud Services.

2. Click **Explore** and go to **Recovery Plans** in the left pane.

3. Select the recovery plan that you want to validate.

4. Click **Validate** from the **Actions** drop-down menu.

5. In the **Validate Recovery Plan** dialog box, do the following.

   a. In **Primary Location**, select the primary location.

   b. In **Recovery Location**, select the recovery location.

   c. Click **Proceed**.
   The validation process lists any warnings and errors.

6. Click **Back**.
   A summary of the validation is displayed. You can close the dialog box.

7. To return to the detailed results of the validation, click the link in the **Validation Errors** column.
   The selected recovery plan is validated for its correct configuration.

Protection and Manual DR (Xi Leap)

Manual data protection involves manually creating recovery points, replicating recovery points, and recovering guest VMs at the recovery site. You can also automate some of these tasks. For example, the last step—that of manually recovering guest VMs at the recovery site—can be performed by a recovery plan while the underlying recovery point creation and replication can be performed by protection policies. Conversely, you can configure protection policies to automate recovery point creation and replication, and then recover guest VMs manually at the recovery site.

Creating Recovery Points Manually (Out-of-Band Snapshots)

About this task

To create recovery points manually in Xi Cloud Services, do the following.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **VMs** in the left pane.
3. Select the VMs for which you want to create a recovery point.
4. Click **Create Recovery Point** from the **Actions** drop-down menu.
5. To verify that the recovery point is created, click the name of the VM, click the **Recovery Points** tab, and verify that a recovery point is created.

Replicating Recovery Points Manually

You can manually replicate recovery points only from the site where the recovery points exist.

About this task

To replicate recovery points manually from Xi Cloud Service, do the following.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **VMs** in the left pane.
3. Click the VM whose recovery point you want to replicate, and then click **Recovery Points** in the left.
   The **Recovery Points** view lists all the recovery points of the VM.
4. Select the recovery points that you want to replicate.
5. Click **Replicate** from the **Actions** drop-down menu.

6. In the **Replicate** dialog box, do the following.
   a. In **Recovery Location**, select the location where you want to replicate the recovery point.
   b. In **Target Cluster**, select the cluster where you want to replicate the recovery point.
   c. Click **Replicate Recovery Point**.

Recovering a Guest VM from a Recovery Point Manually

You can recover a VM by cloning a VM from a recovery point.

About this task

To recover a VM from a recovery point at Xi Cloud Services, do the following.

Procedure

1. Log on to Xi Cloud Services.
2. Click **Explore** and go to **VMs** in the left pane.
3. Click the VM whose recovery point you want to replicate, and then click **Recovery Points** in the left.
   The **Recovery Points** view lists all the recovery points of the VM.
4. Select the recovery point from which you want to recover the VM.
5. Click **Restore** from the **Actions** drop-down menu.
6. In the **Restore** dialog box, do the following.
   a. In the text box provided for specifying a name for the VM, specify a new name or do nothing to use the automatically generated name.
   b. Click **Restore**.

> Warning: The following are the limitations of the manually recovered VMs (VMs recovered without the use of a recovery plan).
>
> - The VMs recover without a VNIC if the recovery is performed at the remote site.
> - VM categories are not applied.
> - NGT needs be reconfigured.

Entity Synchronization Between Paired Availability Zones

When paired with each other, availability zones (sites) synchronize disaster recovery configuration entities. Paired sites synchronize the following disaster recovery configuration entities.

Protection Policies

A protection policy is synchronized whenever you create, update, or delete the protection policy.

Recovery Plans

A recovery plan is synchronized whenever you create, update, or delete the recovery plan. The list of availability zones (sites) to which Xi Leap must synchronize a recovery

plan is derived from the VMs that are included in the recovery plan. The VMs used to derive the availability zone list are VM categories and individually added VMs.

If you specify VM categories in a recovery plan, Leap determines which protection policies use those VM categories, and then synchronizes the recovery plans to the availability zones specified in those Protection Plans.

If you include VMs individually in a recovery plan, Leap uses the recovery points of those VMs to determine which protection policies created those recovery points, and then synchronizes the recovery plans to the availability zones specified in those protection policies. If you create a recovery plan for VM categories or VMs that are not associated with a protection policy, Leap cannot determine the availability zone list and therefore cannot synchronize the recovery plan. If a recovery plan includes only individually added VMs and a protection policy associated with a VM has not yet created VM recovery points, Leap cannot synchronize the recovery plan to the availability zone specified in that protection policy. However, recovery plans are monitored every 15 minutes for the availability of recovery points that can help derive availability zone information. When recovery points become available, Xi Leap derives the availability zone by the process described earlier and synchronizes the recovery plan to the availability zone.

VM Categories used in Protection Policies and Recovery Plans

A VM category is synchronized when you specify the VM category in a protection policy or recovery plan.

Issues such as a loss of network connectivity between paired availability zones or user actions such as unpairing of availability zones followed by repairing of those availability zones can affect VM synchronization.

> Tip: Nutanix recommends to unprotect all the VMs on the availability zone before unpairing it to avoid getting into a state where the entities have stale configurations after repairing of availability zones.

If you update VMs in either or both availability zones before such issues are resolved or before unpaired availability zones are paired again, VM synchronization is not possible. Also, during VM synchronization, if a VM cannot be synchronized because of an update failure or conflict (for example, you updated the same VM in both availability zones during a network connectivity issue), no further VMs are synchronized. Entity synchronization can resume only after you resolve the error or conflict. To resolve a conflict, use the **Entity Sync** option, which is available in the web console. Force synchronization from the availability zone that has the desired configuration. Forced synchronization overwrites conflicting configurations in the paired availability zone.

> Note: Forced synchronization cannot resolve errors arising from conflicting values in VM specifications (for example, the paired availability zone already has a VM with the same name).

If you do not update entities before a connectivity issue is resolved or before you pair the availability zones again, the synchronization behavior described earlier resumes. Also, pairing previously unpaired availability zones trigger an automatic synchronization event. For recommendations to avoid facing such issues, see Entity Synchronization Recommendations (Xi Leap) on page 246.

Entity Synchronization Recommendations (Xi Leap)

Consider the following recommendations to avoid inconsistencies and the resulting synchronization issues.

- During network connectivity issues, do not update entities at both the availability zones (sites) in a pair. You can safely make updates at any one site. After the connectivity issue is

resolved, force synchronization from the site in which you made updates. Failure to adhere to this recommendation results in synchronization failures.

You can safely create entities at either or both the sites as long as you do not assign the same name to entities at the two sites. After the connectivity issue is resolved, force synchronization from the site where you created entities.

- If one of the sites becomes unavailable, or if any service in the paired site is down perform force synchronization from the paired availability zone after the issue is resolved.

Forcing Entity Synchronization (Xi Leap)

Entity synchronization, when forced from an availability zone (site), overwrites the corresponding entities in paired sites. Forced synchronization also creates, updates, and removes those entities from paired sites.

About this task

The availability zone (site) to which a particular entity is forcefully synchronized depends on which site requires the entity (seeEntity Synchronization Between Paired Availability Zones on page 245). To avoid inadvertently overwriting required entities, ensure to force VM synchronization from the site in which the entities have the desired configuration.

If a site is paired with two or more availability zones (sites), you cannot select one or more sites with which to synchronize entities.

To force entity synchronization from Xi Cloud Services, do the following.

Procedure

1. Log on to Xi Cloud Services.
2. Click the settings button (gear icon) at the top-right corner of the window.
3. Click **Entity Sync** in the menu.
4. In the **Entity Sync** dialog box, review the message at the top of the dialog box, and then do the following.
   a. To review the list of entities that will be synchronized to an **AVAILABILITY ZONE**, click the number of **ENTITIES** adjacent to an availability zone.
   b. After you review the list of entities, click **Back**.
5. Click **Sync Entities**.

4

MIGRATING GUEST VMS FROM A PROTECTION DOMAIN TO A PROTECTION POLICY

You can protect a guest VM either with a protection domain in Prism Element or with a protection policy in Prism Central. If you have guest VMs in protection domains, migrate those guest VMs to protection policies to orchestrate their disaster recovery using Leap.

Before you begin

Migration from protection domains to protection policies is a disruption process. For successful migration,

- Ensure that the guest VMs have no on-going replication.
- Ensure that the guest VMs do not have volume groups.
- Ensure that the guest VMs are not in consistency groups.

About this task

To migrate a guest VM from a protection domain to a protection policy manually, perform the following procedure.

> Tip: To automate the migration using a script, refer *KB 10323*.

Procedure

1. Unprotect the guest VM from the protection domain.

   > Caution: Do not delete the guest VM snapshots in the protection domain. Prism Central reads those guest VM snapshots to generate new recovery points without full replication between the primary and recovery Nutanix clusters. If you delete the guest VM snapshots, the VM data replicates afresh (full replication). Nutanix recommends keeping the VM snapshots in the protection domain until the first recovery point for the guest VM is available on Prism Central.

   > Caution: Use the automated script for migrating guest VMs from a large protection domain. A large protection domain consists of more than 500 guest VMs. If you migrate the guest VMs manually from a large protection domain, the VM data replicates afresh (full replication).

2. Log on to Prism Central and protect the guest VMs with protection policies individually (see Adding Guest VMs Individually to a Protection Policy on page 136) or through VM categories.

COPYRIGHT

Copyright 2022 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

License

The provision of this software to you does not grant any licenses or other rights under any Microsoft patents with respect to anything other than the file server implementation portion of the binaries for this software, including no licenses or any other rights in any hardware or any devices or software that are used to communicate with or in connection with this software.

Conventions

Convention	Description
`variable_value`	The action depends on a value that is unique to your environment.
`ncli> command`	The commands are executed in the Nutanix nCLI.
`user@host$ command`	The commands are executed as a non-privileged user (such as nutanix) in the system shell.
`root@host# command`	The commands are executed as the root user in the vSphere or Acropolis host shell.
`> command`	The commands are executed in the Hyper-V host shell.
`output`	The information is displayed as output from a command or in a log file.

Version

Last modified: March 17, 2022 (2022-03-17T10:36:34+05:30)

v3.0 | April 2021 | RA-PC21

REFERENCE ARCHITECTURE

Nutanix Hybrid Cloud Reference Architecture

NUTANIX
YOUR ENTERPRISE CLOUD

Copyright

Copyright 2021 Nutanix, Inc.

Nutanix, Inc.
1740 Technology Drive, Suite 150
San Jose, CA 95110

All rights reserved. This product is protected by U.S. and international copyright and intellectual property laws. Nutanix and the Nutanix logo are registered trademarks of Nutanix, Inc. in the United States and/or other jurisdictions. All other brand and product names mentioned herein are for identification purposes only and may be trademarks of their respective holders.

Contents

1. Introduction....5
   - Vision for Private, Hybrid and Multicloud....5
   - Design Objectives....7
   - Audience....11
   - Design Decisions....11
   - How To Use This Reference Architecture?....11

2. Architecture Overview....13
   - Physical Layer....13
   - Virtualization Layer....18
   - Management Layer....19
   - Business Continuity Layer....20
   - Automation Layer....22
   - Security and Compliance Layer....23

3. High-Level Design Considerations....27
   - Choosing A Datacenter Architecture....27
   - Choosing a Hypervisor....35
   - Choosing a Cluster Deployment Model....38
   - Choosing How You Will Scale....43
   - Choosing a Control Plane....43
   - Choosing the Right Licensing and Support....47
   - Choosing the Right Hardware for Your Workload Types....52

4. Detailed Technical Design....56
   - Required Software Versions....56
   - Physical Layer Design....57
   - Virtualization Layer Design....121
   - Management Layer Design....128
   - Security Layer Design....135
   - Automation Layer Design....149
   - Operations Design....154

5. Multi-Datacenter Design for Nutanix Core and BC/DR....161
   - Choosing the Right BC/DR Solution....161

Multi-Datacenter Design..... 163
Replication Options.....165
Disaster Recovery Decision Tree..... 175
ROBO Patterns.....188
Control Planes.....191

6. Incorporating Optional Nutanix Products and Services.....206
Prism Operations..... 207
Nutanix Clusters for Hybrid and Multicloud Operations.....215
Nutanix Flow for Microsegmentation.....226
Nutanix Files for Integrated File Services..... 233
Nutanix Objects for Integrated Object Storage..... 246
Nutanix Mine for Integrated Backup..... 258
Calm Application Orchestration..... 263
Karbon for Kubernetes and Container Support on Nutanix.....273
Era for Database as a Service.....282

7. Conclusion.....293

Appendix..... 294
Table of Design Decisions..... 294

List of Figures.....305

List of Tables.....310

1. Introduction

Vision for Private, Hybrid and Multicloud

The Nutanix vision for cloud computing environments started in the datacenter with innovative solutions for private cloud, which greatly reduced the complexity and effort required to deploy and manage software-defined storage, compute, and networking infrastructure. In recent years, Nutanix's vision has expanded to include architectures for hybrid and multicloud that offer more alternatives to optimize costs.

Which type of cloud to use for a specific use case depends on a variety of characteristics of multi-tiered applications. These can be summarized in terms of:

- Use Case: web servers, application servers, databases, etc.
- Workload Type: virtual machines (VMs) and/or containers.
- Storage Format: block, object, file.
- Security Policies: requirements that could rule out the option to host workloads on particular public clouds.
- Unique Requirements: items like machine learning hardware or media transcoding that can only be met by the unique features of one specific Public Cloud offering (for example serverless computing, analytics, PaaS, and so on).

These characteristics ultimately shape an organization's cloud architectures and drive the decisions regarding where to host each application tier in a particular service.

Historically, the hosting architecture for multi-tiered applications didn't change without a labor-intensive migration. Nutanix innovation gives IT organizations the flexibility to dynamically place workloads. The Nutanix distributed architecture natively includes the ability to provision the same application blueprints in four different cloud configurations:

© 2021 Nutanix, Inc. All rights reserved

Table 1: Cloud Configurations

Cloud Configuration	Description
Private	Hosted on prem on AHV and/or VMware.
Hybrid	Some hosted on prem on AHV and/or vSphere and some hosted in AWS, Azure, or GCP.
Public (natively)	Hosted in AWS, Azure, or GCP as a native public cloud service offering.
Public (bare metal)	Hosted as bare metal in AWS running Nutanix Acropolis Operating System (AOS).

Nutanix further extends these cloud configurations by enabling the following four architectural principles:

Table 2: Architectural Principles

Architectural Principle	Description
Application Mobility	Multi-tiered applications can be aligned across all cloud providers to maximize architectural symmetry and promote application mobility.
Increased Standardization	Use of the same hardened OS gold image and business process orchestration (e.g. ITSM integration, approvals, emails, showback/ chargeback, etc.) across all cloud configurations.
Policy-Based Security Governance	Hosted in AWS, Azure, or GCP as a native public cloud service offering.
Policy-Based Cost Governance	Optimize cost by hosting workloads on the platform that meets the requirements of the service and has the lowest Total Cost of Ownership (TCO).

© 2021 Nutanix, Inc. All rights reserved

All of these principles contribute to an advanced hybrid and multicloud strategy that simplifies IT, stretches budgets, and accelerates time to value. This document provides the architecture and design-driven decisions to help our customers realize this strategic vision.

Figure 1: Nutanix Hybrid and Multicloud Strategy

Design Objectives

The objective of this document is to define, explore, and develop key design decisions required when implementing private, hybrid, or multicloud solutions based on the Nutanix platform. The objective can be further broken down as follows:

- To identify and enumerate the key design decisions that need to be documented in order to support a robust design methodology and practice.
- To explore each design decision, evaluating key viable options, tools, and methods for implementation and management so that organizations can make informed decisions relating to their specific design requirements.

Since simplicity is a key principle of all Nutanix products, some requirements may be met through native platform architecture without the need for superfluous design decisions, which are often required by competing platforms. The objective of this document is therefore not to educate readers about Nutanix features and functions, even though this may naturally occur as a side

© 2021 Nutanix, Inc. All rights reserved

benefit. In these cases, this document will explain how these requirements are addressed by Nutanix-native features.

The design objectives of this document for the Hybrid Cloud implementations are as follows:

Table 3: Design Objectives of this Document for Private Cloud Implementations

Design Objectives	Description
Key objective	The platform is capable of hosting and provisioning workloads.
Workload Types	Virtual Machines and containers
Scope of deployment	Greenfield deployment that is adaptable to brownfield deployments with workload migrations.
Cloud type	Hybrid cloud
Number of regions	• Two regions. Region 1 provides two availability zones for resiliency. Region 2 provides long-term backup retention for recovery from disasters impacting all of Region 1. • Services must be available across two availability zones during normal production. Services must be available within one availability zone during disaster recovery. • A minimum distance of 300km between Region 1 (primary) and Region 2 (DR). • Provide running workloads and disaster recovery services in same region.

© 2021 Nutanix, Inc. All rights reserved

Design Objectives	Description
Availability	• SLA 24/7 • Uptime 99.999% across availability zones › Applies to certain production workloads › Scheduled downtime not included › 99% uptime required during disaster recovery scenario • RPO – 0 min across availability zones › Applies to certain production workloads › RPO = 0 can be achieved at the application layer for some services but the infrastructure layer must support RPO = 0 • RTO – 5 min across availability zones. › Applies to certain production workloads
Disaster Recovery	• DR RPO = 60 min between primary availability zones and DR availability zone • DR RTO = 24h between primary availability zones and DR availability zone • Best effort RTO in the region where production workloads are running and the disaster recovery capabilities fails.
Minimum number of nodes per cluster	3 nodes.
Maximum number of workloads	Unlimited number of workloads dependent on the pod based constructs described herein.
Types of Clusters	Management, Workload, Storage Heavy, Edge Clusters.
Virtualization	• Nutanix AHV and VMware ESXi hypervisors. • While Nutanix supports Microsoft Hyper-V, it is not considered in this design.

© 2021 Nutanix, Inc. All rights reserved

Design Objectives	Description
Management Plane	Nutanix Prism Element, Nutanix Prism Central, and VMware vCenter.
Scope	• Sizing recommendations and methodology for the amount of software-defined storage, compute and networking. • Physical implementation of storage, compute, and networking. • Logical configuration of clusters. • Scalability methods and recommendations. • Automation relating to cluster build, expansion, and lifecycle management. • Management and operations aspects such as capacity management, reporting, upgrades, backup and restore, monitoring and alerting, and logging.
Authentication, authorization, and access control	Microsoft Active Directory users and groups tied to Role Based Access Control.
Security Policy & Enforcement	• Least privilege access policies will be implemented so that end-users and administrators will need to be members of groups in order to be able to perform secure aspects of their job function. • Certificates are signed by a trusted certificate authority (CA). • Hardening of platform associated with hypervisor, control plane, and data plane. • Security policy definition and enforcement of drift away from defined policies and checksum verification. • Separation of traffic classes such as management and application.

© 2021 Nutanix, Inc. All rights reserved

Audience

This document for Multi Datacenter design is intended for infrastructure architects, infrastructure administrators, and infrastructure operators who want to deploy and manage datacenters based on Nutanix Enterprise Cloud, to address requirements for availability, capacity, performance, scalability, business continuity, and disaster recovery.

Design Decisions

This document makes recommendations and guides the reader to appropriate decisions where possible. In cases where a design decision is required, this document describes possible options.

Table 4: Design Decision Example

NET-001	Title Of Design Decision
Justification	Justification to support why the decision was made.
Implication	Additional implications as a result of the design decision.

Note: Appendix I: Table of Design Decisions includes a list of all the design decisions described throughout this document.

How To Use This Reference Architecture?

This document is subdivided into five major sections as follows:

Architectural Overview

Introduces key Architecture concepts that will be discussed throughout this design.

Design Considerations

Discusses key design considerations that will vary for each customer. Customers will be required to make decisions that will influence the design and build of their end-solution.

Detailed Design

Identifies key design decisions and in most cases determines the optimal configuration and decision that will be used for validating the design. For design decision, alternate options may be discussed along with pros and cons. For good reasons, Customers may choose to deviate from the decisions made in this document. The decisions made in this section are recommended by Nutanix, however they are by no means meant to be the only method, and Nutanix recognizes that alternate decisions may be appropriate depending on specific requirements.

Multi-site Design, Disaster Recovery and Business Continuity

Articulates considerations for designing across multiple data centers, and the related DR/BC architecture.

> Note: This section does not provide detailed operational guides or runbooks.

Incorporating Optional Nutanix Services

Describes the design considerations for additional services that can optionally be deployed in a Nutanix environment to address specific requirements. The list of capabilities covered includes: management automation, Nutanix deployment in public cloud, advanced security capabilities, file and object data services, integrated backup options, application orchestration, container support, and database-as-a-service (DBaaS).

© 2021 Nutanix, Inc. All rights reserved

2. Architecture Overview

This section describes the high-level Nutanix architecture, including major concepts and design elements that anyone designing a Nutanix deployment should understand. If you are already familiar with Nutanix hyperconverged infrastructure (HCI) and Nutanix software, you can skip this section.

The diagram below shows the high-level architecture covered in this document. This overview explains the elements in each layer. Later sections will explore the design decisions necessary for each layer.

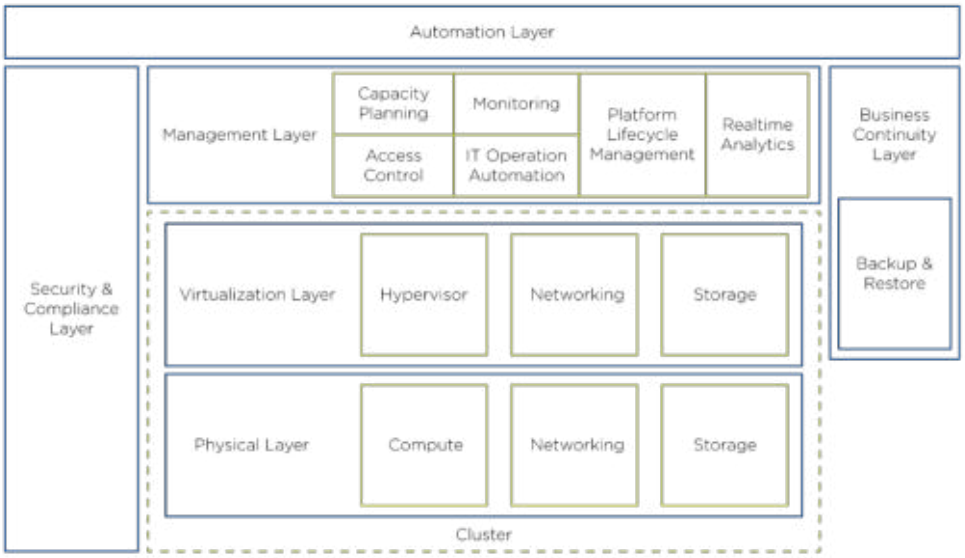


Figure 2: Architecture Overview

Physical Layer

Because the Nutanix Enterprise Cloud architecture is based on hyperconverged infrastructure, the physical layer is significantly different than it would be in a traditional datacenter architecture. Understanding the differences will allow you to make the best hardware choices for your Nutanix deployment.

© 2021 Nutanix, Inc. All rights reserved

Hyperconverged Infrastructure

Nutanix HCI converges the datacenter stack including compute, storage, storage networking, and virtualization, replacing the separate servers, storage systems, and storage area networks (SANs) found in conventional datacenter architectures and reducing complexity. Each node in a Nutanix cluster includes compute, memory, and storage, and nodes are pooled into a cluster. The Nutanix Acropolis Operating System (AOS) software running on each node pools storage across nodes and distributes operating functions across all nodes in the cluster for performance, scalability, and resilience.

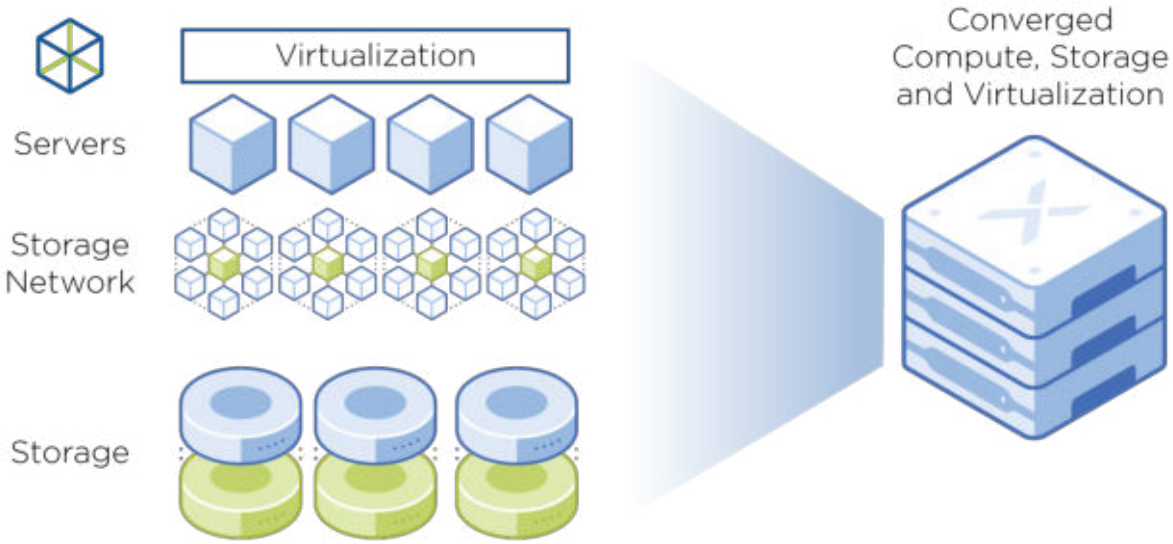


Figure 3: Hyperconverged Infrastructure

A Nutanix node runs an industry-standard hypervisor and the Nutanix Controller VM (CVM). The Nutanix CVM provides the software intelligence for the platform and is responsible for serving IO to running VMs.

© 2021 Nutanix, Inc. All rights reserved

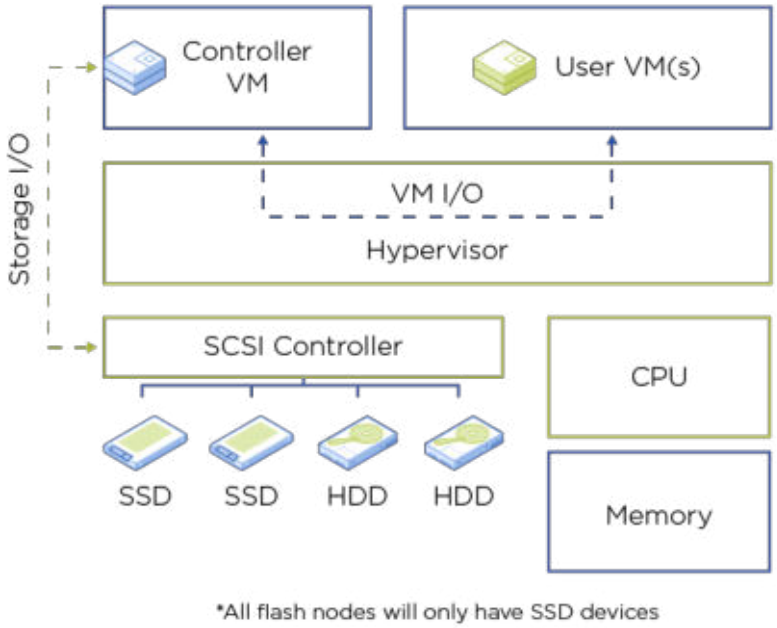


Figure 4: A Nutanix Node

Hardware Choice

Nutanix Enterprise Cloud provides significant choice when it comes to hardware platform selection. Available options include:

- Nutanix NX appliances.
- OEM appliances from leading vendors such as Dell, Lenovo, HPE, IBM, and Fujitsu.
- Other third-party servers from a wide range of vendors.

The Nutanix Support Portal contains the most up-to-date information on supported systems.

Hardware is available in a variety of chassis configurations from various vendors. Options range from multi-node chassis for high density to single-node rackmount chassis.

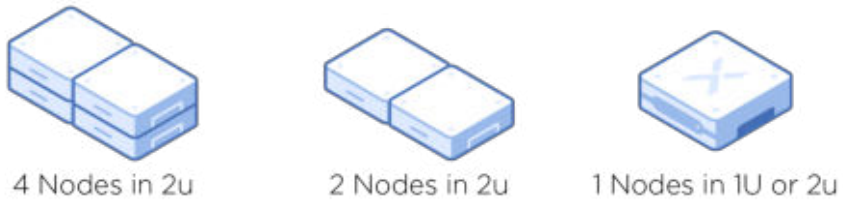


Figure 5: Chassis Configurations

Nutanix commonly refers to all of these chassis configurations as a block.

Compute

Sizing systems to meet compute needs in a Nutanix environment is similar to other architectures. However, it's important to ensure that your design provides enough compute (CPU/RAM) to support the CVM.

Storage

Nutanix nodes offer a range of storage configurations:

- Hybrid nodes combine flash SSDs for performance and HDDs for capacity.
- All-flash nodes utilize traditional flash SSDs.
- NVMe nodes utilize NVMe SSDs.

Different node types can be mixed in the same cluster. More information is provided in the document Product Mixing Restrictions.

For data resiliency, Nutanix uses replication factor (RF), maintaining 2 or 3 data copies. This approach enables a Nutanix cluster to be self-healing in the event of a drive, node, block, or rack failure. In a Nutanix cluster consisting of multiple blocks, RF can enable block awareness. Data copies are distributed across blocks to protect against the failure of an entire block. In configurations spanning multiple racks, RF can similarly provide rack awareness with resilience to a rack outage. For more information on Nutanix data resiliency, please refer to the Nutanix Bible.

Compression, deduplication and erasure coding (EC-X) can be enabled to increase data efficiency and save capacity.

Data locality and intelligent tiering ensure that the data associated with a VM is preferentially stored on that VM's local node. Active data is stored on the fastest media, delivering performance and eliminating the need for ongoing performance tuning.

Networking

Fast, low-latency and highly available networking is a key element of this document. The distributed storage architecture relies on the performance and resilience of the physical network. A good design provides high performance while maintaining simplicity.

In the detailed design section of this document we address common network topologies, selection of physical switches, and recommended connections between hosts and the physical network.

Cluster Design

A Nutanix cluster is the management boundary of the storage provided to a group of workloads. A Nutanix deployment can be architected to support either (a) mixed workloads in a single Nutanix cluster; or (b) dedicated clusters for each workload type in a block and pod design.

Designs that choose dedicated clusters may include any of the following cluster types:

Management clusters

Designed to run VMs that support datacenter management such as:

- Nutanix Prism Central.
- VMware vCenter.
- Active Directory Domain Controllers.
- Other management workloads, such as DNS, DHCP, NTP, Syslog.

Management clusters reside in the management workload domain. In this document, a management cluster occupies a separate rack or is distributed across multiple racks.

Workload clusters

© 2021 Nutanix, Inc. All rights reserved

Reside in a virtual infrastructure workload domain and run tenant virtual machines. You can mix different types of compute clusters and provide separate compute pools to address varying SLAs for availability and performance.

Storage clusters

Storage-only clusters provide dedicated data services to tenants. These are typically deployed for use cases focused on Object, File, or Block-level storage.

Edge/ROBO clusters

Reside at an edge and/or ROBO deployment to run virtual machines or ROBO workloads. These are typically distinguished from normal workload clusters by their small size and limited external bandwidth.

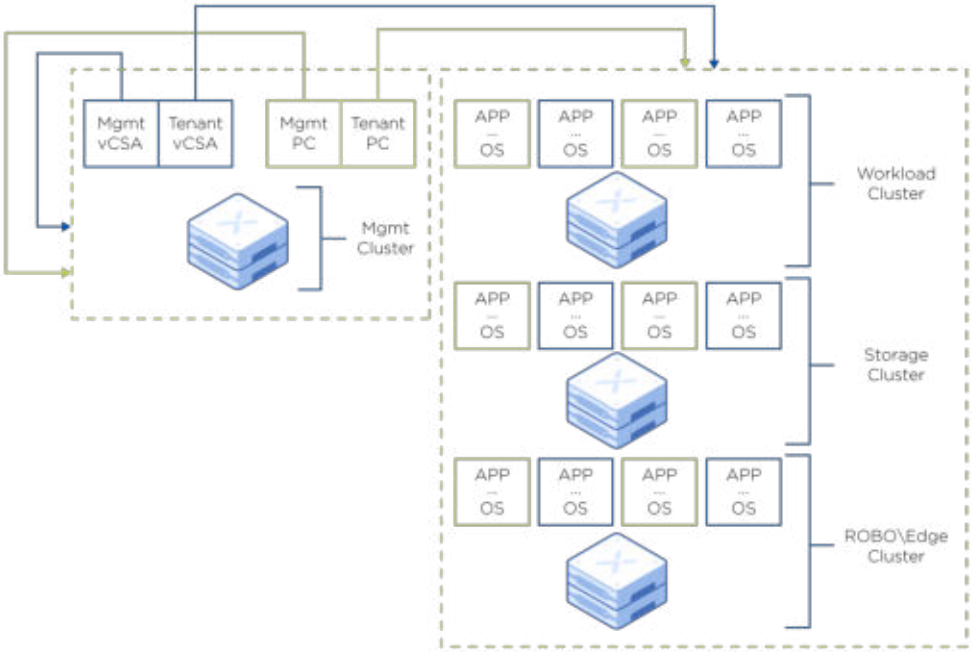


Figure 6: Cluster Design

Virtualization Layer

The virtualization layer sits logically above the physical layer, controlling access to compute, network, and storage resources. This document provides a choice

between two hypervisors: Nutanix AHV and VMware ESXi. Both are enterprise-grade hypervisors, filling a similar set of requirements and use cases. Nutanix AHV is included at no additional cost with every Nutanix node.

> Note: Nutanix HCI supports Microsoft Hyper-V, however this document does not describe deployment of this hypervisor.

Management Layer

The management layer is a key differentiator for Nutanix.

Nutanix Prism provides simplified end-to-end management for Nutanix environments. Prism combines multiple aspects of datacenter management into a single consumer-grade design that provides complete infrastructure and virtualization management, operational insights, and troubleshooting. Prism largely eliminates the need for separate management tools. All Prism functionality is also accessible via REST API.

The Prism family consists of three products that extend core capabilities:

Prism Element

The core Nutanix management platform enables management and monitoring at the cluster level for all infrastructure (compute, storage, networks) and virtualization. Key functionality of Prism includes:

- Full VM, storage, and hypervisor management.
- Network visualization.
- Role-based access control (RBAC).
- Nutanix 1-click upgrades. Prism orchestrates and streamlines platform upgrades, keeping track of the changes. Can upgrade all Nutanix software and firmware running in a Nutanix environment plus the ESXi hypervisor.

Prism Central

Enables management and monitoring of multiple Nutanix Prism Element clusters from a central interface.

Prism Pro and Ultimate

© 2021 Nutanix, Inc. All rights reserved

Adds advanced capabilities to the Prism platform, including performance anomaly detection, capacity planning, custom dashboards, reporting, advanced search capabilities, and task automation.

The Prism family of products are an integral part of a Nutanix cluster and do not require separate infrastructure. Prism Central runs as a separate VM, or as a cluster of 3 VMs for additional scale and resilience. More details on Nutanix management are provided later. Nutanix deployments that use the AHV hypervisor can be fully managed by Prism.

Nutanix deployments that use VMware vSphere should also include VMware vCenter Server. This is the centralized monitoring and resource management software for VMware virtual infrastructure. It performs a number of tasks, including resource provisioning and allocation, performance monitoring, workflow automation, and user privilege management.

Automated IT Operations

Prism Pro allows administrators to automate routine operational tasks, reducing administrator effort and time while increasing the quality of results. To provide this automation, Nutanix X-Play enables “if-this-then that” (IFTT) features that allow admins to create Playbooks that define automation actions that run when a particular trigger occurs.

The most common type of trigger is alert-based, where a system-defined or user- defined alert causes an action to occur. An alert could be something as simple as crossing a designated CPU or memory threshold. Other triggers can be manual; the associated playbook does not take action until an admin explicitly tells it to. With a manual trigger, an admin selects an entity such as a VM, and the specified playbook executes against it. Manual triggers allow the admin to control when and where the automation takes place. (See the section Prism Operations for more information on automation with Prism.)

Business Continuity Layer

Nutanix HCI is built with resiliency in mind including redundancy for power and other hardware components and the ability to architect resiliency for entire

datacenters. Nutanix provides multiple ways to provide business continuity including backup/restore and disaster recovery:

There are five main categories where Nutanix provides native protection, via hypervisor, third-party software, or a combination:

- Hardware
  - Self-healing from disk failure
  - Redundancy of key components
- Node
  - Node failure resiliency
  - Non-disruptive upgrades
- Data
  - Tunable redundancy
  - Nutanix Mine provides secondary storage integration and works with leading backup vendors.
- VM
  - VM-centric Protection
  - High Availability (HA)
  - ADS/DRS
  - Eco-system integration such as Nutanix Mine
- Datacenter/Site
  - Multi-datacenter replication including ROBO
  - Multi-site replication including ROBO
  - Cloud Connect
  - Metro Availability / Sync Replication
  - Near-sync
  - Async
  - Nutanix Xi Leap provides cloud-based DR-as-a-service (DRaaS).

© 2021 Nutanix, Inc. All rights reserved

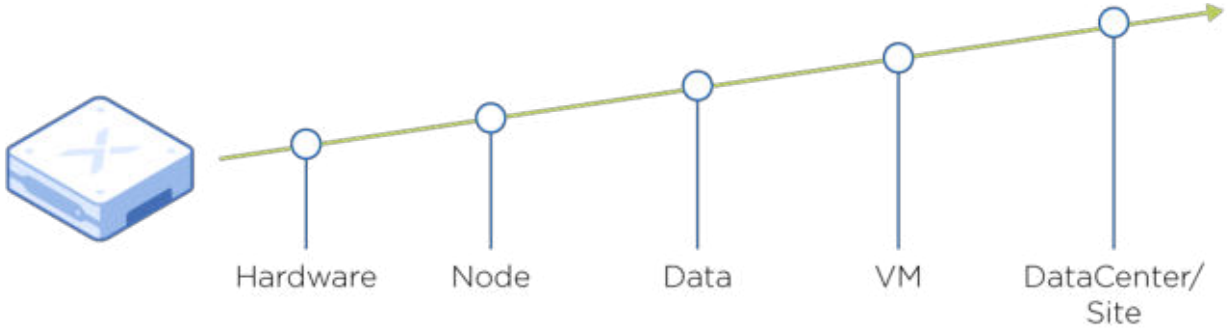


Figure 7: Business Continuity Layer

This document describes the configuration and use of Nutanix-native capabilities. Other solutions mentioned may be added to the completed design but are beyond the scope of this document.

For a list of supported 3rd party backup solutions see: Nutanix Technology Alliance Partners.

Automation Layer

Automation and orchestration are increasingly recognized as critical to IT success. By simplifying infrastructure management across the entire lifecycle, automating operations, and enabling self-service, Nutanix helps you deploy datacenter infrastructure that delivers a high degree of scalability, availability, and flexibility.

Nutanix improves efficiency with meaningful automation, self-service, and integration with development pipelines.

- Flexible task automation: Nutanix Prism Pro provides a low-code/no-code, visual approach to task automation, enabling any administrator to build, maintain, and troubleshoot automations. Common admin tasks, like adjusting resources allocated to a VM in response to a constraint, are easily automated. Even the most complex, multi-step procedures can be turned into one- click operations.
- Self-service with no loss of control: Enterprise teams want self-service access to infrastructure and services to accelerate time to market. With Nutanix Calm, you can create blueprints that model applications and tasks and

publish them to an internal marketplace or add them to a growing collection of pre-integrated, blueprints on the Nutanix Marketplace. Application owners and developers can request IT services from the marketplace whenever needed. (See the section Calm Application Orchestration for more information.)

- Simplified development: Nutanix eliminates the complexity of test and development automation, allowing developers and administrators to work more efficiently. Your team can deploy and maintain a fully automated CI/CD pipeline with continuous application deployment across on-premises and cloud locations.

DevOps is a way to standardize processes and improving communication and collaboration between development and operations teams in an enterprise organization. DevOps methodology is not discussed in this design, but the constructs and the logical components included are elements that will help you develop/improve your DevOps strategy.

Security and Compliance Layer

Designing for your security requirements is paramount to delivering robust solutions. This is another area where this document is well differentiated from conventional datacenter architectures.

Figure 8: Designing for Security and Compliance

The Nutanix security approach is not predicated on having hundreds of different configuration options that you must set to achieve a secure environment. Nutanix takes a security-first approach including a secure platform, extensive automation, and a robust partner ecosystem. There are additional configuration options available if you need to add an extra layer of security based on business and or technical requirements.

Nutanix enables you to maintain a continuous security baseline to meet regulatory requirements more easily. Powerful security automation, called Security Configuration Management Automation (SCMA), monitors the health of storage and VMs, automatically healing any deviations from this baseline.

Nutanix provides customers with the ability to evolve from point-in-time security baseline checking to a continuous monitoring/self-remediating baseline to ensure all CVM/AHV hosts in a cluster remain baseline compliant throughout the deployment lifecycle. This new innovation checks all components of the documented security baselines (STIGs), and if found to be non-compliant, sets it back to the supported security settings without customer intervention.

In addition, data-at-rest encryption features and a built-in Key Management Server (KMS) further add our robust security capabilities.

These security features are discussed later in the Detailed Design section, and you can find more information online in the Nutanix Bible.

© 2021 Nutanix, Inc. All rights reserved

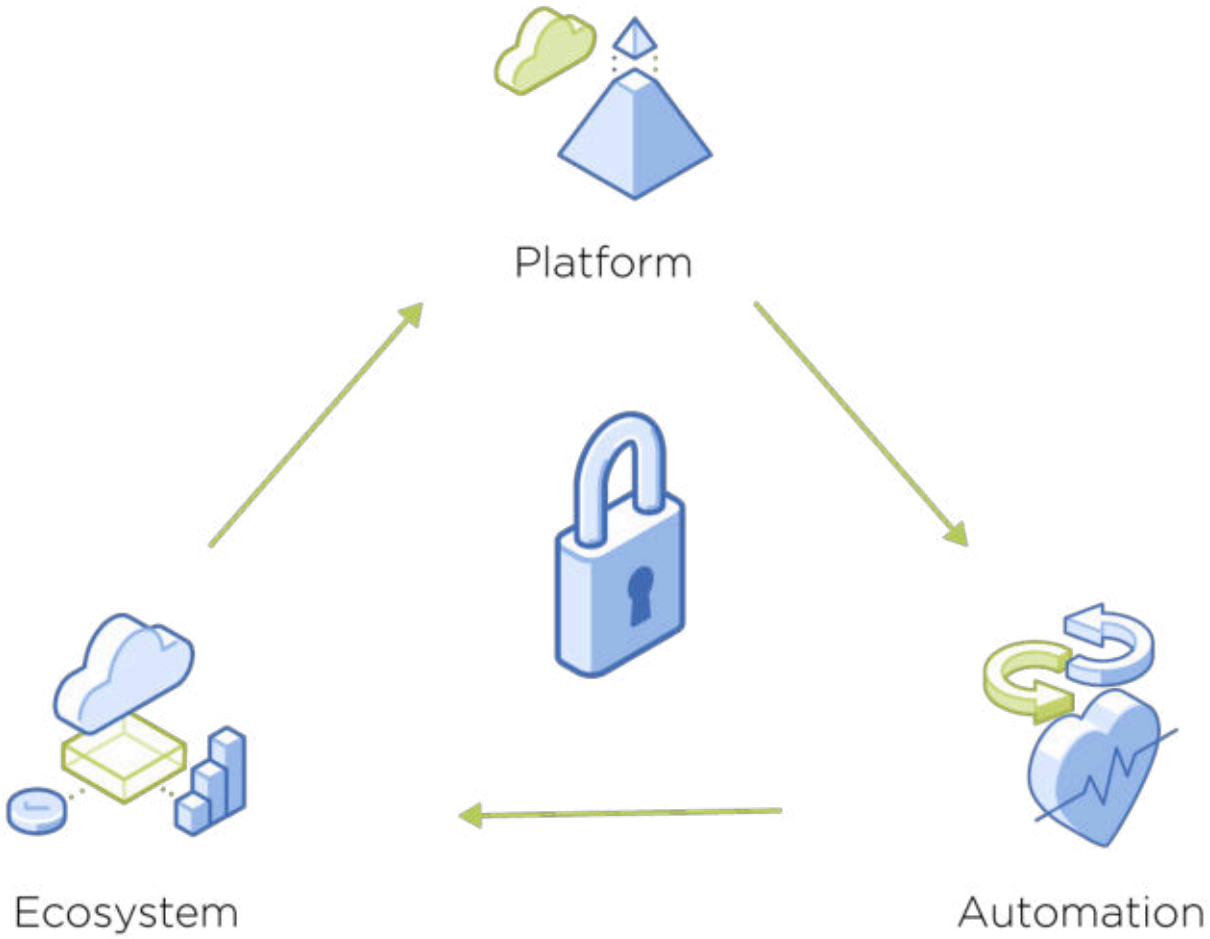


Figure 9: Nutanix Security

Nutanix incorporates security into every step of its software development process, from design and development to testing and hardening. Nutanix security significantly reduces zero-day risks. One-click automation and a self-healing security model ensure ongoing security maintenance requires much less effort.

Nutanix provides the following international security standards:

- 508 Compliant
- FIPS 140-2 Level 1
- National Institute of Standards and Technology (NIST) 800-53

© 2021 Nutanix, Inc. All rights reserved

- TAA Compliant

3. High-Level Design Considerations

A key feature of Nutanix Enterprise Cloud is choice. Nutanix customers have the flexibility to choose their preferred hardware vendor, CPU architecture, hypervisor, and more. This design is intended to help guide you to the best choices for your organizational needs.

There are a number of high-level design decisions that must be made before proceeding to a detailed technical design. This section provides the information necessary to help you make the following decisions:

- Choosing a datacenter architecture.
- Choosing a hypervisor.
- Choosing a cluster deployment model.
- Choosing a hardware platform.

Choosing A Datacenter Architecture

What availability must be delivered and what type of failures must the design protect against. These are key input requirements for a design since they will have an immediate impact on regions, availability zones, and data centers required plus budget.

Area and Availability Definitions

Region

A region is a geographical area with one or more availability zones (AZs). Regions are independent from each other so that failures in one region should not affect another region. Typical examples would be US east coast, US west coast, Europe north, Europe south, Asia east, Asia south.

© 2021 Nutanix, Inc. All rights reserved

Figure 10: Regions

Availability Zone

A region holds one or more availability zones. Each availability zone contains one or more data centers.

Availability zones are implemented such that normal failures (such as a power plant failure) in one zone will not affect another. Natural and manmade disasters such as catastrophic earthquakes and nuclear strikes may disable more than one availability zone in a region.

Figure 11: Availability Zones

Datacenter

Datacenters host hardware, management, and end user applications/services including network routers, network switches, firewalls, load balancers, physical servers running Nutanix software, and potentially third-party hypervisor(s).

© 2021 Nutanix, Inc. All rights reserved

Figure 12: Datacenters

Datacenter Architectures

Nutanix provides built-in capabilities to provide support for a multi datacenter operating model, independent of the datacenters implementation architecture which usually can be divided into single location and multiple locations.

The different datacenter location models discussed can be leveraged using private or public implementation models, or a combination of the two typically referred to as a hybrid strategy.

Single Location

- Active
  - Applications are active within the datacenter
  - Backups are stored in the same datacenter as where the applications are running.
  - Long time archiving at offsite location
  - Limited protection against datacenter failure.

Figure 13: Single Location

Multiple Locations

- Active – Active
  - Applications are active in both datacenters during normal production.
  - Both sites can provide disaster avoidance and disaster recovery for each other.

Figure 14: Multiple Locations

- Active – Passive
  - Applications are typically running from one datacenter only during normal production.
  - To make failover easier and more predictable the passive datacenter typically runs a limited number of infrastructure services during normal production.
  - The passive site provides disaster avoidance and disaster recovery functionality for the active site.

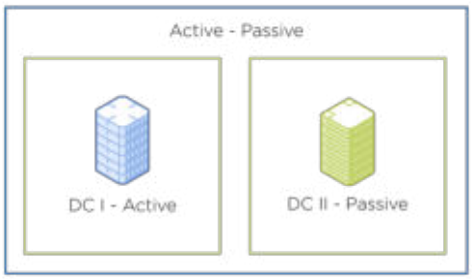


Figure 15: Active - Passive

© 2021 Nutanix, Inc. All rights reserved

- Active-Active-Passive
  - Applications are active in two datacenters during normal production.
  - The active datacenters provide disaster avoidance for each other.
  - To make failover easier and more predictable the passive datacenter typically runs a limited number of infrastructure services during normal production.
  - The passive site provides disaster recovery functionality.

Figure 16: Active - Active - Passive

Remote Office/Branch Office Architectures

In addition to a multiple datacenter model Nutanix can be used to provide business continuity for remote office branch office (ROBO) implementation models with a datacenter as a fan-in/central starting point. including e.g.:

- Datacenter – ROBO
  - Multiple ROBO sites replicating data to the datacenters.
  - Datacenter acts as disaster recovery for the ROBO sites.

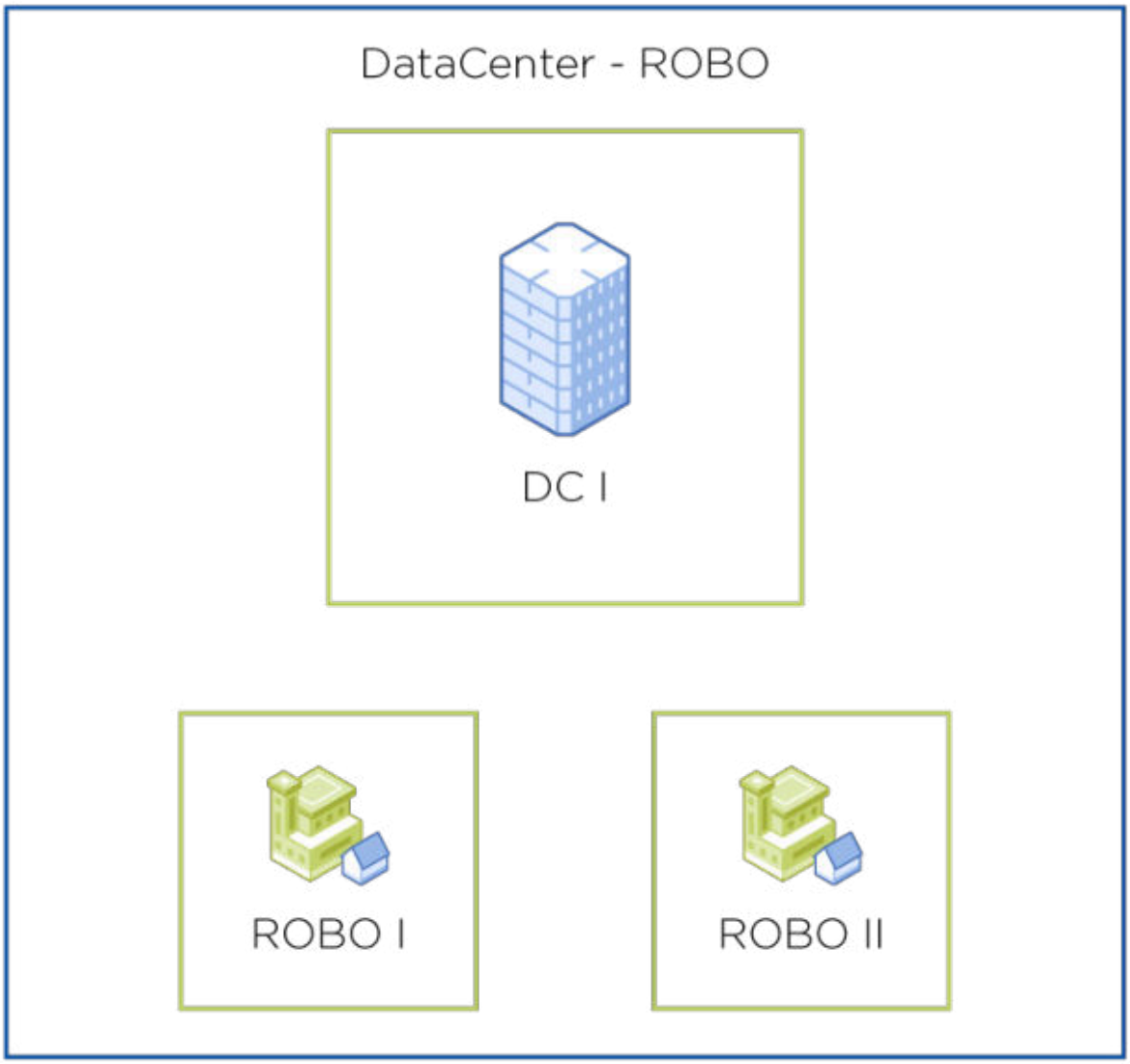


Figure 17: Datacenter ROBO

© 2021 Nutanix, Inc. All rights reserved

- Chain structure with Datacenter-ROBO-ROBO
  - Datacenter acts as disaster recovery for the closest ROBO site or sites
  - The first tier ROBO site/sites connected to the datacenter acts as disaster recovery sites for the next chain of ROBO sites. Some organizations use the term "regional data center" for these locations.

Figure 18: Chain structure with Datacenter-ROBO-ROBO

ROBO designated sites are typically a small footprint and differentiated from datacenters due to low bandwidth & high latency connections between datacenters and the ROBO sites. For larger customers, there will also be a proportionally high number of clusters/sites relative to the VM count.

Data can be replicated to more than one location for more advanced needs. An example of this policy might use NearSync between branches located in the same availability zone, and Asynchronous to a datacenter in another region

When choosing a ROBO architecture, keep in mind the restrictions that are most impactful:

- 100 millisecond or less latency between prism central and each cluster under management
- A scaled-out prism central can handle 300 clusters under management

- Accurate NTP access
- WAN restricted topology vs open internet access

Table 5: Number of Regions to be used

Region-001	NUMBER OF REGIONS TO BE USED
Justification	
Implication	

Table 6: Number of Availability Zones to be used

AZ-001	NUMBER OF AVAILABILITY ZONES TO BE USED
Justification	
Implication	

Table 7: Number of Datacenters to be used

DC-001	NUMBER OF DATACENTERS TO BE USED
Justification	
Implication	

This following table maps out the capabilities which can be applied to the different datacenter plus datacenter & ROBO architectures based on RPO and RTO requirements.

Table 8: RPO and RTO requirements

Nutanix Feature	RPO	RTO	Protect Against	Environment(s)
Time Stream / VM Snapshot	Minutes	Minutes	Minor Incident	Single Nutanix Cluster

© 2021 Nutanix, Inc. All rights reserved

Nutanix Feature	RPO	RTO	Protect Against	Environment(s)
Cloud Connect	Hours	Hours	Minor Incident	External Cloud
Asynchronous / Remote Replication	Minutes	Minutes	Major Incident	Multiple Nutanix Clusters
NearSync replication	Seconds	Minutes	Major Incident	Multiple Nutanix Clusters
Synchronous Replication	Zero	Minutes	Major Incident	Multiple Nutanix Clusters
Metro Availability	Zero	Near Zero	Major Incident	Multiple Nutanix Clusters

Typically, the multiple Nutanix clusters specified in the Environments column are placed in different datacenters but can be placed in same datacenter but in different fire zones.

Nutanix provides a variety of Node options and cluster sizes. Make sure to respect the current limitations:

- AOS Snapshot Frequency for Nutanix Nodes
- Single-Node Replication Target Requirements and Limitations

Choosing a Hypervisor

Nutanix supports a range of server virtualization options including: Nutanix AHV, VMware vSphere, and Microsoft Hyper-V. This design covers the deployment of either Nutanix AHV or VMware vSphere.

Nutanix AHV is included with AOS and delivers everything you'd expect from an enterprise virtualization solution: high performance, flexible migrations, integrated networking, security hardening, automated data protection and disaster recovery, and rich analytics. With robust, integrated management features, AHV is a lean virtualization solution.

A significant advantage of AHV as part of this validated design is the elimination of virtualization as a separate management silo and all of the

© 2021 Nutanix, Inc. All rights reserved

complexity that entails. When designing a vSphere environment, for example, decisions must be made about the number of vCenters, the type of deployment mechanism, how to enable HA, and so on. Extensive ongoing training is often required for internal staff or consultants in order to effectively design, deploy, and upgrade the environment. AHV avoids these challenges.

AHV enables:

- Checkbox high availability configuration (vs. complex percentage or slot size config)
- No virtual SCSI devices required (vs. manually configured multiple SCSI devices for maximum performance).
- Distributed networking by default (vs. deciding between Standard or Distributed Switch).
- Automatic CPU Masking (vs. manual in vSphere).
- No need for Host Profiles.
- Fewer design choices.
- Simple control plane lifecycle.
- Simplified 1-click upgrades.
- Single support path.

VMware vSphere is a proven virtualization platform used by many organizations and has a robust ecosystem. It is a complex platform with many design choices and settings to tune; it often requires the purchase of additional licenses.

When deciding which hypervisor to use in your deployment, the choice comes down to which solution best meets your business and technical requirements and your budget. Key areas to consider when choosing a hypervisor include:

- Operating system support (including legacy and current OSes).
- Third-party virtual appliance support.
- Security/hardening baseline.

© 2021 Nutanix, Inc. All rights reserved

- Integration with third party products (backup, software-defined networking, anti-virus, security tools, etc.).
- Availability of automation tools.
- Staff skill set and training (architecture, administration, daily operations).
- Scalability of the solution.
- Licensing costs and model (AHV adds no licensing costs to a Nutanix deployment).
- Integration with the existing environment.
- Migration of existing VMs.
- Hypervisor and management plane technical features and performance.
- Simplicity or complexity of the solution.
- Features or products that support only one hypervisor or the other.
- Satisfaction with existing hypervisor platform(s).
- ROI/TCO of the full stack.
- Simplified support model (i.e. single support vendor vs multiple).
- Time to Deploy/Time to Value.

Freedom of choice is a key tenet of Nutanix. After considering these factors, it is likely that one hypervisor platform stands out as the best option for your deployment. No matter which hypervisor you choose, the solution is backed by world-class Nutanix support and a full ecosystem of node types. (Hardware selection is discussed in the section Platform Considerations).

Table 9: Choose a Hypervisor

VRT-001	CHOOSE NUTANIX AHV OR VMWARE ESXI AS THE HYPERVISOR FOR YOUR DEPLOYMENT
Justification	
Implication	

© 2021 Nutanix, Inc. All rights reserved

Choosing a Cluster Deployment Model

When designing Nutanix clusters for your deployment, there are several important design considerations:

- Choosing whether to deploy a separate management cluster?
- Choosing whether your cluster(s) will run mixed workloads or be dedicated to a single workload?

Separate Management Clusters

A separate management cluster is not a requirement of this design. If you are planning a smaller-scale deployment, it doesn't always make sense to design, purchase, and operate a separate cluster for a limited set of management VMs.

However, the infrastructure applications and services that run on a management cluster are critical. At a larger scale, there are several reasons to separate these management workloads:

- Availability: Separating management workloads on their own cluster(s) makes it easier to ensure they are always available.
- Security: you may wish to more strictly control access to management clusters and workloads, including additional controls such as firewalls, dedicated networks, more stringent role-based access (RBAC), and possibly others. While this can be accomplished on mixed clusters, it is easier to monitor and manage these additional security controls on a separate physical cluster.
- Operations: RBAC prevents those that are not authorized from interacting with important management services. Physical separation prevents any accidental or malicious actions that could compromise the availability or performance of important infrastructure services.
- Performance: Performance is just as important for infrastructure services as any other workload. Having a dedicated management cluster(s) simplifies troubleshooting of performance issues and reduces the potential for hard to diagnose workload conflicts. It eliminates the possibility that changes in the management space will affect application workloads and vice versa. In a highly elastic cloud environment, there may be workload expansions that

© 2021 Nutanix, Inc. All rights reserved

occur via self-service or automated events. Depending on whether resource control policies are in effect, this can cause resource contention in a shared cluster.

> Note: You do not need a separate management cluster for each use case or environment unless you have strict security or compliance requirements that make it necessary. This means that a single management cluster can support multiple use cases such as: EUC deployments, private cloud, and general server virtualization environments.

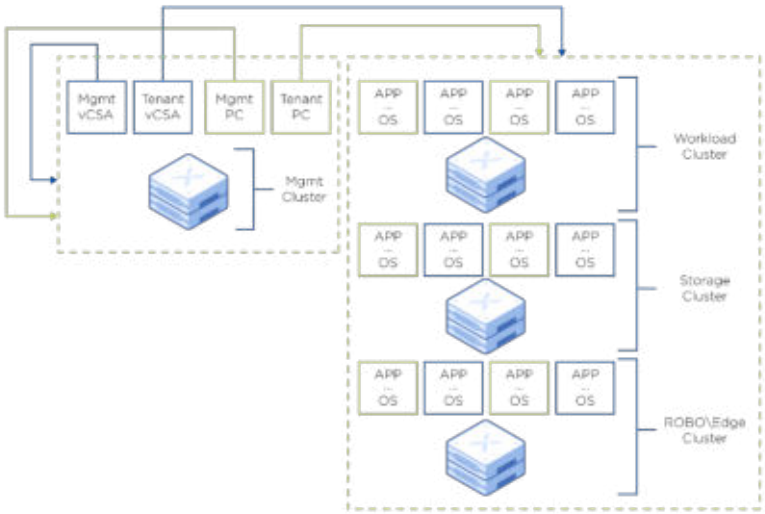


Figure 19: Separate Management Clusters

Should You Deploy Two Management Clusters?

In large-scale deployments, the management cluster can be split to create separate failure domains. With two management clusters, you can place redundant components into each cluster, enabling a higher level of availability. Should there be an issue with one of the management clusters, the other remains available to service requests.

Table 10: Management Cluster Architecture

PFM-001	MANAGEMENT CLUSTER ARCHITECTURE: DEPLOY A SEPARATE MANAGEMENT CLUSTER OR SHARE A CLUSTER WITH OTHER WORKLOADS. WHEN CHOOSING A SEPARATE MANAGEMENT CLUSTER, CONSIDER A REDUNDANT CONFIGURATION.
Justification	
Implication	

Will You Deploy Dedicated or Mixed Clusters?

The decision whether to mix workloads within a cluster or to dedicate a cluster for each type of workload is usually a question of scale. For example, if you have 200 general server VMs, a small Exchange deployment, and 10 average-sized database VMs, mixing the workloads in a single cluster is common and can be easily managed. However, if any or all of these workloads increase by 5-10x, the complexity of sizing and operating the mixed environment goes up dramatically.

Operating large-scale mixed environments creates a number of unique challenges. You have to decide whether you are willing to manage the challenges of mixed workloads or dedicate clusters for each workload. Here are the main factors to consider:

- Performance and capacity: The resource demands of different applications can vary widely. You need to understand the needs of each application when mixing workloads since the chance for conflicts to occur is increased. There may also be wildly different performance and capacity needs between applications which could require different node configurations within a single cluster. Unless you are going to isolate certain workloads to particular nodes, each node within a cluster needs to be able to handle the average daily mix of applications that might be running on it.
- VM resource-sizing requirements: The CPU and memory sizing for general server VMs, VDI VMs, business-critical application VMs, etc. can vary widely. While it's fairly easy to account for memory sizing, CPU sizing is more complex. Each of these workloads consumes different amounts of CPU, and may require much different levels of CPU overcommit, if any.

If you have large groups of VMs with widely different resource requirements, it's typically better to build clusters to contain more uniformly sized VMs. From a hypervisor HA standpoint, you may require additional resources within a mixed cluster to ensure the ability to failover. This can also increase the day 2 operational support effort, since it may require manually tuning HA settings and increased monitoring to ensure HA resources remain in compliance.

- Software licensing: The most common reason for dedicated clusters is software licensing. There are a variety of reasons why using dedicated clusters make sense from a licensing standpoint. Here are two common examples:
  - Operating system licensing: Windows and Linux vendors may offer "all-you-can- eat" license models for licensing at the host level. Therefore, it makes more sense to have clusters dedicated to either Windows or Linux VMs, to minimize licensing costs.
  - Database licensing: Database licenses are frequently based on either CPU cores or sockets. These licenses can be expensive, and you often have to license all the nodes in a cluster to enable DB VMs to run on every node. Once again, you probably don't want to run other workloads on the cluster since that reduces the return on your license investment.

In addition, you may want to run nodes hosting database workloads on different hardware than nodes hosting general server VMs. For instance, having fewer CPU cores running at a higher clock speed may reduce your overall licensing costs while still providing the necessary compute power.

This is provided for informational purposes only, please refer to your licensing agreement with Microsoft for more information.

- Security: In many projects, security constraints are an overriding factor. When it comes to the security of mixed vs dedicated clusters, there are a few design considerations that are important to consider as you decide whether logical or physical separation is adequate to address your security requirements:
  - Operations: RBAC is the primary means of controlling access and management of infrastructure and VMs in a mixed cluster. Dedicated

© 2021 Nutanix, Inc. All rights reserved

clusters prevent non-ap proved parties from gaining any access whatsoever.

- Networking: A mixed cluster typically relies on separate VLANs and firewall rules per workload to control access. A dedicated cluster would only have the required networks presented to it for a single workload and would likely limit who and what has network access. Both approaches can control and limit access to cluster resources, but dedicated goes a step further by providing complete physical network isolation for those that require it.

If you are able to address the performance/capacity, resource-sizing, licensing, and security constraints discussed above, it is possible to successfully design and operate clusters of any size to run mixed workloads. The key is to thoroughly understand the requirements of each workload and size with that knowledge.

Recommendation

At scale, it is our recommendation to use dedicated clusters to the greatest extent possible for the reasons discussed above. Utilize mixed workload clusters where it fits.

Including Storage-Only Nodes in Clusters

Storage-only nodes contribute storage capacity and I/O performance within a cluster. Storage only nodes are available for clusters running either AHV or ESXi. Storage-only nodes can be any node type, but typically are configured with just enough CPU and memory resources to run the CVM since there aren't any application VMs running on these nodes. These nodes are a member of the AOS cluster but are not visible to the hypervisor cluster for non-storage functions, so the hypervisor won't schedule other VMs to run on them.

Table 11: Mixed or Dedicated Workloads Per Cluster

PFM-002	MIXED OR DEDICATED WORKLOAD PER CLUSTER
Justification	
Implication	

Choosing How You Will Scale

This document clearly lays out the Nutanix opinionated design for deploying HCI clusters and building a private or hybrid cloud for any type of workload. A challenge in any design, whether you're starting with one cluster or dozens, is how to scale the environment to reach your goal. To accomplish this, you need an overarching architectural plan that creates a repeatable process that can be followed by the organization. Defining this in advance removes confusion and simplifies future deployment. An overarching architectural plan allows you to track progress and ensure compliance with the design.

For this design, we recommend that an at-scale deployment should include:

- A control plane.
- A block and pod architecture.

Choosing a Control Plane

Within your design there should be a primary control plane where the majority of daily operational tasks are performed. In this design:

- The primary control plane is Prism Central.
- VMware vCenter is also required for those using VMware ESXi.

Each of these control planes has a maximum size that dictates the number of VMs, nodes, and clusters that can be managed by an instance.

Block and Pod Architecture

A repeatable architecture is needed to ensure safe and efficient scaling. This design uses the pod and block architecture because it's both familiar and easily consumed. We have adapted it to support Nutanix deployments. This section explains the different parts of the architecture and how they can be used to scale a deployment to any level.

What is a Pod?

In this design, a pod is a group of resources that are managed by a single Prism Central instance. The diagram below shows a single pod containing four

© 2021 Nutanix, Inc. All rights reserved

building blocks. A pod is not bound by physical location limitations; all of its resources could be at a single site or at multiple sites. Examples of multiple sites include: a traditional multi-datacenter design, a hybrid cloud design,or a ROBO architecture.

Building Block

In this design, a build block is equivalent to a Nutanix cluster. Each of the clusters can run a single dedicated workload or mixed workloads. There is no need for building blocks to be uniform within your design.

Recommendation

For each workload:

- Establish whether it will have a dedicated cluster or share a mixed cluster with other applications.
- Define the maximum size of this type of building block. While building blocks for different workloads do not need to be the same, they certainly can be.
- Determine the maximum size of a building block based on:
  - The scale at which you will deploy workloads.
  - The need for failure domains.
  - Operational considerations such as upgrade timing.

These topics are discussed later in the Detailed Technical Design section of this document.

© 2021 Nutanix, Inc. All rights reserved

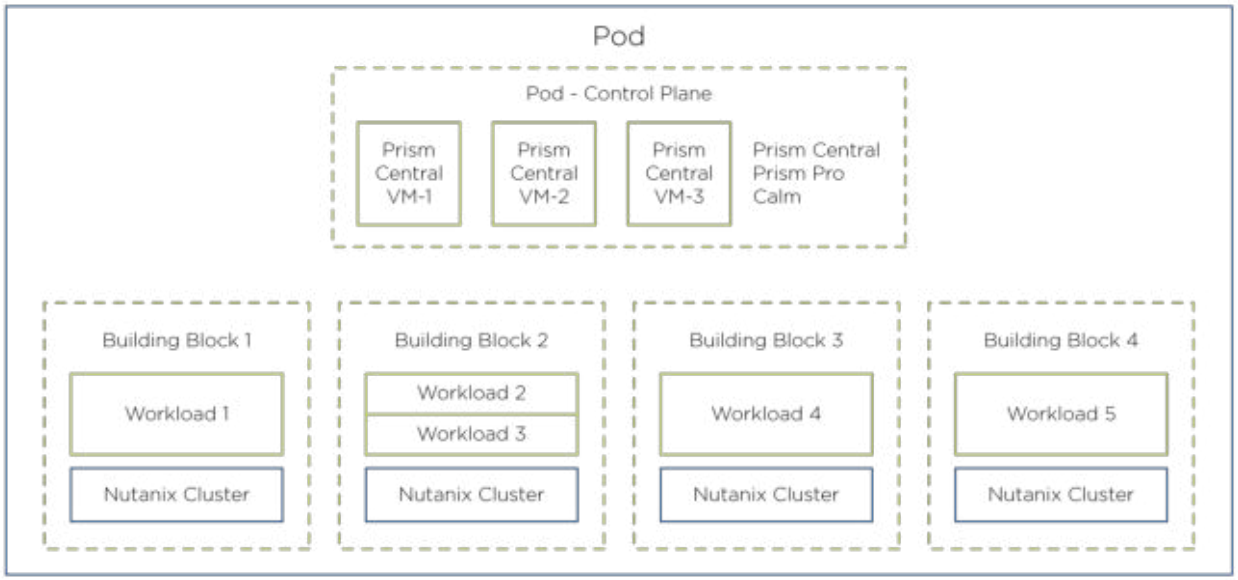


Figure 20: Block and Pod Architecture

Scaling the Block and Pod Architecture

You can scale an individual pod up to the maximums that Prism Central supports for the AOS version you are deploying. For the AOS version specified in this document, a scale out Prism Central deployment with large sized VMs can manage up to:

Prism Central Limits (assumes Scale-Out PC configuration):

- 25,000 VMs
- 200 clusters
- 1,000 nodes

If any of these limits are reached, a pod is considered full. For example, with an EUC deployment, the VM limit will likely be reached first since you typically have high VM density resulting in large numbers of VMs with fewer nodes and clusters. A large ROBO environment might hit the cluster count limit because you tend to have many sites each with a small cluster and a few VMs.

At very large scale (e.g. thousands of nodes), it can make sense to have pods dedicated to each workload, but for environments that have Nutanix deployed for multiple workloads, a pod will typically contain multiple applications.

Once a pod reaches a scaling limit, start a new pod with at least one building block. The new pod scales until it also reaches a scaling limit, and so on.

The building blocks within a pod scale in a similar fashion. A building block is started, and workloads are migrated onto it until it reaches its determined max size, and a new building block is started. New building blocks can be as small as 3 nodes, the minimum to start a Nutanix cluster, or any size up to the max size you've specified for that building block.

The starting size for each building block and the increments for scaling them are organizational decisions:

- For smaller or more agile organizations, starting small and scaling incrementally often makes sense.
- Larger organizations may prefer to deploy a new building block fully populated and migrate workloads onto it as schedules dictate.
- Although 3 nodes is the minimum size for a cluster, using 4 nodes provides a higher level of redundancy during maintenance and failure conditions.
- For ROBO and edge use cases, the starting size can be as small as one, two, or three nodes depending on requirements.

The diagram below illustrates a simple VDI building block example. With VDI it is easy to think in terms of number of users and the node count in a cluster. In this example, the building block is a 16-node cluster supporting 1,500 users. This works out to 100 users per node plus an additional node for HA. With the first building block full, a request for an additional 500 users requires a new building block to be started. This building block is then scaled up to its max size before starting a third.

© 2021 Nutanix, Inc. All rights reserved

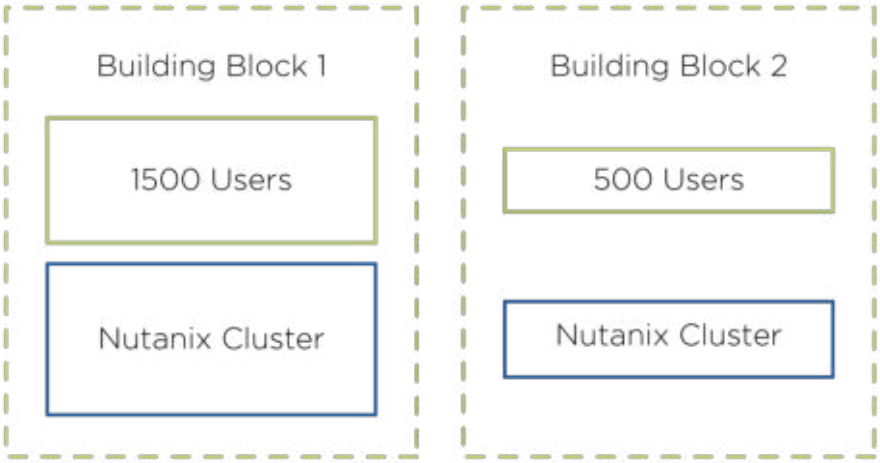


Figure 21: Example VDI Building Block

By having well established and documented design decisions for pod size and building block size, the architecture and operational teams are free to keep scaling without the need to revisit decisions to satisfy each resource expansion request.

Choosing the Right Licensing and Support

Now that you've chosen a hypervisor, decided on mixed or dedicated workloads, and established your block and pod architecture, you can make informed hardware decisions. When evaluating the different Nutanix platform options available, there a few key decision points:

- Choosing Your Software licensing and Support.
- Choosing Your Platform Vendor.

Software Licensing and Support Considerations

Nutanix nodes are available in two different purchasing/licensing options: appliances or software-only:

- Appliances are available directly from Nutanix or through our OEM relationships.
  - Appliance based licensing is referred to as “life of device licensing”, meaning it’s only applicable to the appliance it was purchased with.
  - The manufacturer of the appliance takes all support calls for software and hardware issues. For example, if you choose Dell appliances, Dell will take all support calls and escalate to Nutanix for software support as needed. With Nutanix NX appliance all support calls go directly to Nutanix.
- The Software Only option de-couples software and support licensing from the underlying hardware. This enables:
  - License portability. The same license can continue to be used when the underlying hardware is changed, such as a hardware vendor change or a node refresh. (note: licenses are portable for like-for-like hardware replacements. If the hardware specification of the nodes change, then there might be the need for additional licenses.).
  - Deployment on additional supported hardware platforms. Hardware is per our qualified list of platforms, which can be found on the Hardware Compatibility List (HCL).
  - Direct software support from Nutanix while the server vendor provides hardware support.
  - Another type of Software Only Licensing is the Core-Licensing option, which is best for those customers who prefer the benefits of the software-only model but want hardware from a specific OEM server vendor. Core licensing enables customers to purchase software only licenses and buy hardware from any of the appliance vendors. For example, XC-Core utilizes Dell XC OEM appliances but de-couples software and hardware support.

Table 12: Nutanix Software Licensing Level

PFM-003	SELECT NUTANIX SOFTWARE LICENSING LEVEL
Justification	
Implication	

Vendor Considerations

When it comes to selecting a hardware vendor, there are a number of factors to evaluate:

- Brand loyalty. This can be a strong factor in an IT buying decision. There may be purchasing commitments or discounts in place at an organizational level driving this loyalty, or you may simply have been happy with past experiences.
- Support Quality. The quality of the support experience can also be a factor in your hardware evaluation. For hardware failures, makes sure the hardware vendor responds quickly and can provide parts reliably within the contracted response time. The overall support experience is important. A vendor should be easy to contact, be responsive to requests, and provide resolution in a timely manner.
- Hardware Quality. The reliability and quality of hardware is of obvious importance. Today's servers are all very similar internally; they use many of the same components with just a few proprietary components for each vendor. The reliability of the leading server vendors is pretty similar so your decision may be contingent on your organization's past experiences.
- Operational Experience. When it comes to the ongoing operational experience, what does it take to carry out day 2 operations to support the lifecycle of the physical server and the vendor toolset (if any). This includes: monitoring server health, reporting and upgrading firmware, and monitoring for component issues and failures. Virtually all of the server vendors offer tools for these activities and when combined with the power of AOS and Prism, the experience is pretty similar. Nutanix Lifecycle Manager (LCM) offers firmware reporting and management for all Nutanix appliance options.

© 2021 Nutanix, Inc. All rights reserved

- Configuration Options. Physical configuration options may weigh strongly in choosing among server vendors, and in choosing a model type for each workload. Factors include:
  - Number of sockets.
  - CPU options.
  - Number of storage bays.
  - Network connectivity options.
  - Storage media options. While there is generally a level of parity between server vendors, a particular vendor may offer something the others do not, or one vendor may offer the latest options more quickly when new components are released. You may require or prefer a specific type of network card or need nodes with a large number of storage bays. Storage considerations, such as the number of SSD/NVMe options available, RDMA capabilities, and support for large-capacity media may be important for specific workload requirements.
- Physical form factor. This is another common decision point since it can affect the amount of space consumed in a rack, the power draw, the number of network connections, and the number of internal expansion slots. Availability of internal expansion slots may limit the number and type of network cards that can be deployed, and whether a node is capable of accepting GPU cards and the number of GPU cards it's capable of supporting. When it comes to different form factors, there are:
  - High-density chassis that offer either four or two physical nodes in 2U of rack space. These are popular options for a variety of workloads that do not require extensive internal expansion or a large number of storage bays.
  - Standard rackmount servers, typically with a 1U or 2U chassis and one physical server per chassis. These provide much wider capabilities in the number of storage bays available, the number of internal expansion slots available, and may also support more memory.

© 2021 Nutanix, Inc. All rights reserved

Table 13: Physical Node Vendor

PFM-004	SELECT PHYSICAL NODE VENDOR
Justification	
Implication	

Mixed Configurations and Node Types

Nutanix clusters allow significant flexibility in terms of the node types and configurations you can utilize in a single cluster. This allows clusters to be operated and expanded over time without artificial constraints. A cluster can be expanded with different node configurations to accommodate new workloads or when previous nodes are no longer available.

Considerations include:

- Node models. Mixing node models within a cluster is a fairly regular occurrence. While it's possible to have the same CPU, memory and storage configuration in two different node types, it's not required in order to mix them in the same cluster.
- CPU configurations. Mixing nodes within a cluster with different CPU configurations such as core count, clock speed, or CPU generation is supported. This can be to address changing application requirements, inventory availability, financial constraints, time of purchase, or other factors.

While there is no limit to the drift between configurations, it's a commonly accepted best practice to keep the core counts and memory configuration of nodes within a cluster at similar levels. Using different CPU generations in the same cluster can limit the feature set / functionality of newer CPUs. The lowest common denominator is the level of the oldest CPU generation within a cluster. Having mostly uniformly configured nodes in a cluster makes it easier for humans to double check the HA and capacity planning recommendations of automated tools.

- Storage media. Having nodes with different storage configurations is also supported. Variations at the storage layer can include varying the size or number of SSDs or adding all-flash nodes to a hybrid cluster.

© 2021 Nutanix, Inc. All rights reserved

When introducing a storage configuration within a cluster that dramatically increases the amount of storage, you must ensure there is ample failover capacity to rebuild the largest node. For example, suppose the existing nodes in a cluster have 10TB of capacity each and you want to add node(s) with 40TB of capacity. Initially, it's best to add a pair of these 40TB nodes. Subsequent large-capacity-node additions can then be done one at a time.

Choosing the Right Hardware for Your Workload Types

When it comes to sizing Nutanix clusters for different application workloads there are many options. Most workloads can successfully run on any of the Nutanix and OEM models available, but some models and configurations may be a better fit than others.

Once you know the requirements of your workload(s) you can use the Nutanix Sizer to determine the best configuration for your cluster(s).

Nutanix Sizer is a web-based application that is available to Employees, partners and select customers. Sizer allows for the architect to input application and workload requirements and the node and cluster configurations are automatically calculated.

Different hardware platforms offer different characteristics to address differences in workloads. The following sections provide guidance on model selection, and performance considerations.

Model Types

The various Nutanix and OEM appliance models and hardware compatibility list (HCL) servers provide the flexibility to identify the right solutions to meet financial, space, and performance requirements for different projects. There may be differences in terms of the number and types of models available, but the same level of flexibility is generally available from all vendors. Some vendors offer fewer models but allow greater configuration flexibility.

Generally speaking, there are four different groups that servers fall into, translating to different use cases:

- General workloads and EUC: These are by far the most popular nodes deployed in Nutanix clusters. They can handle the vast majority of workloads

including: general server virtualization, business-critical applications, VDI, and most others. There are a mix of form factors available.

- ROBO and Edge: These are similar to the general workload options, with the exception that they may offer few options for CPU and storage as they are optimized for these edge use cases.
- Storage dense: For workloads that require large amounts of storage capacity, storage-dense nodes offer a larger number of storage bays and dense media options, possibly with fewer CPU options. The physical configuration of these nodes is optimized for workloads such as Nutanix Files, Nutanix Objects, or to be utilized as storage-only nodes.
- High performance: The most demanding workloads and business-critical applications (BCA) may require additional CPU resources and storage performance. For these workloads there are nodes offering additional CPU configurations in terms of core count and clock speed, as well as quad-socket configurations. It's common for these models to offer as many as 24 storage bays to allow for more flash devices or hard drives for workloads that can utilize the added performance characteristics.

Each of the above model alternatives offer one or several of the available physical form factors along with the density and performance characteristics discussed.

Nutanix software does not require any complex tuning or configuration to support the different workloads, but there are plenty of hardware options to tailor your selections to different use cases.

Performance Considerations

Nutanix and AOS meet the performance demands of different workloads without continuous performance tuning. The Nutanix HCI storage fabric is powerful and intelligent enough to handle nearly any type of workload.

However, different cluster design and configuration options still yield performance benefits. Selecting the appropriate node model and configuration to meet application and solutions requirements is an important design decision. The primary design considerations for performance are:

© 2021 Nutanix, Inc. All rights reserved

- Number of drives. The number of hard disk drives (HDDs) or flash devices in a node can dramatically affect its performance characteristics. However, simply picking the node with the most device bays won't improve the performance of every workload.
  - Write-heavy workloads benefit from additional storage devices to provide performance and consistency. Other workload characteristics such as read/write ratio and I/O size should also be considered.
  - Workloads such as VDI typically have minimal capacity requirements but higher IOPS demands. It's common to utilize nodes with partially populated storage bays and as few as 2 flash devices per node, providing the right amount of storage capacity while still exceeding performance demands.
- All Flash. All-flash configurations are available from Nutanix, OEMs, and supported third-party server vendors. All-flash clusters utilize only SSDs, and these configurations provide higher IOPS and a more consistent I/O profile. While all flash configurations have become common, they are not absolutely necessary for every workload and use case.
- NVMe. There are a number of new technologies available now and coming soon that offer additional performance capabilities. NVMe is the first of these to be widely available and offers a number of benefits over SSD. New flash technology allows NVMe devices to deliver higher levels of I/O with lower latencies.
- RDMA. To realize the full benefits of NVMe, nodes are typically configured with remote direct memory access (RDMA). RDMA allows one node to write directly to the memory of another node. This is done by allowing a VM running in the user space to directly access a NIC, which avoids TCP and kernel overhead resulting in CPU savings and performance gains.
- Size of flash tier. In hybrid configurations containing SSD and HDD devices, the bulk of the performance comes from the flash tier. Therefore, it's important to understand the workload being deployed on a hybrid cluster. The data an application accesses frequently is typically referred to as the working set. The flash tier in hybrid clusters should be sized to meet or exceed the size of the working set for all of the applications that will run on

© 2021 Nutanix, Inc. All rights reserved

the cluster. There is no penalty for having too much flash in a cluster but not having enough can result in inconsistent performance.

Table 14: Node Models

PFM-005	SELECT NODE MODEL(S) PER USE CASE
Justification	
Implication	

Note: As your organization works through the Detailed Technical Design elements in the following section, be prepared to revisit your model decisions to fine-tune CPU, memory, and storage configurations.

© 2021 Nutanix, Inc. All rights reserved

4. Detailed Technical Design

With the necessary high-level design decisions made, including hypervisor, deployment model, and hardware platform and models, you can now plan the technical design for your Nutanix deployment.

This section provides technical guidelines for each layer of the design stack. Disaster recovery and business continuity guidelines are provided in chapter 5. Where possible, we've organized the sections so that you don't have to spend time reading material that doesn't apply. For instance, if you are not deploying VMware, you can skip all sections that are applicable only to it.

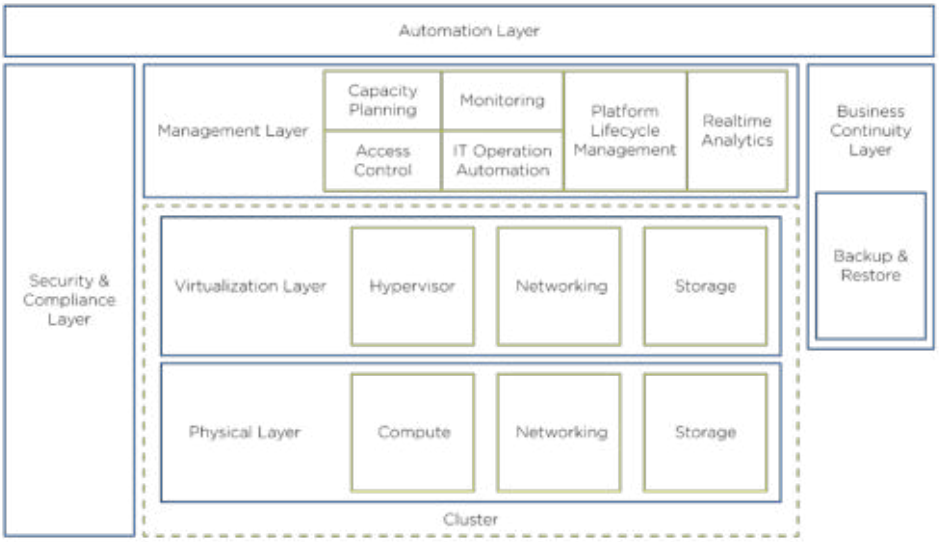


Figure 22: Planning the Technical Design

Required Software Versions

This design assume that you will be running the following software versions:

Table 15: Software Versions

Name	Version	Description
Prism Central	Latest version	Use latest version available (pc.2020.7 at the time of writing).
Nutanix AOS	Latest LTS	Use latest LTS available (5.15 at the time of writing).
Hypervisor Options		
Nutanix AHV	Latest LTS	Use latest LTS available (5.15 at the time of writing).
VMware vSphere	6.7 or 7.0	
VMware vCenter	6.7 or 7.0	Recommended for vSphere deployments.

Physical Layer Design

This section guides you through the process of designing all physical aspects of your Nutanix deployment, including:

- Compute and Storage Design.
- Networking.
- Choosing cluster size.
- Failure domain considerations.
- Designing workload domains and cluster layout.

Choosing the Optimum Cluster Size

This section provides guidance for cluster sizing. When designing a Nutanix cluster it's important to consider more than just technical limits and recommendations. There are other important considerations including: hardware vendor recommendations, security, and operational requirements that may dictate the optimal cluster size.

© 2021 Nutanix, Inc. All rights reserved

The following table provides a comprehensive list of the design considerations that pertain to cluster size:

Note: These considerations apply to cluster sizing. Additional information for each area is provided in later sections.

Table 16: Cluster Size Design Considerations

AREA	LIMITING FACTOR(S)	CONSIDERATIONS
Operations/ manageability	Maintenance window	Define/validate your maintenance window and make sure the cluster upgrade process fits in the window. Example: • Maintenance window: 12h. • Full single node upgrade: at least 45m (Hardware, FW, AOS, and hypervisor). Time depends on hardware vendor. • Maximum cluster size: 12 -15 nodes.

© 2021 Nutanix, Inc. All rights reserved

AREA	LIMITING FACTOR(S)	CONSIDERATIONS
Security & compliance	Security zones within an organization or business unit	Collect all relevant security and compliance requirements. Example: • Multiple security zones: Internet DMZ, PROD, test & dev, inner-DMZ • Every security zone may dictate different cluster sizes: › Internet-facing DMZ clusters usually have a smaller number of nodes and smaller number of workloads to minimize impact of a security breach or DDoS attack. › For example, Dev/Test clusters that have lower criticality may also have less strict change management policies and can therefore have a large numbers of hosts.
Vendor recommendations and limitations	• Hypervisor limitations • Management plane limitations • Vendor recommendations	Each product has limitations and vendor recommendations. Make sure you do not cross boundaries set by the vendor. See the vendor limitations/ recommendations table.

© 2021 Nutanix, Inc. All rights reserved

AREA	LIMITING FACTOR(S)	CONSIDERATIONS
Business continuity	• RPO • RTO • Backup window	Collect BC/DR requirements, RPO, RTO, backup and restore time window, backup system performance statistics. Example: • RPO 24h • RTO 48h Ensure you can recover/restore from backup and restart workloads within 48h. A cluster where total storage capacity exceeds technical capabilities of the backup system could fail to meet desired RTO.

© 2021 Nutanix, Inc. All rights reserved

AREA	LIMITING FACTOR(S)	CONSIDERATIONS
Workload considerations	• Application architecture • Application licensing • Application criticality	Verify application architecture with the application team/vendor, including HA, DR, scale-in vs scale-out, and performance requirements. Consider licensing model for each application and its implications. Example #1: Oracle or MS SQL licensing Licensing models are based on physical core count. Design clusters for database performance and capacity requirements to avoid cluster oversizing and minimize license costs. Example #2: Application has its own HA or DR If application can provide native HA and/or DR, RPO/RTO considerations described under Business Continuity (above) may not apply.

© 2021 Nutanix, Inc. All rights reserved

AREA	LIMITING FACTOR(S)	CONSIDERATIONS
Networking	• Total available network switch ports. • Available network switch ports per rack.	Available physical ports per rack and rack row is important when choosing cluster size and number of clusters. Example: • 96 ports (10GbE) available per rack. • 48 ports (1Gbps) available per rack. • 2 x 10GbE uplinks per Nutanix host. • 1 x 1Gbps uplink for Out-of-Band management. Maximum nodes per rack is 48 (total capacity of the TOR switches).
Datacenter Facility	• Available server rooms. • Total power and cooling. • Power and cooling per rack. • Available total rack units. • Available rack units per rack. • Floor weight capacity.	Power and cooling is one of the most important factors limiting Nutanix cluster size and node density. You have to ensure you do not exceed any hard limits when designing your cluster layout. When calculating power consumption and thermal dissipation use maximum values provided by vendor. Typical datacenter rack is 42U. Some datacenters have racks up to 58U.

Maximums and Minimums: AHV Deployments

The following table shows the maximum limits for management plane software components in AHV deployments:

Table 17: Maximum Limits for Management Plane Software Components in AHV Deployments

Management Software	Max. # of Hosts	Max # of VMs	Notes
Nutanix Prism Central	Up to 200 clusters or 1000 nodes or 25000 VMs.		Assumes Prism Central scale out is deployed.

The following table provides guidance regarding minimum and maximum number of nodes supported in a single Nutanix AHV cluster:

Table 18: Min and Max Number of Nodes in a Single AHV Cluster

Min # of Nodes	Max # of Hosts	Hypervisor	Recommendation
1	No Limit. Cluster size based on a variety of factors.	Nutanix Acropolis Hypervisor (AHV).	Single and dual-node clusters are for ROBO only.

Maximums and Minimums: VMware vSphere Deployments

The following table shows the maximum limits for management plane software components in VMware deployments:

Table 19: Maximum Limits for Management Plane Software Components in VMware Deployments

Management Software	Max. # of Hosts	Max # of VMs	Notes
Nutanix Prism Central	Up to 200 clusters or 1000 nodes or 25000 VMs.		Assumes Prism Central Scale-Out.
VMware vSphere vCenter	2000 hosts	25000 powered on	

The following table provides guidance regarding minimum and maximum number of nodes supported in single VMware cluster:

Table 20: Min and Max Number of Nodes in a Single VMware Cluster

Min # of Nodes	Max # of Nodes	Hypervisor	Recommendation
2	64	VMware vSphere ESXi	Dual-node clusters are for ROBO workloads only. Maximum number of nodes in single Nutanix cluster with VMware ESXi is limited by hypervisor version. For more details refer to official VMware documentation.

Table 21: Number of Clusters

PFM-006	NUMBER <<TYPE AND SIZE>> OF CLUSTERS
Justification	
Implication	

Failure Domain Considerations

Failure domains are physical or logical parts of a computing environment or location that is adversely affected when a device or service experiences an issue or outage.

The device or services that are affected can greatly affect the size of the failure domain and its potential impact. For example, a router generally has a bigger failure domain than a wireless access point since more endpoints rely on a single router than a single access point. Identifying possible failure domains and keeping the size of failure domains small or manageable where possible, reduces the chance of widespread disruption.

Building redundancy within and/or across failure domains is an important method to help mitigate the risks of failure.

When designing a Nutanix deployment, you can take steps to mitigate risk for each of the following failure domains:

- Drives

- Nutanix node
- Nutanix block
- Management plane
- Nutanix cluster.
- Datacenter rack and server room
- Datacenter

Nutanix clusters are resilient to a drive, node, block, or rack failure, which is enabled by Redundancy Factor 2, the default. Redundancy Factor 3 can enable simultaneous drive, node, block, or rack failures with the right architecture. After drive, node, block or rack failure, a Nutanix cluster self-heals to reach the desired redundancy factor and rebuilds resilience for additional subsequent failures.

> Note: You can configure your Nutanix environment to be fault tolerant to node, block, and rack failures. This is described later in the section: Data Redundancy and Resiliency. Mitigating the risks of network failure domains is described in the section: Networking.

The Management Plane

One of the most important failure domains, and one that is often overlooked by architects, is the management plane. The more workload domains managed by a single management plane, the bigger the impact of a failure. When deploying the management plane, consider the following risk mitigations to reduce the impact of a failure.

Table 22: Management Plane Risk Mitigations

AREA	RISK MITIGATION
Availability	• Design and deploy management plane to be highly available. • At a minimum, design to meet the availability. requirements of the managed workload or service with the highest uptime requirement.

AREA	RISK MITIGATION
Limit the impact	• Confine workload domain to a single datacenter or site. • Confine workloads domain to a defined security zone. • Ensure the API gateway is always available because other 3rd party integrations may rely on it. (e.g. 3rd party backup vendor integration.).
Access Control	Configure built-in RBAC to restrict access to management platform resources.

Note: The section Management Layer provides more details on management plane configuration.

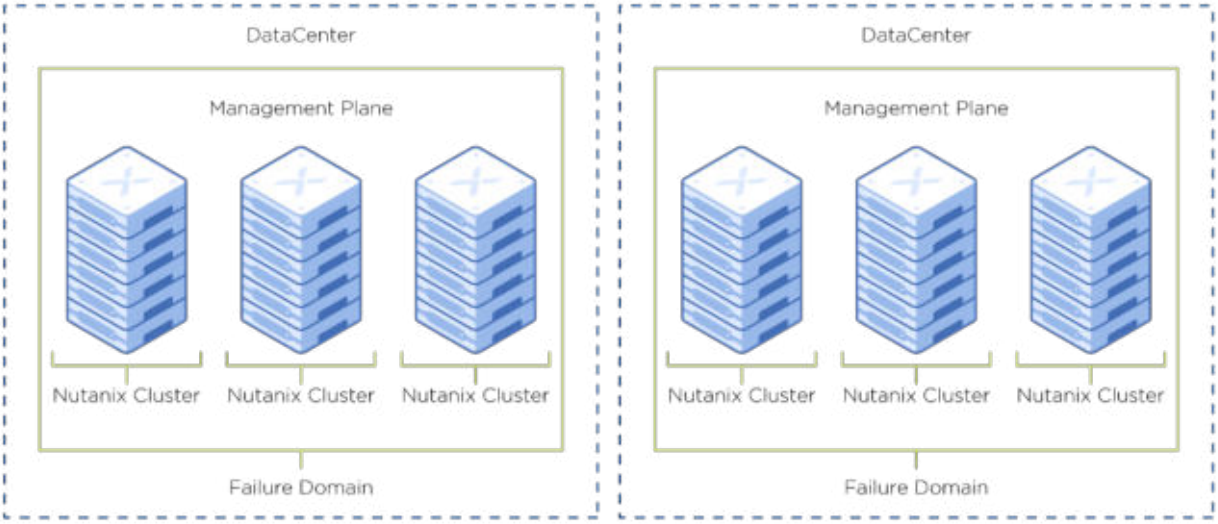


Figure 23: Management Plane Failure Domain

The Nutanix Cluster as a Failure Domain

When evaluating each Nutanix cluster as a failure domain, you have to consider the risks and potential effects in terms of both the size of the cluster (as described above in Choosing the Optimum Cluster Size) and the workloads running in the cluster, including whether the cluster is running mixed or dedicated workloads.

© 2021 Nutanix, Inc. All rights reserved

Large clusters result in larger failure domains and potentially higher business impacts, since they typically host considerably more workloads. To mitigate the risk of data unavailability or service disruption, design for redundancy at a cluster level to protect data and services, as described in the following table:

Table 23: Designing for Redundancy at the Cluster Level

AREA	RISK MITIGATION
Power	Redundant power from two different power supplies.
Networking	Redundant TOR switches. Redundant upstream connectivity to TOR switches from each Nutanix node.
Cluster	Leverage Nutanix scale-out architecture and data protection capabilities to replicate data to a second Nutanix cluster in the event that one cluster fails.
Application	Deploy application across multiple clusters.

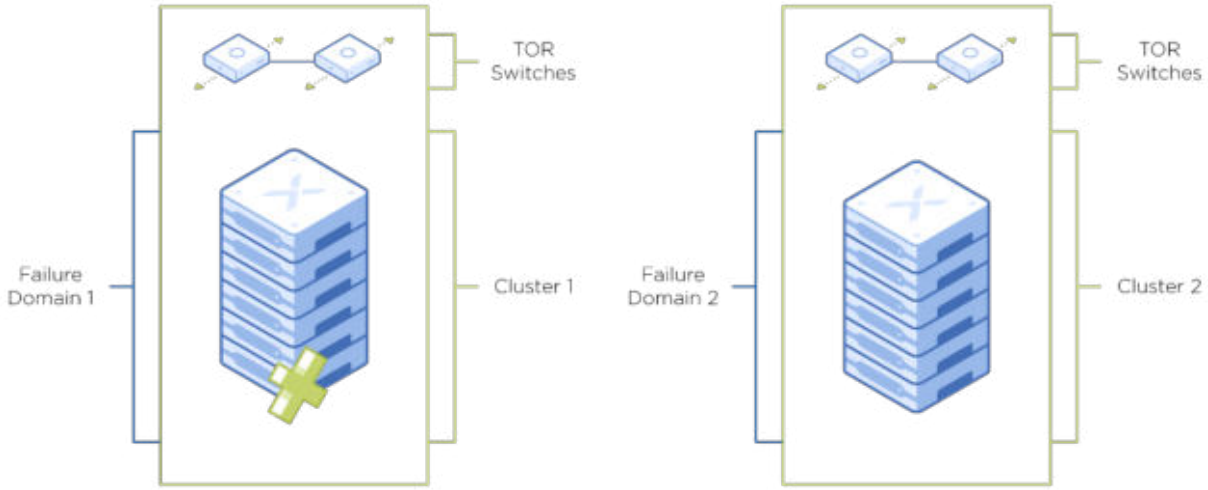


Figure 24: The Nutanix Cluster as a Failure Domain

Datacenter Rack and Server Room Failure Domains

When considering the datacenter rack failure domain, the primary mitigations are redundant power from two different power supplies to each rack, redundant TOR switches, and redundant network uplinks.

When considering the datacenter server room failure domain, it is critical to examine all datacenter components to ensure they are not shared among multiple server rooms, as described in the following table:

Table 24: Datacenter Rack and Server Room Risk Mitigation

AREA	RISK MITIGATION
Power	Redundant power
Cooling	Independent cooling
Server Room	Provide redundant server room within separate firezone (in the same datacenter)
Application	Place application in multiple server rooms. For example, active directory domain controllers in separate server rooms.

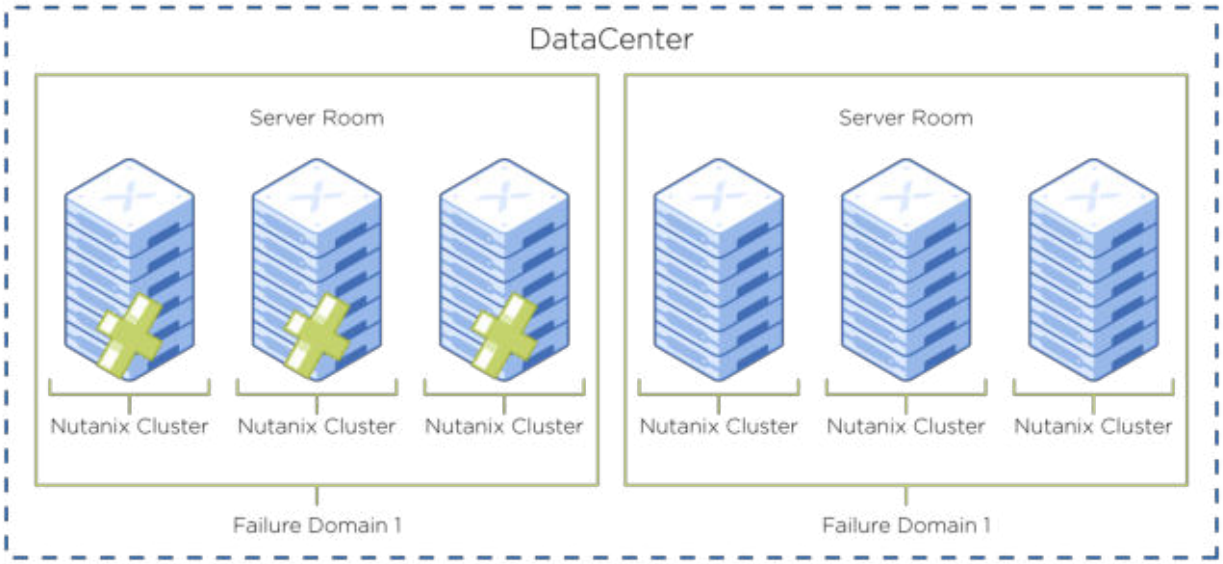


Figure 25: Datacenter Rack and Server Room Failure Domains

Datacenter Building

When considering the datacenter failure domain, it is critical to examine the redundancy of all connections to the outside to ensure they are not shared, as described in the following table:

Table 25: Datacenter Building Risk Mitigation

AREA	RISK MITIGATION
Power	Redundant power supply from different suppliers.
Cooling	Independent/redundant cooling system for each datacenter room.
Network Connectivity	Redundant networking connectivity between datacenter buildings from different providers Redundant internet connectivity from different ISP
Datacenter buildings/server room	Where possible, utilize multiple datacenter buildings and/or server rooms in separate datacenter buildings so a single server room does not affect all of production. Distribute multi-component services equally across datacenter server rooms.

© 2021 Nutanix, Inc. All rights reserved

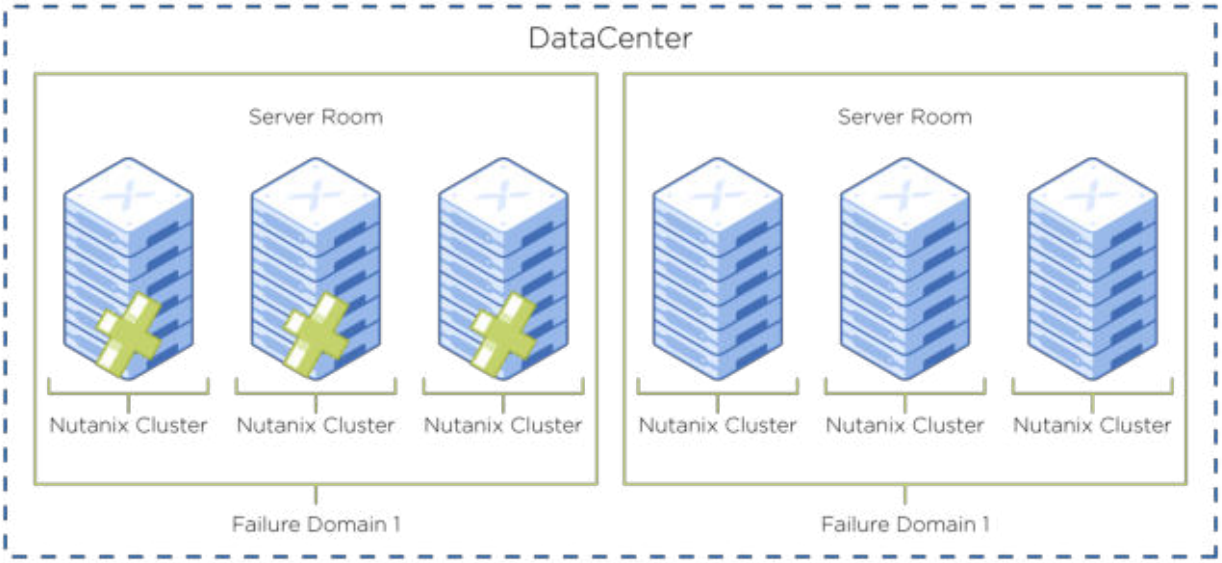


Figure 26: Datacenter Building Failure Domain

Designing Datacenter Locations

Based on typical enterprise requirements the logical implementation for DR across regions, availability zones and datacenters will use the following layout.

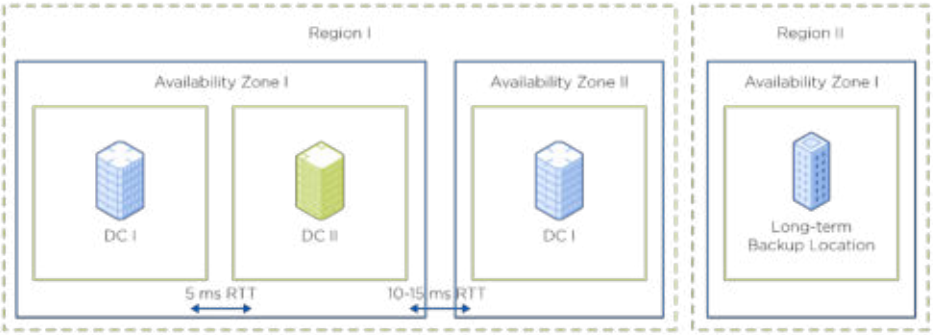


Figure 27: Designing Datacenter Locations

Regions

Table 26: Two region implementation for DR

REGION-001	TWO REGIONS ARE USED, SUCH AS: REGION 1 (PRIMARY REGION) RUNS ACTIVE WORKLOADS AND PROVIDES TWO AVAILABILITY ZONES FOR DISASTER RECOVERY AND REGION 2 (SECONDARY REGION) PROVIDES A LOCATION FOR LONG-TERM BACKUP RETENTION.
Justification	Regional separation of running workloads and disaster recovery capabilities plus long-term backup retention.
Implication	It may take a long time to restore services in or from Region 2 when Region 1 fails but the likelihood of an entire region failing is low.

Table 27: Single region implementation for DR

REGION-002	ONE REGION HOSTS PRIMARY AND DISASTER RECOVERY AVAILABILITY ZONES.
Justification	• Addresses the requirement to have active workloads and disaster recovery capabilities in a region separate from where the long-term backups are stored. • Services must be available in one region only based on end-user location • Staff only available in one region • Budget constraints • One region can host multiple datacenters within 100 km and provide RTT of maximum 5ms which will be required to meet infrastructure RPO 0. • One region can host datacenter separation of more than 200 KM which is required for disaster recovery purposes.

REGION-002	ONE REGION HOSTS PRIMARY AND DISASTER RECOVERY AVAILABILITY ZONES.
Implication	• If Region 1 fails, the services has to be rebuilt from backups from outside Region 1. • Staff has to travel to Region 2 to build up the services from backups, or backups must be relocated to Region 1 when Region 1 becomes available after the failure, and the services must be rebuilt/restored.

Table 28: Long term retention backups are placed in Region 2

REGION-003	LONG TERM RETENTION BACKUPS ARE PLACED IN REGION 2.
Justification	Addresses the requirement to have long-term backups stored in a separate region from active workloads and provides disaster recovery capabilities in case of regional disasters.
Implication	• If Region 1 fails, the services has to be rebuilt from backups from outside Region 1. • Staff has to travel to Region 2 to build up the services from backups, or backups must be relocated to Region 1 when Region 1 becomes available after the failure, and the services must be rebuilt/restored.

Availability Zones

Table 29: Three availability zones in two regions

AZ-001	THREE AVAILABILITY ZONES IN TWO REGIONS WILL BE USED, SUCH AS: DC1 AND DC2 IN REGION 1 RUNNING WORKLOADS, AZ2 IN REGION 1 FOR DISASTER RECOVERY PURPOSES, AND AZ1 IN REGION 2 WHERE LONG-TERM BACKUPS ARE STORED.
Justification	Required to provide the separation required

© 2021 Nutanix, Inc. All rights reserved

AZ-001	THREE AVAILABILITY ZONES IN TWO REGIONS WILL BE USED, SUCH AS: DC1 AND DC2 IN REGION 1 RUNNING WORKLOADS, AZ2 IN REGION 1 FOR DISASTER RECOVERY PURPOSES, AND AZ1 IN REGION 2 WHERE LONG-TERM BACKUPS ARE STORED.
Implication	• Many availability zones to manage. • Multiple network connections to manage between availability zones.

Table 30: Two availability zones in Region 1

AZ-002	TWO AVAILABILITY ZONES IN REGION 1 WILL HOST PRIMARY WORKLOADS.REGION 2 DR SITE WILL HOST ANY WITNESS FUNCTIONALITY REQUIRED FOR SERVICE AVAILABILITY.
Justification	• Addresses the requirement to have applications available across multiple availability zones during normal production • Distance is more than 100km and provides RTT of less than 5ms meaning it is technically possible to meet an RPO = 0 requirement.
Implication	Network connections between availability zones are critical for service availability and are also exposed to more risk compared to intra-availability-zone connections.

Table 31: One availability zone in Region 1 to provide disaster recovery

AZ-003	ONE AVAILABILITY ZONE IN REGION 1 WILL BE USED TO PROVIDE DISASTER RECOVERY CAPABILITIES.
Justification	Addresses the requirement to provide disaster recovery across availability zones within the same region.
Implication	No availability zone redundancy during a disaster recovery scenario.

© 2021 Nutanix, Inc. All rights reserved

Table 32: One availability zone in Region 2 hosts backups.

AZ-004	ONE AVAILABILITY ZONE IN REGION 2 HOSTS LONG-TERM BACKUPS.
Justification	Addresses the requirement to have long-term backups stored in a separate region from running workloads, and disaster recovery capabilities are provided.
Implication	Staff has to travel to Region 2 to build up the services from backups, or backups must be relocated to Region 1 when Region 1 becomes available after the failure, and the services must be rebuilt/restored.

Designing Workload Domains

This document uses workload domains as building blocks. Each workload domain consists of a set of Nutanix nodes that are managed by the same management instance (Nutanix Prism Central/VMware vCenter) and connected to the same network domain.

Workload Domain Architecture

A single workload domain can include different hardware and software combinations, have homogeneous ESXi or AHV nodes, or mix ESXi (running VMs) and AHV (not running VMs) in a single cluster. These can be configured to satisfy redundancy, performance, and capacity requirements.

Normally, single workload domains occupy a single rack. However, you can aggregate multiple workload domains in a single rack or span a single workload domain across multiple racks.

Workload Domain Rack Mapping

Mapping workload domains to datacenter racks is not one to one. While the workload domain is a repeatable building block, a rack is a unit of size. Workload domains and datacenter racks can have different characteristics; you map workload domains to datacenter racks according to the use case and rack specifications.

Nutanix Cluster Layout

Single Rack, Single Workload Domain

One workload domain can occupy a single datacenter rack. All nodes from a workload domain are connected to a single pair of TOR switches.

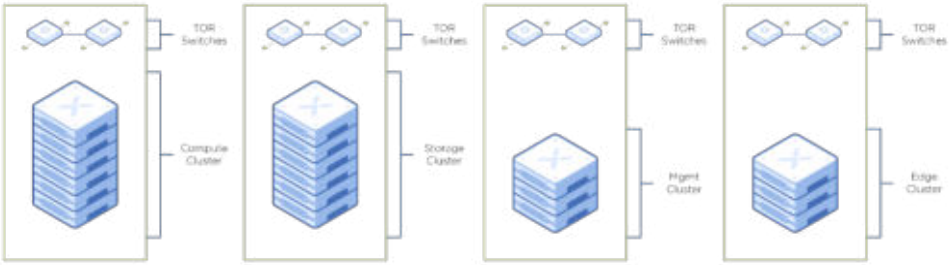


Figure 28: Single Rack Single Workload Domain

Single Rack, Multiple Workload Domains

One workload domain can occupy a single datacenter rack. All nodes from a workload domain are connected to a single pair of TOR switches.

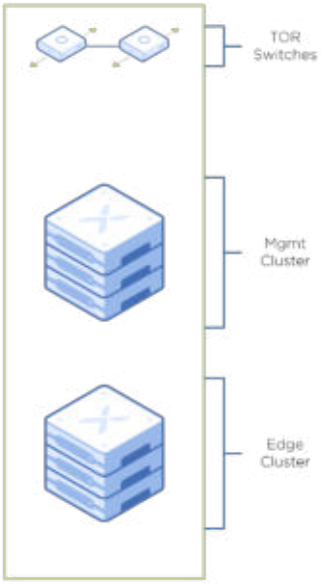


Figure 29: Single Rack, Multiple Workload Domains

Multiple Racks, Single Workload Domain

A single workload domain can span multiple racks. For example, to provide an additional level of data protection (using rack awareness/fault tolerance, see

the section Data Redundancy and Resiliency) or if a single workload domain is bigger than a single rack can contain.

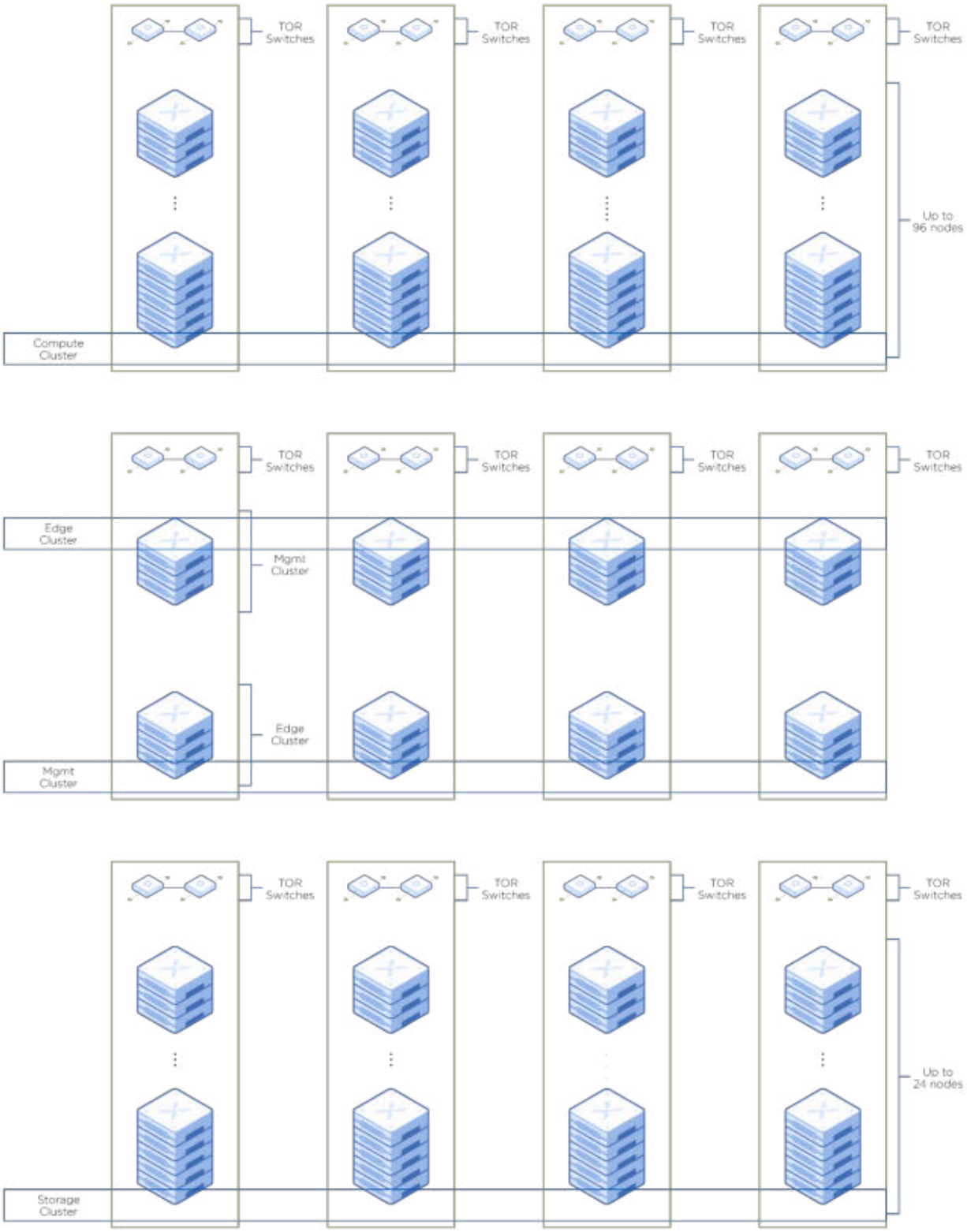


Figure 30: Multiple Racks, Single Workload Domain

Pros and Cons

A single workload domain can span multiple racks. For example, to provide an additional level of data protection (using rack awareness/fault tolerance, see the section:

Data Redundancy and Resiliency) or if a single workload domain is bigger than a single rack can contain.

Table 33: Pros and cons

	PROS	CONS
Single Rack/	• Default redundancy level. • Simple to consume.	May use rack space inefficiently.
Single Workload Domain	• Efficient space usage. • Default redundancy level. • Simple to consume.	Increases impact of rack failure.
Single Rack/ Multiple Workload Domains	• Efficient space usage. • Increases resiliency for each workload domain. • Decreases impact of rack failure.	Complexity.

Table 34: Workload Domains Spanning Single or Multiple Racks

PFM-007	DECIDE WHICH WORKLOAD DOMAINS WILL SPAN A SINGLE OR MULTIPLE RACKS
Justification	
Implication	

Networking

Well-designed networks are critical to a Nutanix deployment's resilience and performance.

A Nutanix cluster can tolerate multiple simultaneous failures because it maintains a set redundancy factor and offers features such as block and rack awareness. However, this level of resilience requires a highly available, redundant network connecting a cluster's nodes. Protecting the cluster's read and write storage capabilities also requires highly available connectivity between nodes. Even with intelligent data placement, if network connectivity between more than the allowed number of nodes breaks down, VMs on the cluster could experience write failures and enter read-only mode.

To optimize I/O speed, Nutanix clusters choose to send each write to another node in the cluster. As a result, a fully populated cluster sends storage replication traffic in a full mesh, using network bandwidth between all Nutanix nodes. Because storage write latency directly correlates to the network latency between Nutanix nodes, any increase in network latency adds to storage write latency.

Physical Switches

A Nutanix environment should use datacenter switches designed to handle high- bandwidth server and storage traffic at low latency. Do not use switches meant for deployment at the campus access layer. Campus access switches may have 10 Gbps ports like datacenter switches, but they are not usually designed to transport a large amount of bidirectional storage replication traffic. Refer to the Nutanix physical networking best practices guide for more information.

The deployment size and purpose also influence physical switch choice. Datacenter switches with large buffers are critical in a large AOS cluster that is going to grow beyond eight nodes or host storage-intensive applications. In smaller clusters or ROBO deployments that have fewer than eight nodes or do not host write-intensive applications, the switch may not experience buffer contention, and you can relax these switch restrictions. There are also some switch types you should never use for any Nutanix deployment because of

oversubscription or other architecture choices; we list examples of these as well.

Datacenter switches should have the following characteristics:

- Line rate: Ensures that all ports can simultaneously achieve advertised throughput.
- Low latency: Minimizes port-to-port latency as measured in microseconds or nanoseconds.
- Large per-port buffers: Accommodates speed mismatches from uplinks without dropping frames.
- Nonblocking, with low or no oversubscription: Reduces chance of drops during peak traffic periods.
- 10 Gbps or faster links for Nutanix CVM traffic: Only use 1 Gbps links for additional user VM traffic or when 10 Gbps connections are not available, such as in a ROBO deployment. Limit Nutanix clusters using 1 Gbps links to eight nodes maximum.

Switch manufacturers' datasheets, specifications, and white papers can help identify these characteristics. For example, a common datacenter switch datasheet may show a per-port buffer of 1 MB, while an access layer or fabric extension device has a per-port buffer of around 150 KB. During periods of high traffic, or when using links with a speed mismatch (such as 40 Gbps uplinks to 10 Gbps edge ports), a smaller buffer can lead to frame drops, increasing storage latency.

The following table is not exhaustive, but it includes examples of model lines that meet the above requirements for high-performance or large clusters. Models similar to the ones shown are also generally good choices.

Table 35: Examples of Recommended Switch Models

EXAMPLES OF HIGH PERFORMANCE SWITCH MODELS		
Arista 7050X3	Arista 7160	Arista 7170
Arista 7280	Cisco Nexus 9000	Cisco Nexus 7000

Cisco Nexus 7000	Dell S5200-ON	HPE FM3810
HPE FM3132Q	Juniper QFX-5100	Lenovo NE2580O
Mellanox SN2010	Mellanox SN2100	Mellanox SN2410

The following are examples of switches that do not meet high-performance datacenter switch requirements but are acceptable for ROBO clusters and clusters with fewer than eight nodes or low performance requirements:

Table 36: Examples of Robo Or Smb Switch Models

EXAMPLES OF ROBO OR SMB SWITCH MODELS		
Arista 7050	Arista 7150S	Cisco Nexus 3000
Cisco Catalyst 9000	Cisco Catalyst 3000	HPE FM2072

The following are examples of switches that are never acceptable for any Nutanix deployment:

Table 37: Switches that are not acceptable

SWITCH	REASON NOT RECOMMENDED
Cisco Nexus 2000 (Fabric Extender)	Highly oversubscribed with small per-port buffers.
10 Gbps expansion cards in a 1 Gbps access switch	10 Gbps expansion cards provide uplink bandwidth for the switch, not server connectivity.

Each Nutanix node also has an out-of-band connection for IPMI, iLO, iDRAC, or similar management. Because out-of-band connections do not have the same latency or throughput requirements of VM or storage networking, they can use an access layer switch.

Note: Nutanix recommends an out-of-band management switch network separate from the primary network to ensure management availability. Configure server-facing ports in the management network as access ports and do not use VLAN trunking for these ports. Access to this critical management network should be restricted.

© 2021 Nutanix, Inc. All rights reserved

Table 38: Using a Large Buffer Datacenter Switch at 10GBPS or Faster

NET-001	USE A LARGE BUFFER DATACENTER SWITCH AT 10GBPS OR FASTER
Justification	Achieves high performance for critical storage and VM traffic converged on the same network fabric.
Implication	Requires data center switches that may be more expensive than campus or access layer switches.

Network Topology

In a greenfield environment, Nutanix recommends a leaf-spine network topology because it is easy to scale, achieves high performance with low latency, and provides resilience. A leaf-spine topology requires at least two spine switches and two leaf switches. Every leaf connects to every spine using uplink ports.

There are no connections between the spine switches or between the leaf switches in a conventional leaf-spine design. To form a Nutanix cluster it's critical that all nodes are in the same broadcast domain, thus in any leaf-spine design all leaf switches connecting nodes in a cluster should carry the Nutanix VLAN. This can be accomplished with physical connections between switches, using an overlay network, or using a pure layer 2 design. The following example shows a pure layer 2 design or overlay.

© 2021 Nutanix, Inc. All rights reserved

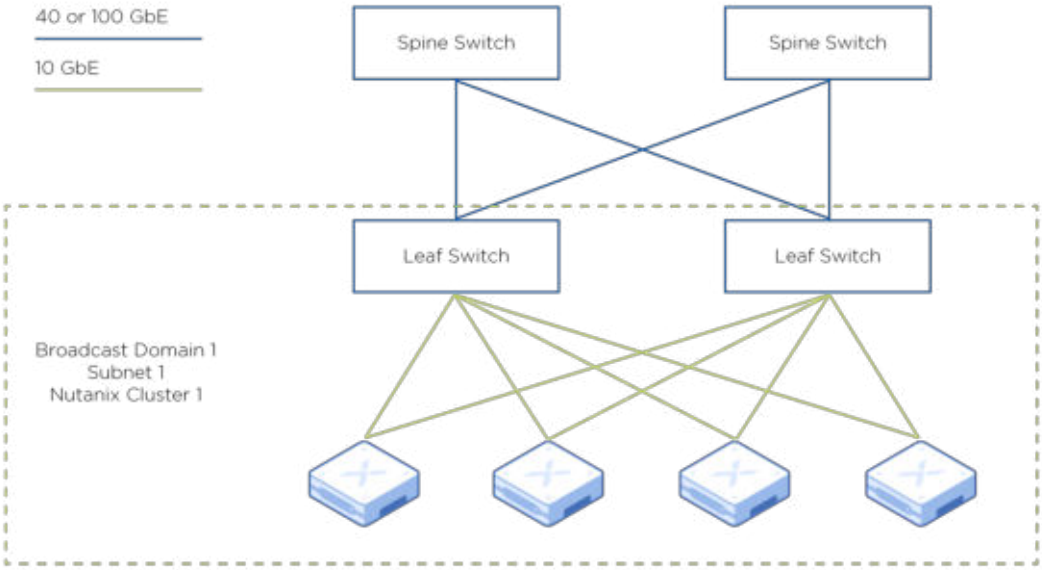


Figure 31: Pure Layer 2 Design

You may also choose a leaf-spine topology that uses links between switches to guarantee layer 2 connectivity between Nutanix nodes.

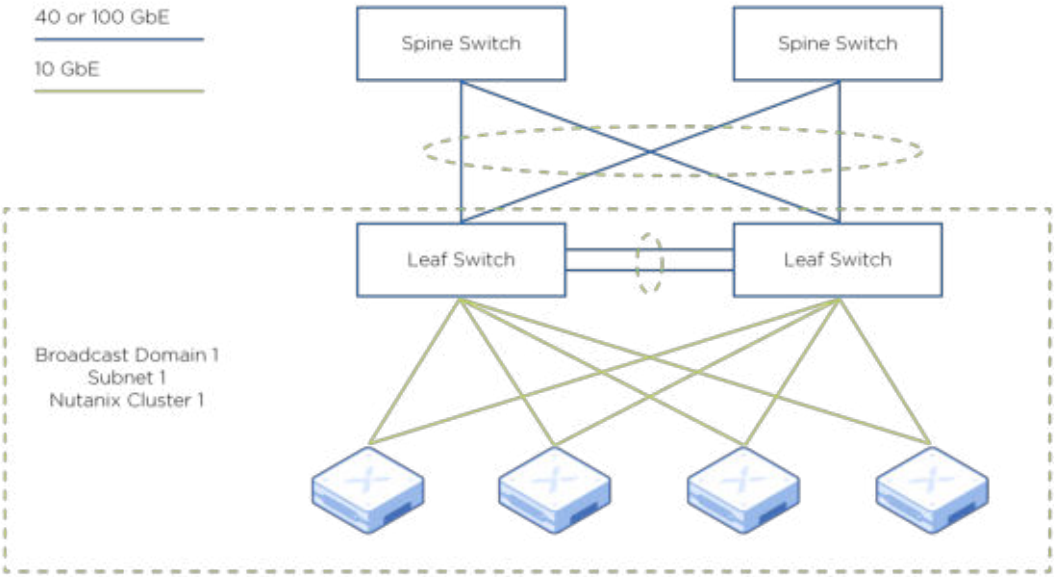


Figure 32: Leaf Spine Topology

Use uplinks that are a higher speed than the edge ports to reduce uplink oversubscription. To increase uplink capacity, add spine switches or uplink ports as needed.

The core-aggregation-access (or multi-tier) network design is a modular layout that allows you to upgrade and scale layers independently. Nutanix clusters perform well in the core-aggregation-access topology, but extra caution should be taken around scaling the Nutanix cluster.

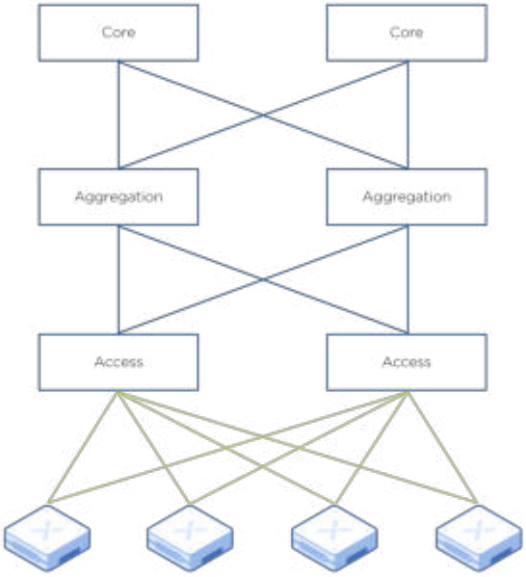


Figure 33: Core Aggregation Access Design

In pre-existing environments, you may not have full control over the network topology, but your design should meet the following requirements:

Guidelines:

- Networks must be highly available and tolerate individual device failures.
- Ensure that each layer of the network topology can tolerate device failure.
- Avoid configurations or technologies that do not maintain system availability during single device outages or upgrades such as stacked switches.
- Ensure that there are no more than three switches between any two Nutanix nodes in the same cluster. Nutanix nodes send storage replication traffic to

each other in a distributed fashion over the top-of-rack network. One Nutanix node can therefore send replication traffic to any other Nutanix node in the cluster.

- The network should provide low and predictable latency for this traffic. Leaf-spine networks meet this requirement by design. For the core-aggregation-access model, ensure that all nodes in a Nutanix cluster share the same aggregation layer to meet the three-switch-hop rule.

Oversubscription occurs when an intermediate network device or link does not have enough capacity to allow line rate communication between the systems connected to it. For example, if a 10 Gbps link connects two switches and four hosts connect to each switch at 10 Gbps, the connecting link is oversubscribed. Oversubscription is often expressed as a ratio—in this case 4:1, as the environment could potentially attempt to transmit 40 Gbps between the switches with only 10 Gbps available. Achieving a ratio of 1:1 is not always feasible.

Recommendation 1

- Keep the oversubscription ratio as small as possible based on budget and available capacity.

In a typical deployment where Nutanix nodes connect to redundant top-of-rack switches, storage replication traffic between CVMs traverses multiple devices.

Recommendation 2

- To avoid packet loss due to link oversubscription, ensure that the switch uplinks consist of multiple interfaces operating at a faster speed than the Nutanix host interfaces. For example, for nodes connected at 10 Gbps, interswitch connections should consist of multiple 10, 40, or 100 Gbps links.
- Connect all Nutanix nodes that form a cluster to the same switch fabric. Do not stretch a single Nutanix cluster across multiple, disconnected switch fabrics. A switch fabric is a single leaf-spine topology or all switches connected to the same switch aggregation layer. Every Nutanix node in a cluster should therefore be in the same L2 broadcast domain and share the same IP subnet.

© 2021 Nutanix, Inc. All rights reserved

See section Security > Network Segmentation for service placement information and design decisions.

Recommendation 3

- Use native, or untagged, VLANs for the hypervisor host and CVM for ease of initial configuration. Ensure that this untagged traffic is mapped into the CVM and hypervisor VLAN only on the required switch ports to reduce risk.
- Use tagged VLANs for all guest VM traffic and add the required guest VM VLANs to all connected switch ports for hosts in the Nutanix cluster.
- Limit guest VLANs for guest VM traffic to the smallest number of physical switches and switch ports possible to reduce broadcast network traffic load.

Table 39: Using a Leaf-Spine Network Topology for New Environments

NET-002	USE A LEAF-SPINE NETWORK TOPOLOGY FOR NEW ENVIRONMENTS.
Justification	Achieves high performance for critical storage and VM traffic and is easy to scale.
Implication	Requires more network connections between switches and may require new network design.

Table 40: Populate Each Rack with Two 10GBE or Faster TOR Switches

NET-003	POPULATE EACH RACK WITH TWO 10GBE OR FASTER TOR SWITCHES
Justification	Simplifies the design, follows the leaf/spine model, and provides high performance and high availability to the network.
Implication	Increases rack space requirements, costs, and the number of leafs.

Table 41: Avoid Switch Stacking to Ensure Network Availability During Individual Device Failure

NET-004	AVOID SWITCH STACKING TO ENSURE NETWORK AVAILABILITY DURING INDIVIDUAL DEVICE FAILURE.

© 2021 Nutanix, Inc. All rights reserved

Justification	Critical network infrastructure must be redundant between all CVMs in the same Nutanix cluster.
Implication	More network devices may be required to provide high availability and reduce single points of failure.

Table 42: Use No More Than Three Switches Between Two Nutanix Nodes in the Same Cluster

NET-005	ENSURE THAT THERE ARE NO MORE THAN THREE SWITCHES BETWEEN ANY TWO NUTANIX NODES IN THE SAME CLUSTER.
Justification	Storage latency is directly related to network latency, so the network distance between nodes in the same cluster must be reduced.
Implication	A Nutanix cluster cannot span multiple sites or switch fabrics, instead use data replication technologies instead to provide high availability between sites.

Table 43: Reduce Network Oversubscription to Achieve a Near-1:1 Ratio

NET-006	REDUCE NETWORK OVERSUBSCRIPTION TO ACHIEVE AS CLOSE TO A 1:1 RATIO AS POSSIBLE.
Justification	Dropped network packets or a congested network will immediately impact storage performance and must be avoided in the design phase.
Implication	Higher bandwidth paths must be created using more uplink paths of faster speed between switches.

Table 44: Configure CVM and Hypervisor VLAN on Server-facing Switch Ports

NET-007	CONFIGURE THE CVM AND HYPERVISOR VLAN AS NATIVE, OR UNTAGGED ON SERVER FACING SWITCH PORTS.
Justification	Newly added nodes use untagged traffic for discovery and will work out of the box, reducing manual server configuration.
Implication	Network switches must be configured to accept untagged frames on ports facing Nutanix servers and place the traffic into the CVM and hypervisor VLAN.

© 2021 Nutanix, Inc. All rights reserved

Table 45: Use Tagged VLANs on Switch Ports for All Guest Workloads

NET-008	USE TAGGED VLANS ON THE SWITCH PORTS FOR ALL GUEST WORKLOADS.
Justification	Workloads should be separated from each other and from the CVM and hypervisor network using VLANs.
Implication	Multiple VLANs and IP subnets are required.

Broadcast Domains

Performing layer 3 routing at the top of rack, creating a layer 3 and layer 2 domain boundary within a single rack, is a growing network trend. Each rack is a different IP subnet and a different layer 2 broadcast domain. This layer 3 design decreases the size of the layer 2 broadcast domain to remove some common problems of sharing a large broadcast domain among many racks of servers, but it can add complexity for applications that require layer 2 connectivity.

In contrast, the traditional layer 2 design shares a single broadcast domain or VLAN among many racks of switches. In the layer 2 design, a switch and server in one rack share the same VLANs as a switch and server in another rack. Routing between IP subnets is performed either in the spine or in the aggregation layer. The endpoints in the same switch fabric have layer 2 connectivity without going through a router, but this can increase the number of endpoints that share a noisy broadcast domain.

© 2021 Nutanix, Inc. All rights reserved

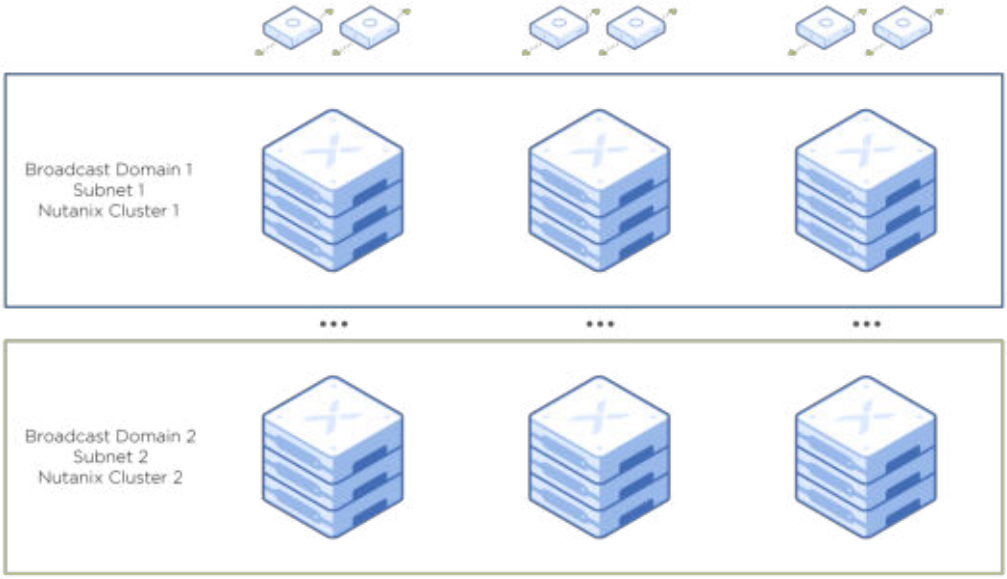


Figure 34: Layer 2 Network Design

Nutanix recommends a traditional layer 2 network design to ensure that CVMs and hosts can communicate in the same broadcast domain even if they are in separate racks. The CVM and host must be in the same broadcast domain and IP subnet.

If a layer 3 network design is chosen, there are two possible ways to make this work with a Nutanix deployment:

- Keep all Nutanix nodes in a cluster inside the same rack.
- Create an overlay network in the switch fabric that creates a virtual broadcast domain that is shared across racks for the Nutanix CVMs and hosts.

Keeping all of the Nutanix nodes in a single cluster in a single rack has the limitation of losing resilience against rack failure. Creating an overlay in the network fabric:

- Requires special configuration.
- May require specific network hardware to achieve.
- Increases configuration complexity.

Spreading a broadcast domain across racks virtually may also raise concerns that the original intent of the layer 3 design, reducing broadcast domain scope, is now violated.

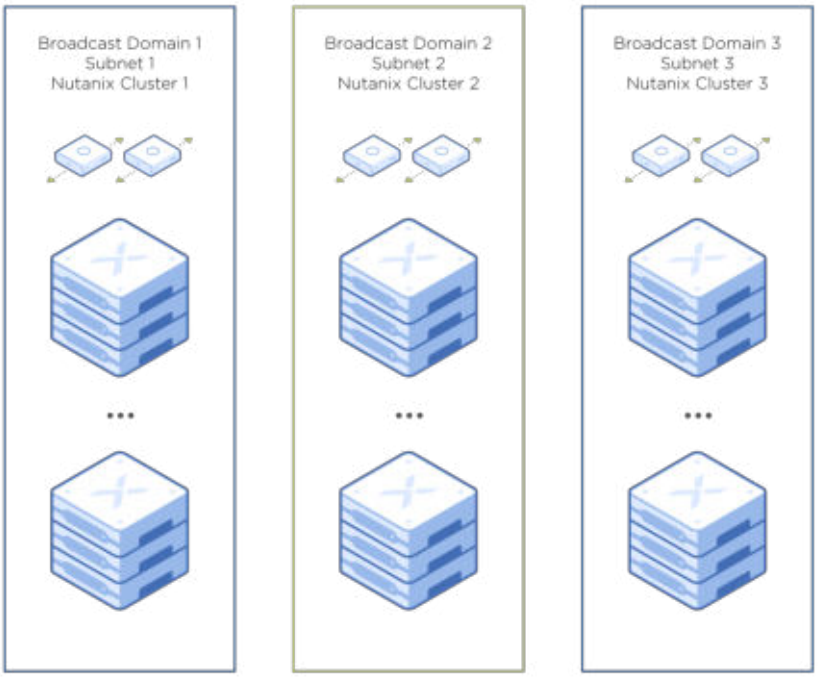


Figure 35: Layer 3 Network Design

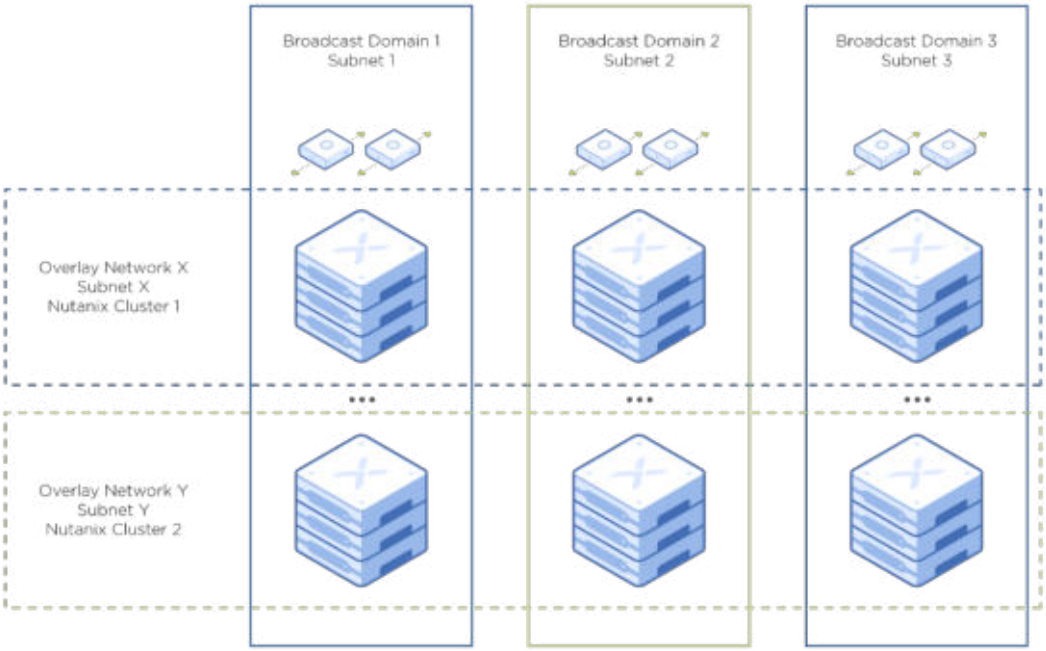


Figure 36: Layer 3 Network Design with Overlay

Table 46: Use a Layer 2 Network Design

NET-009	USE A LAYER 2 NETWORK DESIGN
Justification	Nutanix CVMs and hypervisor hosts must be in the same broadcast domain to function properly after a CVM becomes unavailable.
Implication	A broadcast domain failure or storm, can span multiple racks.

Scaling the Network

To scale the leaf-spine network:

- Add two leaf switches and a management switch for every rack.
- Add spine switches as needed.

Because there are no more than three switch hops between any two Nutanix nodes in this design, a Nutanix cluster can easily span multiple racks and still use the same switch fabric. Use a network design that brings the layer 2

broadcast domain to every switch that connects Nutanix nodes in the same cluster.

The following example shows a leaf-spine network using either an overlay or a pure layer 2 design, with no connections are required between leaf switches.

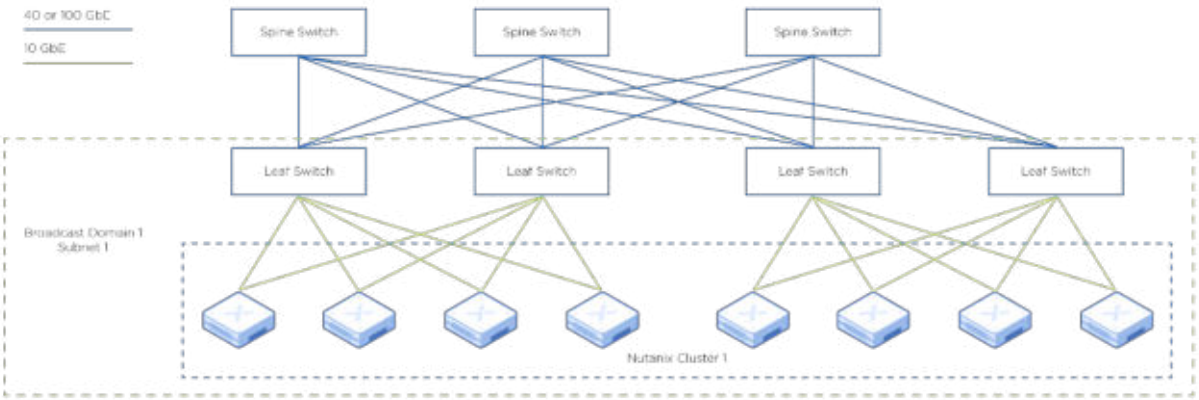


Figure 37: Logical View of Leaf Spine Scale

You may choose a leaf-spine design where a pair of leaf switches are connected using link aggregation. Discuss scale plans with the switch vendor when scaling beyond two spine switches in this design.

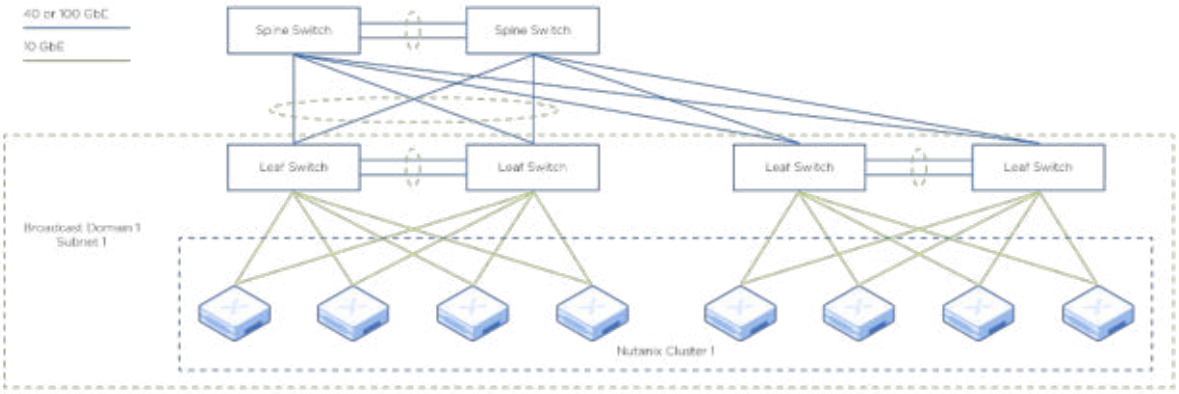


Figure 38: Leaf Switches Connected Using Link Aggregation

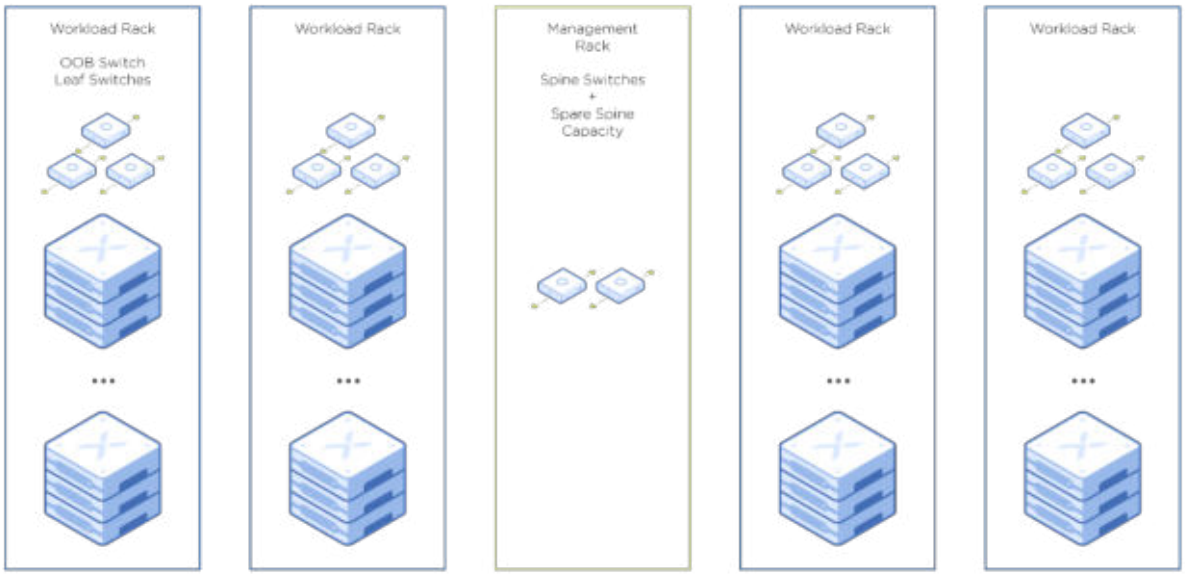


Figure 39: Rack Level View of Leaf Spine

Determining when to add new spine switches is a function of:

- How many ports the spine switch has.
- How many leaf switches (or racks) there are.
- How many connections are formed between each leaf and spine.

It's also important to take into account the throughput capacity of the switch and the throughput required between each rack.

In the previous example with 4 workload racks, assume you have 2 x 32-port spine switches with a single 100 Gbps connection to every leaf. Also assume each TOR leaf switch has 48 x 10 Gbps ports and 6 x 100 Gbps ports.

Spine switch port utilization is at 8 out of 32 ports on each spine, leaving capacity to grow up to 16 racks at each spine (32 spine ports divided by 2 ports per rack).

Each leaf would be using 1 out of the 6 available uplinks and each leaf can support up to 48 connected servers. However, at this point these two spine switches will be processing a lot of traffic and may be overloaded depending on their specifications.

Calculating oversubscription ratios, if you assume that each rack has 24 x 10Gbps dual-connected servers with dual leaf switches, for a total of 240 Gbps bandwidth capacity at each leaf, oversubscription is 2.4:1. This design is oversubscribed and would not be recommended.

One way to reduce oversubscription is to add another 100Gbps uplink from each leaf to the existing spine switches. However, that reduces the total number of supported racks by half to 8, (32 spine ports divided by 4 ports per rack), so you must carefully consider which leaf and spine switches are selected and what scale you would like to achieve.

Another way to grow spine capacity is to add a spine switch, which would then require another uplink from every leaf, and greatly increase the total throughput of the switch fabric without reducing the number of racks supported. If you added another spine switch to bring the total to 3 spines, each leaf switch would add another 100Gbps uplink (reducing oversubscription), but each spine could still support 16 racks (32 spine ports divided by 2 ports per rack). With 6 uplinks on every leaf switch, you can support designs with up to 6 spines in this example.

Scaling the multi-tier network design may require adding another aggregation and access layer to the core. In this case, there would be more than three switch hops between the two access layers.

> Note: The guideline here is to ensure that you add Nutanix nodes in separate aggregation and access layers to separate clusters to keep the number of switch hops between nodes in the same cluster to three or fewer.

In the following figure, Cluster 1 connects to one aggregation layer and Cluster 2 connects to another.

© 2021 Nutanix, Inc. All rights reserved

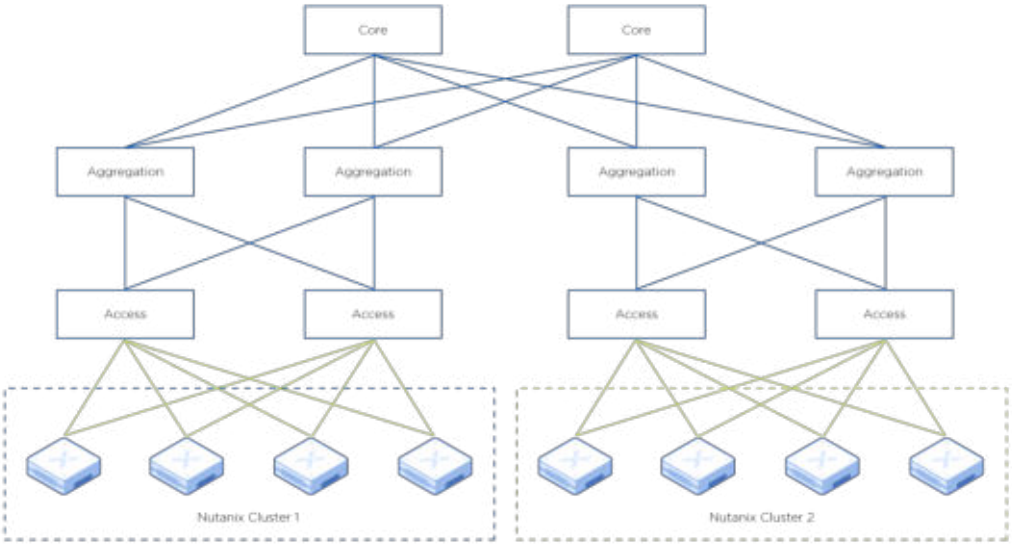


Figure 40: Clusters Connecting to Different Aggregation Layers

Host Networking

AHV Networking

AHV leverages Open vSwitch (OVS) for all VM networking. The virtual switch is referred to as a bridge and uses a br prefix in the name. Read the AHV Networking section on the Nutanix Support Portal for in depth guidance on any settings not covered here.

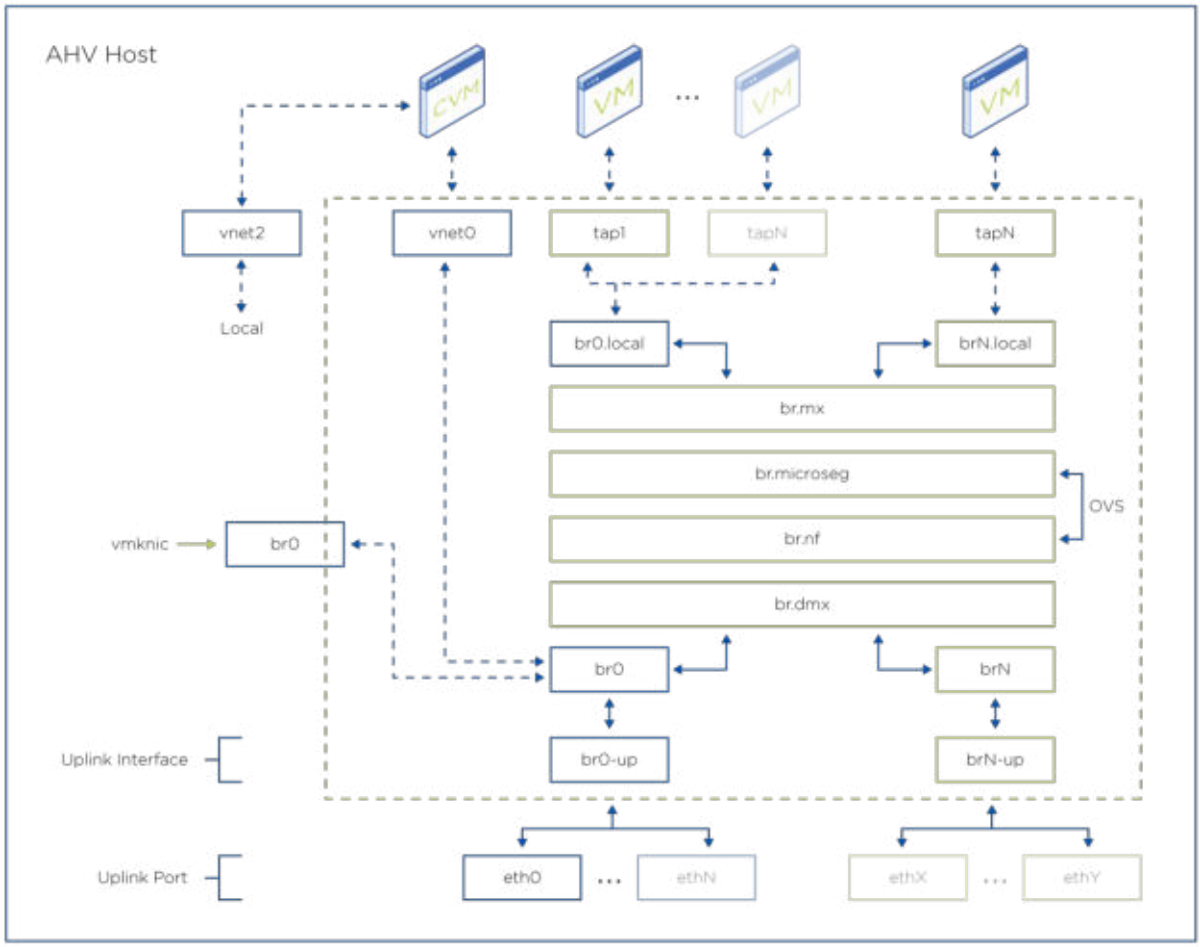


Figure 41: AHV Networking

Bridge

The default bridge, br0, behaves like a layer 2 learning switch that maintains a MAC address table. Additional uplink interfaces for separate physical networks are added as brN, where N is the new bridge number. To ensure that all VM traffic goes through the same set of firewall rules and the same set of network functions, a bridge chain is created with all microsegmentation and network function rules being placed in br. microseg and br.nf respectively.

Traffic from VMs enters the bridge at brN.local. Next, traffic is multiplexed onto the bridge chain, goes through the firewall and network function ruleset, and

is then demultiplexed on the correct brN for physical network forwarding. The reverse path is followed for traffic flowing from the physical network to VMs.

All traffic between VMs must flow through the entire bridge chain, since brN is the only bridge that performs switching. All other bridges simply apply rules or pass traffic up and down the bridge chain.

Recommendation

- Use only the default bridge, br0, with the two fastest network adapters in the host.
- Converge the management, storage, and workload traffic on this single pair of uplink adapters.
- Additional brN bridges should only be added when connection to a separate physical network is required. For example, if the top-of-rack has two pairs of switches, one pair for storage and management, and another pair for workload traffic, it makes sense to create another bridge, br1 and place the workloads on this bridge.
- Other valid use cases include separate physical networks for iSCSI Volumes workloads, backplane intracluster replication, or workloads like virtual firewalls that require access to a physically separate network. In these cases, the additional bridge is used to connect to the separate top-of-rack network. If there is only a single top-of-rack physical network, then a single bridge with VLAN separation is sufficient.

AHV also includes a Linux bridge called virbr0. The virbr0 Linux bridge carries management traffic between the CVM and AHV host. All other storage, host, and workload network traffic flows through the br0 OVS bridge, or additional brN bridges if configured.

> Note: Do not modify the configuration of any bridges inside the AHV host unless following an official Nutanix guide.

Bond

Bonded ports aggregate the physical interfaces on the AHV host. By default, a bond named br0-up is created in bridge br0. After the node imaging process, all interfaces are placed within a single bond, which is a requirement for the foundation imaging process. A bridge can have only a single bond.

Bonds allow for several load-balancing modes, including active-backup, balance-slb and balance-tcp. Link Aggregation Control Protocol (LACP) can also be activated for a bond for link aggregation. The “bond_mode” setting is not specified during installation and therefore defaults to active-backup, which is the default configuration.

Recommendation

Keep the following recommendations in mind for all bond scenarios to prevent undesired behavior and maintain NIC compatibility.

- Ensure that each bond has at least two uplinks for redundancy.
- Do not mix NIC models from different vendors in the same bond.
- Do not mix NICs of different speeds in the same bond.

Uplink Load Balancing

The following bond modes are available:

- active-backup: Default configuration which transmits all traffic over a single active adapter. If the active adapter becomes unavailable, another adapter in the bond will become active. Limits host throughput to the bandwidth of a single network adapter. No additional switch configuration required.
- balance-slb: Distributes VM NICs across adapters in the bond and periodically rebalances for even uplink utilization. Limits VM per-nic throughput to a single network adapter, but allows utilization of multiple host physical interfaces for multiple VM NICs. NOTE: Not recommended due to negative interaction with default IGMP snooping and pruning network settings for multicast.
- balance-tcp with LACP: Distributes VM NIC TCP or UDP session across adapters in the bond. Limits per-nic throughput to the maximum bond bandwidth (number of physical uplink adapters * speed). Requires physical switch link aggregation. Used when LACP negotiation is required by the datacenter team or throughput requirements exceed a single NIC.

Recommendation

- To keep network configuration simple, use the standard 1,500 byte MTU in the hosts, CVMs, and workload VMs. Nutanix does not recommend jumbo

© 2021 Nutanix, Inc. All rights reserved

frames unless specifically required by high-performance Nutanix Volumes iSCSI workloads or specific workload requirements.

- When switching from 1,500-byte frames to 9,000-byte frames, performance improvements are generally not significant unless the workload uses the maximum network bandwidth for read traffic.

For more information on when to use jumbo frames, see the Nutanix Volumes Guide and AHV Networking guides.

Table 47: Connecting NICs to TOR Switches

NET-010	CONNECT AT LEAST ONE 10 GBE OR FASTER NIC TO EACH TOR SWITCH
Justification	Maintains high availability in the event of the loss of one switch.
Implication	Requires at least two TOR switches, which will impact cost and rack space. Increases total bandwidth to each host to 20 Gbps.

Table 48: Using Bridges with Uplinks

NET-011	USE A SINGLE BR0 BRIDGE WITH AT LEAST TWO OF THE FASTEST UPLINKS OF THE SAME SPEED UNLESS MULTIPLE PHYSICAL NETWORKS ARE REQUIRED.
Justification	Simplifies the design, requiring fewer physical switches and switch ports, while relying on high bandwidth of 10Gbps and faster adapters.
Implication	All traffic is sent over a single virtual switch to a single pair of TOR switches.

Table 49: Using NICs of the same vendor within a bond

NET-012	USE NICs OF THE SAME VENDOR WITHIN A BOND
Justification	Ensures NIC compatibility within the bond and prevents undesired failover behavior.
Implication	NIC vendor choice must be considered when initially purchasing or adding NICs.

© 2021 Nutanix, Inc. All rights reserved

Table 50: Using VLANs to separate logical networks

NET-013	USE VLANS TO SEPARATE LOGICAL NETWORKS
Justification	Physical hosts have a limited number of network ports and each port adds complexity. Traffic separation can be logically accomplished without numerous physical ports.
Implication	Enables multiple networks to be used at once without additional hardware, saving on costs, and simplifying the design.

Table 51: Load Balancing

NET-014	USE ACTIVE-BACKUP UPLINK LOAD BALANCING
Justification	Simplifies the design and does not need any additional configuration.
Implication	All traffic is transmitted over one adapter, limiting bandwidth to at least 10 Gbps.

Table 52: Using Standard 1500 Byte MTU

NET-015	USE STANDARD 1,500 BYTE MTU AND DO NOT USE JUMBO FRAMES.
Justification	Simplifies the design and does not need any additional configuration. Reduces risk and complexity from non-standard configuration that only increases performance in certain use cases.
Implication	All network frames are limited to the default 1,500-byte maximum size for interoperability, potentially creating more network overhead for high throughput write workloads.

vSphere Networking

VMware vSphere networking follows many of the same design decisions as AHV networking. The critical design choices for vSphere networking are covered here. See the VMware vSphere Networking Guide for more details.

Nutanix hosts with vSphere ESXi use two virtual switches (vSwitches), named vSwitchNutanix and vSwitch0. vSwitchNutanix is the internal, standard vSwitch used for management and storage traffic between the CVM and the hypervisor.

Recommendations

- Do not modify vSwitchNutanix. vSwitch0 is also a standard vSwitch by default, used for communication between CVMs as well as workload traffic.
- Convert vSwitch0 to the distributed vSwitch following KB 4552 for Converting to the distributed vSwitch allows central management of networking for all hosts, instead of host by host networking configuration. The distributed vSwitch also adds capability for advanced networking functions such as load based teaming, LACP, and traffic shaping.
- Connect at least two of the fastest adapters of the same speed into vSwitch0 and use the "Route Based on Physical NIC Load" load balancing method to ensure traffic is balanced between uplink adapters.
- Connect these adapters to two separate top-of-rack switches to ensure redundancy.
- Do not add more vSwitches unless a connection to another physical network is needed to meet security or workload requirements.
- All CVM storage, hypervisor host, and workload traffic should flow through vSwitch0 using VLAN separate between the workload traffic and all other traffic.
- Use the default 1,500 byte frame size on all uplinks unless there is a specific performance or application requirement that would justify 9,000 byte jumbo frames.

Table 53: Use Virtual Distributed Switch (VDS)

NET-016	USE VIRTUAL DISTRIBUTED SWITCH (VDS)

Justification	Reduces the amount of configuration required in the environment. Makes port group provisioning faster and less error prone. Provides additional features such as LACP and bandwidth shaping. Important: Leave the internal vNetwork Standard Switch used by ESXi host and CVM for internal communication in place, including security settings.
Implication	Ability to configure port groups and other features at the cluster switch level vs. individual nodes. Reliance on vCenter for host and VM network configuration management.

Note: Leave the internal vNetwork Standard Switch used by ESXi host and CVM for internal communication in place, including security settings.

Table 54: Connecting NICs to the TOR Switch

NET-017	CONNECT AT LEAST ONE 10 GBE OR FASTER NIC TO EACH TOP-OF-RACK SWITCH
Justification	Maintains high availability in the event of the loss of one switch.
Implication	Requires at least two ToR switches, which will impact cost and rack space. Increases total bandwidth to each host to 20 Gbps.

Table 55: Using Virtual Switches and Uplinks

NET-018	USE A SINGLE VSWITCH0 WITH AT LEAST TWO OF THE FASTEST UPLINKS OF THE SAME SPEED
Justification	Simplifies the design, requiring fewer physical switches and switch ports, while relying on high bandwidth of 10Gbps and faster adapters.
Implication	All traffic is sent over a single virtual switch to a single pair of top-of-rack switches.

© 2021 Nutanix, Inc. All rights reserved

Table 56: Load Balancing

NET-019	USE ROUTE BASED ON PHYSICAL NIC LOAD UPLINK LOAD BALANCING
Justification	Achieves load balancing and use of both adapters with no switch side configuration required.
Implication	Requires the vSphere Distributed Switch to spread VM NICs among uplinks to reduce the most utilized physical network adapter.

Table 57: Using Standard 1500 Byte MTU

NET-020	USE STANDARD 1,500 BYTE MTU AND DO NOT USE JUMBO FRAMES.
Justification	Simplifies the design and does not need any additional configuration. Reduces risk and complexity from non-standard configuration that only increases performance in certain use cases.
Implication	All network frames are limited to the default 1,500-byte maximum size for interoperability, potentially creating more network overhead for high throughput write workloads.

Compute and Storage Design

Nutanix HCI is a converged storage and compute solution which leverages local hardware components (CPU, memory, and storage) and creates a distributed platform for running workloads.

Each node runs an industry-standard hypervisor and the Nutanix CVM. The Nutanix CVM provides the software intelligence for the platform. It is responsible for serving all IO to VMs that run on the platform.

The following logical diagram illustrates the relationship of the components of a Nutanix node:

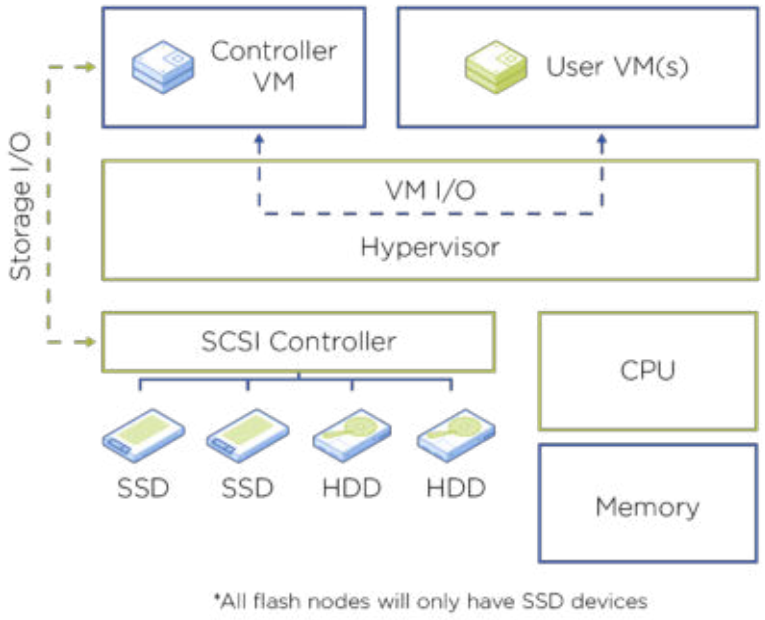


Figure 42: Relationship between the Components of the Nutanix Node

The CVM controls the storage devices directly and creates a logical construct called a storage pool of all disks from all nodes in the cluster. For AHV it uses PCI passthrough and for ESXi it uses VMDirectPathIO.

Compute Design

AHV CPU and Memory Planning

AHV lets you configure vCPUs (similar to sockets) and cores per vCPU (similar to CPU cores) for every VM. Except for the memory used by the CVM and the AHV software, all physical memory can be allocated for use by VMs. AHV guarantees memory allocated to each VM and doesn't overcommit or swap which can impact performance.

The amount of memory required by an AHV host varies based on a number of factors. The host uses from 2GB to 10GB of memory on hosts with 512GB memory.

NUMA and VMs

Processor sockets have integrated memory controllers. A VM has faster access to memory attached to the physical CPU where it is running versus memory attached to another CPU in the same host. This is called Non-Uniform Memory Access (NUMA). As a result, the best practice is to create VMs that fit within the memory available to the CPU where the VM is running whenever possible.

Table 58: Running Non-NUMA-aware Applications on a VM

CMP-001	IF RUNNING A NON-NUMA-AWARE APPLICATION ON A VM, CONFIGURE THE VM'S MEMORY AND VCPU TO FIT WITHIN A NUMA NODE ON AN AHV HOST.
Justification	This prevents the VM's vCPU and memory access from crossing NUMA boundaries.
Implication	Avoiding crossing NUMA boundaries results in better, more predictable performance.

For example, if a host has 2 CPU sockets with 8 cores each and 256GB memory, the host has 2 NUMA nodes, each with:

- 8 CPU cores
- 128GB memory

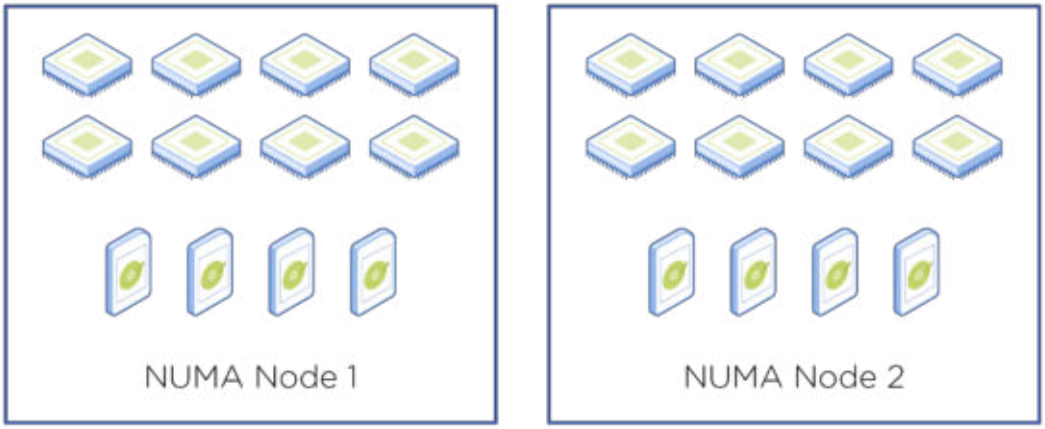


Figure 43: Example Host with 2 NUMA Nodes

For example, a VM that has 4 vCPUs and 64GB memory, will fit within a single NUMA node on the host and achieve the best performance.

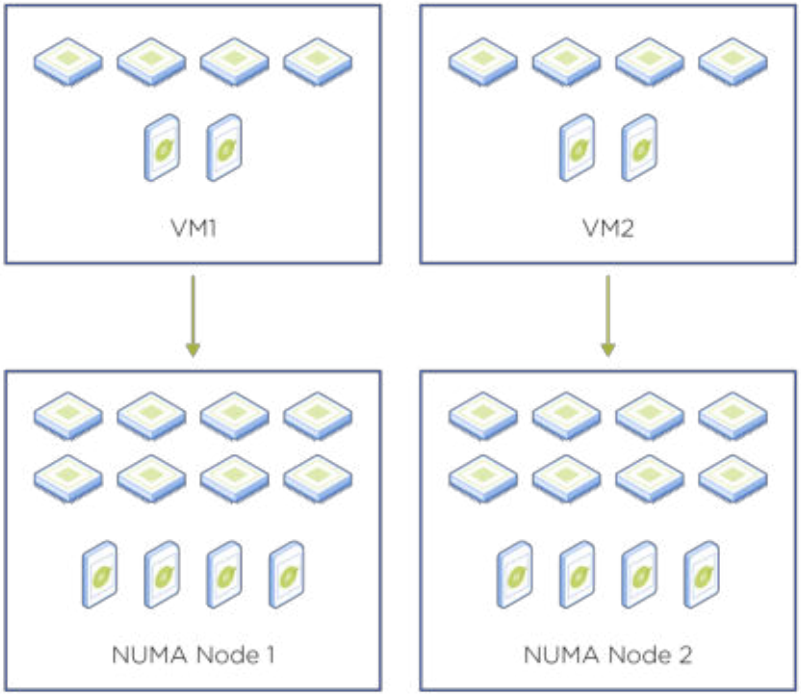


Figure 44: Example of a Single NUMA Node

Virtual Non Uniform Memory Access (vNUMA) and User VMs

The primary purpose of vNUMA is to give large virtual machines or wide VMs (VMs requiring more CPU or memory capacity than is available on a single NUMA node) the best possible performance. vNUMA helps wide VMs create multiple vNUMA nodes.

Nutanix supports vNUMA with both AHV and ESXi. vNUMA requires NUMA aware applications.

Each vNUMA node has virtual CPUs and virtual RAM. Pinning a vNUMA node to a physical NUMA node ensures that virtual CPUs accessing virtual memory see the expected NUMA behavior. Low-latency memory access in virtual hardware (within vNUMA) matches low-latency access in physical hardware (within physical NUMA), and high-latency accesses in virtual hardware (across vNUMA

boundaries) match high-latency accesses on physical hardware (across physical NUMA boundaries).

In AHV, administrators can configure vNUMA for a VM via the aCLI (Acropolis CLI) or REST API; this configuration is VM-specific and defines the number of vNUMA nodes. Memory and compute are divided in equal parts across each vNUMA node. Refer to AHV admin guide for steps on how to do it.

In ESXi, vNUMA is automatically enabled for a VM with >8 vCPU.

Building on the previous example, if a host with 16 CPU cores and 256GB memory has a VM with 12vCPUs and 192GB memory that is not vNUMA configured, the vCPU and memory assignment will span NUMA boundaries.

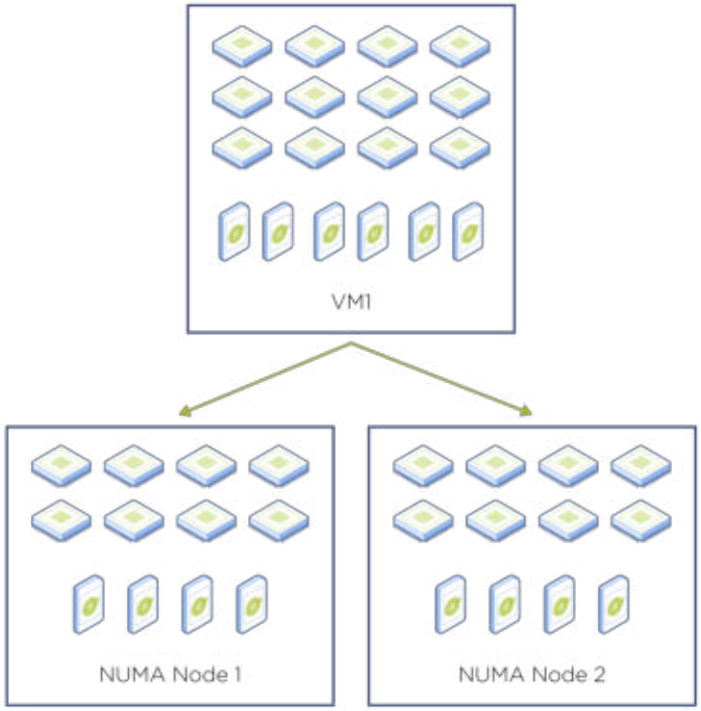


Figure 45: vCPU and Memory Assignment Spanning NUMA Boundaries

To ensure the best performance for wide VMs like this, vNUMA must be configured.

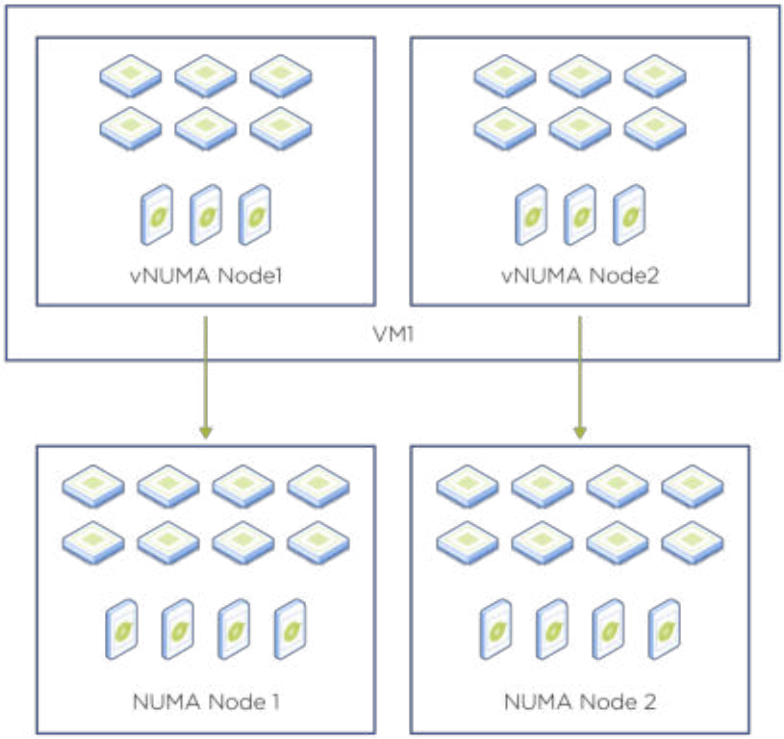


Figure 46: Best Performance vNUMA Configuration for Wide VMs

vNUMA VM configurations require strict fit. With strict fit, for each VM virtual node, memory must fit inside a physical NUMA node. Each physical NUMA node can provide memory for any number of vNUMA nodes. If there is not enough memory within a NUMA node, the VM does not power on. Strict fit is not required for CPU. To determine how many vNUMA nodes to use for a user VM, follow application-specific configuration recommendations provided by Nutanix.

CVM CPU and Memory Considerations

Optimal CVM vCPU values depend on the type of system and how many cores per NUMA node are present in the system. The CVM is pinned to the first physical CPU of the node. The following table gives an overview of minimum vCPU and memory values for the CVM (as of AOS 5.15):

Table 59: Minimum CPU and Memory Values as of AOS 5.15

Platform Storage or Type	Minimum Memory (GB)	Minimum vCPU
Platforms up to 80TB per node	32	8
Platforms up to 120TB per node	36	12
Platforms up to 120TB per node deploying Nutanix Objects	36	12

For specific recommendations on CVM sizing relating to specific workloads (e.g. Microsoft, Oracle, etc) please refer to Nutanix Solution Briefs and Best Practice Guides for those workloads.

Recommendations

Use Nutanix Foundation to set up a new cluster or node and configure the CVM(s). Nutanix Foundation software automatically identifies the platform and model type. Based on that information, Nutanix Foundation configures the appropriate default values for CVM memory and vCPU.

Storage Design

Nutanix HCI combines highly dense storage and compute into a single platform with a scale-out, shared-nothing architecture and no single points of failure. The AOS Distributed Storage Fabric (DSF) appears to the hypervisor like any centralized storage array, however all of the I/Os are handled locally to provide the highest performance.

Storage planning for this document, requires that you understand some of the high-level concepts for DSF:

Storage Pool

A group of physical storage devices within the cluster. This can include HDDs and both NAND and NVMe SSDs. The storage pool spans multiple Nutanix nodes and expands as the cluster expands. A default storage pool is created when a cluster is created.

Container

Logical segmentation of the Storage Pool is provided by containers, which contain a group of VMs or files (vDisks). Configuration options and data management services are configured at the container level. When a cluster is created, a default storage container is configured for the cluster.

Note: When VMware ESXi is used, containers correspond to NFS datastores.

Container Guidelines:

A key design question is how many containers to configure in a cluster:

- Nutanix typically recommends minimizing the number of containers created. For most use cases only 1 container is enough.
- Exceptions can be made if two or more applications running on the same cluster require different data reduction and replication factor values. In this case, additional containers can be created for those workloads.

Data reduction, including compression, dedupe, erasure coding and replication factor (RF) values, is configured on a container basis and multiple containers need to be created when different values are desired.

All Nutanix containers are thin provisioned by default. Thin provisioning is a widely accepted technology that has been proven over time by multiple storage vendors, including VMware. As the DSF presents containers to VMware vSphere hosts as NFS datastores by default, all VMs are also thin provisioned by default. vDisks created within AHV VMs are thinly provisioned by default. Nutanix also supports thick provisioned vDisks for VMs in ESXi by using space reservations. Using thick provisioned vDisks might impact data reduction values on containers since Nutanix reserves the thick provisioned space and it cannot be oversubscribed.

Table 60: Creating vDisks in ESXi

STR-01	WHEN CREATING VDISKS IN ESXI, ALWAYS USE THIN-PROVISIONED VDISKS.
Justification	There is no performance difference between thin-provisioned and thick-provisioned disks. Thin provisioning is more space efficient.
Implication	All Nutanix data reduction technologies can be used on the vDisk.

© 2021 Nutanix, Inc. All rights reserved

Note: Multi-writer VMDK requires thick provisioned vDisks.

vDisk

All storage management is VM-centric, and I/O is optimized at the vDisk level.

The software solution runs on nodes from a variety of manufacturers. A vDisk is any file over 512KB stored in DSF, including vmdks and VM hard disks. vDisks are logically composed of vBlocks.

Recommendations

Use multiple vDisks for an application versus a single large vDisk.

Using multiple vDisks allows for better overall performance due to the ability to leverage multiple OS threads to do more parallel processing versus a single vDisk configuration. Refer to the appropriate application solution guide for application-specific configuration recommendations for how many vDisks to configure.

Nutanix Volumes

In addition to providing storage through the hypervisor, Nutanix also allows supported operating systems, to access DSF storage capabilities directly using Nutanix Volumes. Nutanix designed Volumes as a scale-out storage solution where every CVM in a cluster can present storage volumes via iSCSI. This solution allows an individual application to access the entire cluster, if needed, to scale out for performance.

Volumes automatically manages high availability to ensure upgrades or failures are non- disruptive to applications. Storage allocation and assignment for Volumes is done with volume groups (VGs). A VG is a collection of one or more disks (vDisks) in a Nutanix storage container. These Volumes disks inherit the properties (replication factor, compression, erasure coding, and so on) of the container they reside in. With Volumes, vdisks in a VG are load balanced across all CVMs in the cluster by default.

In addition to connecting volume groups through iSCSI, AHV also supports direct attachment of VGs to VMs.In this, the vdisks are presented to guest OS over the virtual SCSI controller. The virtual SCSI controller leverages AHV Turbo and iSCSI under the covers to connect to the Nutanix DSF. By default,

© 2021 Nutanix, Inc. All rights reserved

the vdisks in a VG directly attached to a VM are hosted by the local CVM. Load balancing of vdisks on direct attached VGs can be enabled via acli.

Load balancing of vDisks in a VG enables I/O-intensive VMs to use the storage capabilities of multiple CVMs. If load balancing is enabled on a VG, AHV communicates directly with each CVM hosting a vDisk.

Each vDisk is served by a single CVM. Therefore, to use the storage capabilities of multiple CVMs, create more than one vDisk for a file system and use OS-level striped volumes to spread the workload. This improves performance and prevents storage bottlenecks, but can impact your network bandwidth.

Guideline: Core use-cases for Nutanix Volumes:

- Shared Disks
  - Oracle RAC, Microsoft Failover Clustering, etc.
- Where execution contexts are ephemeral, and data is critical
  - Containers, OpenStack, etc.
  - Guest-initiated iSCSI which is required for MS Exchange on vSphere (for Microsoft support).
- Bare-metal consumers
  - Stand alone or clustered physical workload.

Hybrid and All-Flash Nodes

A Nutanix cluster can consist of hybrid nodes, all-flash nodes, or a combination. Irrespective of the node type, each node has an OpLog that serves as a persistent write buffer for bursty, random write traffic. The OpLog is similar to a filesystem journal.

An extent store provides persistent bulk storage for DSF. A hybrid node has a combination of SSDs and HDDs with OpLog stored on the SSDs. All-Flash nodes are composed of SSDs. If there is a combination of SATA and NVMe SSDs, OpLog is stored on the fastest medium.

DSF automatically tiers data across the cluster to different classes of storage devices using intelligent lifecycle management (ILM). For best performance, ILM makes sure the most frequently used data is available in memory or in flash on

© 2021 Nutanix, Inc. All rights reserved

the node local to the VM. The default threshold where the algorithm will move data from the hot tier to the cold tier is 75%. Data for down-migration is chosen based on last access time. In All-Flash node configurations, the Extent Store consists only of SSDs and no tiering occurs.

Recommendations:

For hybrid configurations, Nutanix recommends:

- Sizing the SSD tier so that the active data set of application fits in 75% of usable SSD capacity.
- Nutanix has a collector tool that can be run to determine the approximate active data set size.
- Alternatively, use application-specific tools such as: MAP toolkit for MSSQL; AWR reports and Nutanix scripts for Oracle.
- A general rule is to check how much data has been backed up during a month. Assuming a 70%/30% R/W pattern (if the R/W patterns are unknown), multiply the data you get from backups by 4, which would give an approximate value for hot data. That amount should fit within 75% of the usable cluster capacity.

For All-Flash configurations there is no ILM, so there is no equivalent recommendation.

Table 61: SSD Capacity Considerations When Sizing a Cluster

STR-02	WHEN SIZING A HYBRID CLUSTER, MAKE SURE TO HAVE ENOUGH USABLE SSD CAPACITY TO CONTAIN THE ACTIVE DATA SET OF RUNNING APPLICATIONS.
Justification	Keeping active data on SSD ensures the application achieves the best performance. ILM moves data between tiers. The best-case scenario is active data in the SSD tier all the time.
Implication	More SSD capacity may need to be sized to keep active data on the SSD tier.

Recommendations

For hybrid configurations, Nutanix recommends:

- Use All-Flash node configurations for business-critical applications for consistent performance. As mentioned before,
- All-Flash configurations can consist of all SATA SSDs or a combination of SATA and NVMe.
- If the application requires very high IOPS with consistent low latency, there is an option to use RDMA with NVMe. RDMA requires RoCEv2 to be configured on the network fabric; Nutanix makes this setup very easy through Prism.

Table 62: Mixing Nodes from Different Vendors in the Same Cluster

STR-03	DO NOT MIX NODE TYPES FROM DIFFERENT VENDORS IN THE SAME CLUSTER.
Justification	Mixing node types from different vendors is unsupported.
Implication	Nodes from different vendors can be setup in separate clusters and all managed centrally from Prism Central.

Table 63: Mixing NVMe and Hybrid SSD/HDD Nodes in the Same Cluster

STR-04	DO NOT MIX NODES THAT CONTAIN NVME SSDS IN THE SAME CLUSTER WITH HYBRID SSD/HDD NODES.
Justification	The performance of NVMe SSD nodes is significantly higher compared to hybrid nodes and mixing them can potentially create performance issues.
Implication	NVMe Nodes can be setup in their own clusters or added to clusters with All SSD nodes and managed centrally from Prism Central.

Node Selection Recommendations

- If your high-level design decision is to use hybrid nodes, you need to decide the number of HDDs and SSDs per node.

© 2021 Nutanix, Inc. All rights reserved

- For applications that require high performance like DBs and Tier-1 applications, more drives per node is better.
  - OpLog will be spread across up to 8 SSDs, so more SSDs results in overall better and more consistent performance. If NVMe devices are present, OpLog is placed on them upto 8 NVMe SSDs.
  - More SSDs provides more usable space for tiering data.
- For All-Flash nodes, more SSDs provide better performance by making read and write access more parallel, especially for non-optimal workload patterns.
- Use Nutanix Sizer to size usable capacity and obtain node recommendations based on your workload use case.
- Also refer to specific application solution guides and/or best practice guides.

Table 64: HDD to SSD Ratio

STR-05	A MINIMUM 2:1 HDD TO SSD RATIO IS REQUIRED FOR HYBRID CLUSTERS.
Justification	This ensures there is enough bandwidth in the slower storage tier to absorb down migrations triggered by ILM.
Implication	If this rule is not followed, it may result in performance degradation during down migrations.

Note: 4 drive slot and 10 drive slot systems from Dell, Lenovo, and Fujitsu can have 2+2 and 4+6 SSD+HDD configurations respectively.

Table 65: Sizing for Redundancy

STR-06	SIZE FOR N+1 NODE REDUNDANCY FOR STORAGE AND COMPUTE WHEN SIZING. FOR MISSION CRITICAL WORKLOADS THAT NEED HIGHER SLAS, USE N+2 NODE REDUNDANCY.
Justification	Having the storage and compute capacity of an additional node (or two) ensures there is enough storage and compute capacity available for VMs to re-start when a node failure occurs.

© 2021 Nutanix, Inc. All rights reserved

Implication	Additional storage and compute capacity is required for N+1 or N+2 redundancy.

Availability Domains and Fault Tolerance

Availability domains are used to determine component and data placement. Nutanix availability domains are:

- Disk Awareness (always).
- Node Awareness (always).
- Block Awareness (optional).Requires a minimum of 3 blocks for FT1 and 5 for FT2, where a block contains either 1,2 or 4 nodes. With Erasure Coding, minimums are 4 blocks for FT1 and 6 blocks for FT2.
- Rack Awareness(optional).Requires a minimum of 3 racks for FT1 and 5 for FT2, and you must define what constitutes a rack and where your blocks are placed. With Erasure Coding, minimums are 4 racks for FT1 and 6 blocks for FT2.

Closely tied to the concepts of awareness and RF is cluster resilience, defined by the redundancy factor or Fault Tolerance (FT). FT is measured for different entities, including:

- Data
- Metadata
- Configuration data
- Free space

FT=1: A cluster can lose any one component and operate normally. FT=2: A cluster can lose any two components and operate normally.

Depending on the defined awareness level and FT value, data and metadata are replicated to appropriate locations within a cluster to maintain availability.

The following table gives an overview of Data awareness and FT levels supported.

© 2021 Nutanix, Inc. All rights reserved

Table 66: Data Awareness and Fault Tolerance Levels Supported

DESIRED AWARENESS TYPE	FT LEVEL	MIN. UNITS*	SIMULTANEOUS FAILURE TOLERANCE
Node	1	3 nodes	1 node
Node	2	5 nodes	2 nodes
Block	1	3 blocks	1 block
Block	2	5 blocks	2 blocks
Rack	1	3 racks	1 rack
Rack	2	4 racks	2 rack

*The minimum units increase by 1 if EC-X is enabled.

FT Guidelines

- FT=2 implies RF=3 for data and RF=5 for metadata by default.
  - Use this setting when you need to withstand two simultaneous failures and have cluster and VM data still be available.
- FT=1 implies RF=2 for data and RF=3 for metadata by default.
  - Use this setting when the application is fine with having only 2 copies of data. This will keep the cluster and VM data available after one failure.
- Another option is to set FT=2 for a cluster with specific containers set to RF=2.
  - With this configuration, data is RF=2, but metadata is RF=5.
  - For containers that are RF=2, if there are 2 simultaneous failures there is a possibility that VM data becomes unavailable, but the cluster remains up.
  - This is different than the FT=1, RF=2 case where two simultaneous failures will result in a cluster being down.

© 2021 Nutanix, Inc. All rights reserved

Table 67: Fault Tolerance for Workloads/Clusters with Higher SLAs

STR-07	USE FT=2 AND RF=3 FOR WORKLOADS AND CLUSTERS THAT NEED HIGHER SLAS OR FOR CLUSTER SIZES >32.
Justification	Using FT=2 and RF=3 provides additional resilience and capability to survive 2 simultaneous failures.
Implication	Additional storage space is required for FT=2 and RF=3, reducing usable capacity.

Capacity Optimization

Nutanix provides different ways to optimize storage capacity that are intelligent and adaptive to workloads characteristics. All optimizations are performed at the container level, so different containers can use different settings.

Erasure Coding (EC-X)

To provide a balance between availability and the amount of storage required, DSF provides the ability to encode data using erasure codes (EC). Like RAID (levels 4, 5, 6, etc.) where parity is calculated, EC encodes a strip of data blocks across different nodes and calculates parity. In the event of a host and/or disk failure, the parity data is used to calculate any missing data blocks (decoding). In the case of DSF, the data block must be on a different node and belong to a different vDisk. EC is a post-process operation and is done on write cold data (Data that hasn't been overwritten in more than 7 days). The number of data and parity blocks in a strip is chosen by the system based on number of nodes and configured failures to tolerate.

The following table summarizes the expected overhead for EC-X vs standard RF2/RF3 overhead:

Table 68: Expected Overhead for EC-X vs Standard RF2/RF3

	FT1 (RF2 Equivalent)		FT2 (RF3 Equivalent) 3X Storage Overhead
# of nodes	EC strip size	3 nodes	1 node
(data/parity)	EC overhead	EC strip size	2 nodes

(data/parity)	EC overhead	3 blocks	1 block	
4	2/1	1.5X	N/A	N/A
5	3/1	1.33X	N/A	N/A
6	4/1	1.25X	2/2	2X
7	4/1	1.25X	3/2	1.6X
8+	4/1	1.25X	4/2	1.5X

EC-X Guidelines

EC-X provides the same level of data protection while increasing usable storage capacity.

- Turn on EC-X for non-mission-critical workloads and workloads that have a significant amount of write cold data, since erasure coding works on write cold data and provides more usable storage.
- For more information refer to specific application guides.

Compression

DSF provides both inline and post-process data compression. Irrespective of inline or post-process compression, write data coming into oplog that is >4k and shows good compression, will be written compressed in OpLog.

- Inline: Compresses sequential streams of data or large size I/Os (>64K) when writing to the Extent Store.
- Post-Process: Compresses data after it is drained from OpLog to the Extent Store after compression delay is met during the next curator scan.

Nutanix uses the LZ4 and LZ4HC compression algorithms. LZ4 compression is used to compress normal data, providing a balance between compression ratio and performance. LZ4HC compression is used to compress cold data to improve the compression ratio. Cold data is characterized as:

- Regular data: No access for 3 days.
- Immutable data: No access for 1 day.

Compression Guidelines

Compression provides on-disk space savings for applications such as databases, and results in a lower number of writes being written to storage. Turn ON inline compression for all containers.

Table 69: Enabling Inline Compression for All Containers

STR-08	ENABLE INLINE COMPRESSION FOR ALL CONTAINERS.
Justification	Since inline compression operates only on sequential data when writing to extent store, enabling it provides immediate space efficiency. Write data coming into oplog is already compressed.
Implication	Slightly higher CPU consumption for compressing/decompressing data.

Deduplication

When enabled, DSF does capacity-tier and performance-tier deduplication. Data is fingerprinted on ingest using a SHA-1 hash that is stored as metadata. When duplicate data is detected based on multiple copies with the same fingerprint, a background process removes the duplicates. When deduplicated data is read, it is placed in a unified cache, and any subsequent requests for data with the same fingerprint are satisfied directly from cache.

Deduplication Guidelines

Deduplication is recommended to be used for full clones, P2V migrations and persistent desktops.

Table 70: Enabling/Disabling Deduplication

STR-09	ENABLE/DISABLE DEDUPLICATION
Justification	Increases effective capacity of caching layer for dedup-able data like VDI clones.
Implication	Slightly higher CPU consumption for running deduplication processes.

Refer to specific application Solution Guides for determining whether to enable deduplication or not.

Virtualization Layer Design

This section describes technical design considerations for design using either Nutanix AHV or VMware ESXi.

Nutanix AHV

Introduction

The Nutanix hypervisor, AHV, is an attractive alternative hypervisor that streamlines operations and lowers overall costs. Built-in to Nutanix Enterprise Cloud and included at no additional cost in the AOS license, AHV delivers virtualization capabilities for the most demanding workloads. It provides an open platform for server virtualization, network virtualization, security, and application mobility. When combined with comprehensive operational insights and virtualization management from Nutanix Prism and Prism Central, AHV provides a complete datacenter solution.

Control Plane

The control plane for managing the Nutanix core infrastructure and AHV are provided by:

- Prism Element (PE)
  - Localized cluster manager responsible for local cluster management and operations. Every Nutanix Cluster includes Prism Element built-in.
  - 1-to-1 cluster manager
- Prism Central
  - Multi-cluster manager responsible for managing multiple Nutanix clusters to provide a single, centralized management interface. Prism Central is an optional software appliance (VM).
  - 1-to-many cluster manager
  - A dedicated PC is required at the source and target sites for any regions or availability zones using DR runbook automation for failing over workloads.

© 2021 Nutanix, Inc. All rights reserved

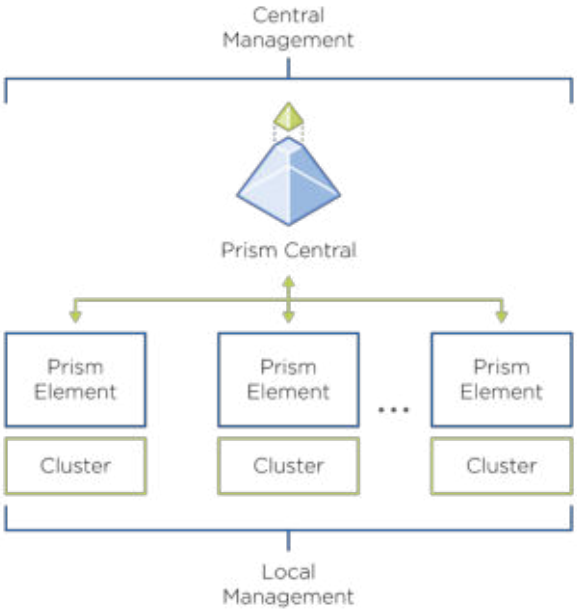


Figure 47: The Nutanix Control Plane

Table 71: Deploying Prism Central for Enhanced Cluster Management

VRT-001	DEPLOY SCALE-OUT PRISM CENTRAL FOR ENHANCED CLUSTER MANAGEMENT.
Justification	Increases the total number of managed objects, increases management plane availability.
Implication	Requires some additional cluster resources to run the necessary VMs.

Table 72: Deploy Prism Central in Each Region or Az Using Runbook Dr Automation.

VRT-002	DEPLOY PRISM CENTRAL IN EACH REGION OR AZ USING RUNBOOK DR AUTOMATION.
Justification	Required for DR runbook automation feature that the source and target sites have separate PC instances controlling them.
Implication	Requires some additional cluster resources to run the necessary VMs.

HA/ADS

Nutanix AHV has built-in VM high availability (VM-HA) and a resource contention avoidance engine called Acropolis Dynamic Scheduling (ADS). ADS is always on and does not require any manual tuning. VM-HA can be enabled and disabled via a simple Prism checkbox. There are two main levels of VM-HA for AHV:

- Best effort. In this HA configuration, no node or memory reservations are required in the cluster. In case of a failure, VMs are moved to other nodes based on the resources/ memory available on each node. This is not a preferred method of HA, since if no resources are available in the cluster after a failure some VMs may not be powered-on. Best effort is the default configuration. Most applicable to non-production environments.
- Guarantee. In this HA configuration, some memory is reserved on each node in the cluster to enable failover of virtual machines from a failed node. The Acropolis service in the cluster calculates the memory to be reserved based on virtual machine memory configuration. All nodes are marked as schedulable with resources available for running VMs. Recommended for production environments. Enable this mode by checking the Enable HA Reservation box in Prism settings.

VM-HA considers memory when calculating available resources throughout the cluster for starting VMs. VM-HA respects Acropolis Distributed Scheduler (ADS) VM-host affinity rules, but it may not honor VM-VM anti-affinity rules, depending on the scenario.

Table 73: Using VM-HA Guarantee

VRT-003	USE VM-HA GUARANTEE
Justification	Ensuring VM high availability is important in production environments. Guarantee should be used in production environments. “Best effort” (default) is optimal in test/ dev environments.
Implication	Reserves enough cluster resources such that an entire host’s worth of VMs can be successfully restarted (RF=2), or two hosts (RF=3). RF=3 must be manually configured, since it is not the default.

For additional information, see: Virtual Machine High Availability.

AHV CPU Generation Compatibility

Similar to VMware's Enhanced vMotion Capability (EVC) which allows VMs to move between different processor generations, AHV will determine the lowest processor generation in the cluster and constrain all QEMU domains to that level. This allows mixing of processor generations within an AHV cluster and ensures the ability to live migrate between hosts.

VMware vSphere

Introduction

VMware vSphere is fully supported by Nutanix with many points of integration. vSphere major and minor releases go through extensive integration and performance testing, ensuring that both products work well together in even the largest enterprises. Rest assured that if you deploy vSphere, the solution is fully supported.

Control Plane

When using vSphere as a part of this document, the control plane is vCenter to manage the vSphere components and Prism Central for managing Nutanix components.

Table 74: Managing ESXi-based Nutanix Clusters

VRT-004	DEPLOY AN HA VCENTER INSTANCE WITH EMBEDDED PSC TO MANAGE ESXI-BASED NUTANIX CLUSTERS.
Justification	Nutanix upgrade automation features, such as 1-click upgrades and LCM, require the advanced control features that vCenter provides. In addition, vCenter Server is required to create and manage the VMware vSphere cluster responsible for DRS and HA.
Implication	Requires additional vCenter licenses and added installation time to fully configure each vCenter instance.

© 2021 Nutanix, Inc. All rights reserved

Table 75: Deploy a vCenter in Region or Az Using Runbook Dr Automation

VRT-005	DEPLOY A vCENTER IN EACH REGION OR AZ USING RUNBOOK DR AUTOMATION.
Justification	Required for DR runbook automation feature that the source and target sites have separate PC and local vCenter instances controlling them.
Implication	Requires some additional cluster resources to run the necessary VMs.

EVC

VMware's Enhanced vMotion Capability (EVC) allows VMs to move between different processor generations within a cluster. EVC mode is a manually configured option, unlike the corresponding feature in AHV. EVC should be enabled at cluster installation time to negate the possible requirement to reboot VMs in the future when it is enabled.

Table 76: Enabling EVC Mode

VRT-006	ENABLE EVC MODE AND SET IT TO THE HIGHEST COMPATIBILITY LEVEL THE PROCESSORS IN THE CLUSTER WILL SUPPORT.
Justification	Even if a cluster is homogenous at the outset, it is likely new nodes will be added over time. Enabling EVC mode when a cluster is built helps ensure future node additions are seamless.
Implication	Requires manual configuration to determine the highest compatible EVC mode.

HA/DRS

VMware HA and DRS are core features which should be utilized as a part of this document. Nutanix best practices dictate that a few HA/DRS configuration settings be changed from the default. These changes are outlined in the design decisions below.

© 2021 Nutanix, Inc. All rights reserved

Table 77: Enabling HA

VRT-007	ENABLE HA
Justification	In production environments VM availability is extremely important. Automatically restarting VMs in case of node failure helps ensure workload availability.
Implication	Requires additional cluster resources to ensure VMs on the failed node(s) can be restarted elsewhere in the cluster.

Table 78: Enabling DRS

VRT-008	ENABLE DRS WITH DEFAULT AUTOMATION LEVEL
Justification	Hot spots can develop in a cluster. DRS moves VMs as needed to ensure optimal performance.
Implication	Initiates vMotion for optimal VM performance in case of resource contention. Moving VMs may temporarily impact Nutanix data locality.

Table 79: Disabling VM PDL and APD Component Protection

VRT-009	DISABLE VM PDL AND APD COMPONENT PROTECTION
Justification	These settings are designed for non-Nutanix storage. They should not be enabled in Nutanix clusters because they can cause storage unavailability. APD monitoring is managed via a Nutanix component.
Implication	Permanent device loss (PDL) and all paths down (APD) monitoring is disabled, and vSphere will not monitor for these conditions.

Table 80: Configuring DAS.IGNOREINSUFFICIENTHBDATASTORE

VRT-010	CONFIGURE DAS.IGNOREINSUFFICIENTHBDATASTORE IF ONE NUTANIX CONTAINER IS PRESENTED TO THE ESXI HOSTS
Justification	In many Nutanix clusters a single datastore is used. The warning message is not applicable in Nutanix environments.
Implication	Eliminates false positive warnings when a cluster uses a single datastore.

© 2021 Nutanix, Inc. All rights reserved

Table 81: Disabling Automation Level for All CVMs

VRT-011	DISABLE AUTOMATION LEVEL FOR ALL CVMS
Justification	CVMs are stored on local storage and cannot be vMotioned to another host. HA does not need to reserve resources for CVMs since they are bound to a single node and cannot be restarted elsewhere in the cluster.
Implication	DRS will not try and move a CVM to another host. Nor will CVM resources be taken into account for HA calculations.

Table 82: Setting Amount of Reserved Resources in the Cluster

VRT-012	SET HOST FAILURES CLUSTERS TOLERATE TO 1 FOR RF2 AND TO 2 FOR RF3.
Justification	Starting in vSphere 6.5, cluster resource percentage is now tied to “host failures cluster tolerates” and automatically adjusts the percentage based on the number of nodes in the cluster.
Implication	Automatically sets the proper amount of reserved resources in the cluster.

Table 83: Setting Host Isolation Response for VMs

VRT-013	SET HOST ISOLATION RESPONSE TO “POWER OFF AND RESTART VMS”
Justification	In an HCI environment all communications, including storage, are via the ethernet network. If a host becomes isolated on the network, VMs need to be moved to a healthy host to continue to function.
Implication	If a host becomes isolated on the network, the VMs on that host will be powered off and restarted on another node in the cluster.

Table 84: Setting Host Isolation Response for the CVM

VRT-014	SET HOST ISOLATION RESPONSE TO “LEAVE POWERED ON” FOR CVMS

© 2021 Nutanix, Inc. All rights reserved

Justification	If there is a transient network disruption, you do not want the CVMs to be powered off.
Implication	If a host becomes isolated on the network, the CVMs will remain on.

Table 85: Disabling HA/DRS for Each Nutanix CVM

VRT-015	DISABLE HA/DRS FOR EACH NUTANIX CVM
Justification	CVM uses local storage and no attempt to restart on other hosts should be performed.
Implication	vSphere will not try and restart the CVM on another host, which is impossible since the CVM uses local storage.

Table 86: Disabling SIOC

VRT-016	DISABLE SIOC
Justification	If SIOC is enabled, it can have negative side effects. These include storage unavailability, creating unnecessary lock files, and complications with Metro Availability.
Implication	Enhanced storage statistics are not available.

Management Layer Design

Control Plane

This document provides a primary control plane where the majority of daily operational tasks are performed. In this design:

- The primary control plane is Prism Central.
- VMware vCenter is also required if you are using VMware ESXi.

Each of these has a maximum size that limits the total number of VMs, nodes, and clusters that can be managed by an instance.

Prism Central

Prism Central (PC) is a global control plane for Nutanix, which includes managing VMs, replication, monitoring, and value-add products such as:

- Nutanix Calm for application-level orchestration. (See the section Calm Application Orchestration for more information.)
- Nutanix Flow for microsegmentation of east-west traffic. (See the section Nutanix Flow for Microsegmentation for more information.)
- Prism Pro for advanced task automation and capacity planning. (See the section Prism Operations for more information.)

These products are installed from Prism Central and managed centrally for all Prism Element clusters associated to the Prism Central instance.

Prism Central plays an important part in a deployment regardless of the scale. There are two deployment architectures for Prism Central that can be utilized and scaled depending on the size and goals of the design.

Single-VM Prism Central Deployment

The single VM deployment option for Prism Central offers a reduced footprint option for designs that do not require high availability for the control plane beyond that provided by the hypervisor. The single Prism Central VM option can be used with a small or large VM sizing. The size directly correlates to the size of environment it can manage.

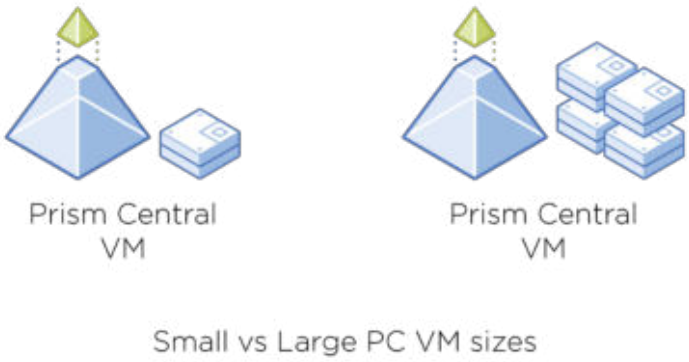


Figure 48: Small vs Large Prism Central VM Sizes

Scale-Out Prism Central Deployment

A scale out PC cluster is comprised of three VMs. All of the members of the cluster are running all of the Nutanix services and products in a distributed architecture.

This distributed architecture allows for the scaled-out Prism Central to manage a larger environment as well as handle more incoming requests from users and the API.

Additionally, it provides a higher level of availability. If one of the VMs is down or performing a maintenance operation, the other members remain available. You have the option of deploying the PC VMs as small- or large-sized VMs.

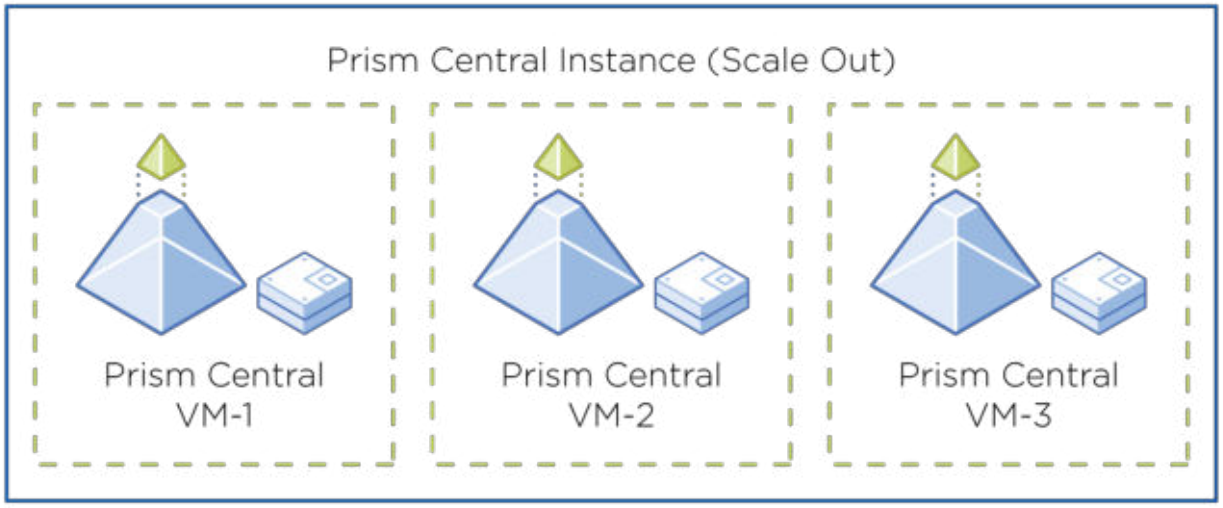


Figure 49: Scale Out Prism Central Deployment

The scale-out Prism Central architecture deploys 3 VMs on the same cluster and is superior in terms of availability and capacity. If scale-out is not the best option from the beginning, it is fairly simple to go from a single PC VM deployment to a scale-out architecture at a later time.

Image Templates

Within any environment operating at scale there is a need to keep approved template images available and in sync between clusters. For AHV clusters, image placement policies should be utilized for this. Image policies are configured to determine the clusters each image can or cannot be utilized in. This makes the initial roll out of new and updated image versions easy.

For ROBO deployments utilizing AHV, image placement policies should be configured to avoid unnecessary data being replicated between locations with limited bandwidth.

Prism Central Recommendations

Table 87: Prism Central Recommendations

Recommendation Area	Description
Prism Central instances	Deploy Prism Central in the following way: • Scale out mode (cluster of 3 VMs). The benefits: • Simplifies capacity planning. • Simplifies platform life cycle management. • Simplifies management for virtual networking. • Reduce management overhead.
Workload Domains	Deploy management workloads such as Prism Central on management clusters.
Access Control	• Integrate with Active Directory. • Leverage AD groups to optimize access management.

VMware vCenter Server

For designs utilizing ESXi as the hypervisor, vCenter is a critical part of the infrastructure for hypervisor cluster management and VM operations. It may be accessed often in environments that are not highly automated. In automated environments, vCenter is needed to perform operations that are sent to it from different orchestration layers.

The vCenter appliance can be deployed in many different sizes. This ultimately determines how large an environment you can support in terms of VMs, hosts,

and clusters. Refer to the official VMware documentation for the version you are deploying to determine the proper sizing for your design.

Beyond the size of the environment, the vCenter appliance can be deployed as a single VM or as a vCenter High Availability (vCenter HA) for additional availability. The option to deploy vCenter HA does not increase the size of the environment it can manage as a single virtual appliance is active at any time. The size of the environment that can be managed is based on the size of the VMs deployed.

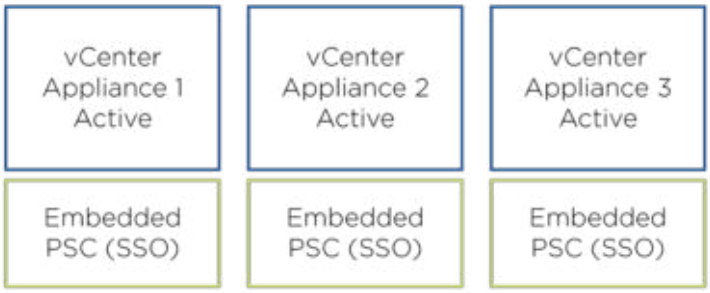


Figure 50: vCenter Appliances

To ensure that the control plane is as highly available as possible, the clustered deployment option is preferred as long as its compatible with other layers of the solution.

vCenter Server Recommendations

Table 88: vCenter Server Recommendations

RECOMMENDATION AREA	DESCRIPTION
vCenter instances	Deploy one vCenter in an HA configuration for all workloads.Configure with Embedded PSC.
Workload domains	Deploy management workloads on management clusters.
Access Control	• Integrate with Active Directory. • Leverage AD groups to optimize access management.

Dependent Infrastructure

There are a variety of other infrastructure services that are necessary for a successful Nutanix deployment, such as NTP, DNS, and AD. You may already have these infrastructure services deployed and available for use when you deploy your Nutanix environment. If they do not exist, then you will probably need to deploy them as part of the new environment

NTP

Network Time Protocol (NTP) is used to synchronize computer clock times including network, storage, compute, and software. If clock times drift too far apart, some products may have trouble communicating across layers of the solution. Keeping time in-sync is beneficial when examining logs from different layers of the solution to determine the root cause of an event.

A minimum of three NTP servers should be configured and accessible to all solution layers with the NTP standard recommendation being five to detect a rogue time source. These include AOS, AHV, and Prism Central, plus vCenter and ESXi if using VMware vSphere as the virtualization layer. Use the same NTP servers for all infrastructure components.

Do not use an Active Directory Domain Controller as an NTP source.

If you are in a Dark site with no internet connectivity, consider using a switch or GPS time source.

Table 89: NTP Servers for Infrastructure Components

DEP-001	A MINIMUM OF THREE NTP SERVERS FOR ALL INFRASTRUCTURE COMPONENTS SHOULD BE PROVIDED.
Justification	Unreliable time services can lead to authentication failures, mis-aligned log entries, and other critical errors.
Implication	Provides reliable time services for all system components.

DNS

Domain Name System (DNS) is a directory service that translates domain names of the form domainname.ext to IP addresses. DNS is important to ensure

that all layers can resolve names and communicate. A highly available DNS environment should be utilized to support this design. At least two DNS servers should be configured and accessible at all layers to ensure components can reliably resolve addresses at all times.

These layers include the following:

- AOS
- AHV
- Prism Central
- VMware ESXi
- VMware vCenter
- Network switches

Table 90: DNS Servers for Infrastructure Layers

DEP-002	A MINIMUM OF TWO DNS SERVERS SHOULD BE CONFIGURED ACCESSIBLE TO ALL INFRASTRUCTURE LAYERS
Justification	DNS is critical for a number of services and infrastructure failures will occur if DNS is unavailable.
Implication	The organization needs to deploy at least two DNS servers and point all services to them.

Active Directory

Active Directory (AD) is a directory service developed by Microsoft for Windows domain networks. AD often serves as the authoritative directory for all applications and infrastructure within an organization. For this design, all of the consoles and element managers will utilize RBAC and use AD as the directory service for user and group accounts. Where possible, AD groups should be utilized to assign privileges for easier operations. User access can then be controlled by adding or removing a user from the appropriate group.

The AD design should be highly available to ensure that directory servers are always available when authentication requests occur.

Logging Infrastructure

Capturing logs at all layers of the infrastructure is very important. For example, if there is a security incident, logs can be critical for forensics. An example of a robust log collector is Splunk, but there are other options. Nutanix recommends that you deploy a robust logging platform that meets your security requirements. All infrastructure logs should be forwarded to the centralized log repository.

Recommendation

It is commonly recommended to store the logs in a different cluster, or location, from where they are being collected. This protects the logs in case of a catastrophic cluster failure, ensuring they can later be used for forensics.

Security Layer Design

Nutanix Enterprise Cloud can be used to build private or multi-tenant solutions, and depending on the use case the security responsibility will vary. The security approach includes multiple components:

- Physical
  - Datacenter access
  - Equipment access (firewalls, load balancers, nodes, network switches, racks, routers)
- Virtual Infrastructure
  - Clusters
  - Nodes
  - Management components
  - Network switches
  - Firewalls
  - Load balancers

© 2021 Nutanix, Inc. All rights reserved

- Threat vectors
  - Active
  - Automated
  - Internal
  - External
- Workloads
  - Applications
  - Containers
  - Virtual Machines

Nutanix Enterprise Cloud infrastructure is designed to deliver a high level of security with less effort. Nutanix publishes custom security baseline documents for compliance, based on United States Department of Defense (DoD) RHEL 7 Security Technical Implementation Guides (STIGs) that cover the entire infrastructure stack and prescribe steps to secure deployment in the field. The STIGs uses machine-readable code to automate compliance against rigorous common standards.

Nutanix has implemented Security Configuration Management Automation (SCMA) which is installed, configured, and enabled during the Nutanix deployment. There are configuration options, however one key differentiator is that Nutanix provides this functionality by default. SCMA checks multiple security entities for both Nutanix storage and AHV. Nutanix automatically reports inconsistencies and reverts them to the baseline.

With SCMA, you can schedule the STIG to run hourly, daily, weekly, or monthly. STIG has the lowest system priority within the virtual storage controller, ensuring that security checks do not interfere with platform performance.

In addition, Nutanix releases Security Advisories which describe potential security threats and their associated risks plus any identified mitigation(s) or patches.

For more information about security, see Building Secure Platforms And Services With Nutanix Enterprise Cloud.

Authentication

Maintain as few user and group management systems as possible. A centrally managed authentication point is preferred to many separately managed systems. Based on that general recommendation you should at a minimum take advantage of the external LDAP support provided by Nutanix components.

Prism Central also provides both LDAP and Security Assertion Markup Language (SAML) and makes it possible for users to authenticate through a qualified Identify Provider (IdP). If none of the options are available Nutanix also provides local user management capabilities.

Table 91: Active Directory Authentication

SEC-001	USE ACTIVE DIRECTORY AUTHENTICATION. THIS APPLIES FOR USER AND SERVICE ACCOUNTS.
Justification	User activity logged for auditing purposes and account security is configured and maintained from a single centralized solution.
Implication	Requires a highly available active directory infrastructure and additional initial configuration.

Table 92: Using SSL/TLS Connection to Active Directory

SEC-002	USE SSL/TLS CONNECTION TO ACTIVE DIRECTORY.
Justification	Increases security by eliminating cleartext exchanges on the network.
Implication	Might require additional configuration.

Certificates

All components facing consumers should be protected with certificate authority (CA) signed certificates. Internal or external signed certificates depend on consumer classifi- cation and what kind of service the specific component provides.

© 2021 Nutanix, Inc. All rights reserved

Table 93: Using a Certificate Authority

SEC-003	USE CERTIFICATE AUTHORITY (CA) THAT ARE CONSIDERED "TRUSTED CA" BY YOUR ORGANIZATION FOR THE COMPONENTS WHERE CERTIFICATES CAN BE REPLACED. THIS CAN BE EITHER INTERNAL OR EXTERNAL SIGNED CERTIFICATES.
Justification	Certificates are required as part of Transport Layer Security (TLS) protocol and are negotiated during session establishment to establish secure session. They provide an extra layer of security and prevent man-in-the-middle attacks.
Implication	Implementation effort and potentially an extra cost if certificates need to be purchased.

Cluster Lockdown

Nutanix cluster lockdown lets you enforce SSH access to the CVMs and hosts using key pairs instead of passwords.

Table 94: Nutanix Cluster Lockdown

SEC-004	DO NOT USE NUTANIX CLUSTER LOCKDOWN.
Justification	Nutanix recommends that access, including SSH, directly to CVM and hypervisor should be restricted to as few entities as possible:Operators. Keep service account passwords secure.Remote workstations. Use e.g. jump hosts.Networks. Use network segmentation via firewall rules or IPtables rules.
Implication	Users can access the CVM with username and password.

If only password-less communication is required, then enable Nutanix cluster lockdown.

VMware Cluster Lockdown

Cluster lockdown can be enabled at the VMware layer as well. Enabling either normal or strict cluster lockdown mode can fail the Nutanix cluster functionality. Make the necessary ESXi configurations to guarantee Nutanix functionality if vSphere cluster lockdown mode is required.

Table 95: vSphere Cluster Lockdown

SEC-005	DO NOT USE VSPHERE CLUSTER LOCKDOWN.
Justification	No requirements exist in this design that justify vSphere cluster lockdown mode.
Implication	Users can access vSphere cluster with username and password.

Hardening

There are certain hardening configurations you can apply for the AHV host and the CVM if required:

- AHV and CVM
  - AIDE. Advanced Intrusion Detection Environment.
  - Core. Enable stack traces for cluster issues.
  - High Strength Passwords. Enforce complex password policy (Min length 15, different by at least 8 characters).
  - Enable Banner. Get specific login message via SSH.
- CVM only
  - Enforce SNMPv3 only.

Table 96: CVM and Hypervisor AIDE

SEC-006	ENABLE CVM AND HYPERVISOR AIDE
Justification	With AIDE enabled, the environment will perform checksum verification of all static binaries and libraries for improved security.
Implication	Additional, very limited, CVM and hypervisor resources will be required.

Table 97: Configuring SCMA

SEC-007	CONFIGURE SCMA TO RUN HOURLY
Justification	To more quickly capture unacceptable configuration drift. Default is to take actions daily.

© 2021 Nutanix, Inc. All rights reserved

Implication	SCMA is lightweight and adds very limited additional load to CVM and hypervisor. Security benefits outweigh the added resources.

Table 98: Unused ESXi Services and Unused Firewall Ports

SEC-008	STOP UNUSED ESXI SERVICES AND CLOSE UNUSED FIREWALL PORTS. Important: Make sure not to stop a service or close a firewall port required by Nutanix, such as SSH and NFS
Justification	To limit the attack surface.
Implication	Additional configuration required.

We recommend that you follow the Hardening Controller VM guide. If you are using AHV, follow the Hardening AHV guide.

For updated Hardening Guides, search portal.nutanix.com for the most recent guide that relates to your version of AOS and AHV.

Internet-facing Services

Security for Internet-facing services is out of the scope of this design, but highly important and mentioned here for awareness.

Internet-facing services are at constant risk of being attacked. Two common types of attacks are denial of service (DoS) and distributed denial of service (DDoS). To help mitigate the potential impact of a DoS or a DDoS attack a design can take advantage of:

- Multiple internet connections (active/backup).
- ISP provided DoS and DDoS filtering.

There are additional ways to implement protection against these types of attacks.

Logging

Logging is critical from an auditing and traceability perspective so making sure the virtual infrastructure, AOS, Prism Central, AHV, ESXi, any additional

software in the environment send their log files to a highly available log infrastructure is critical.

Recommendation

A minimum one of the individual targets making up the highly available logging infrastructure should run outside of the virtual infrastructure itself so that if the virtual infrastructure is compromised, forensic investigations will need to access logs that might not be available on the cluster itself.

A single centralized activity logging solution for auditing purposes and account security should be configured and maintained.

Table 99: Logging

SEC-009	SEND LOG FILES TO A HIGHLY AVAILABLE SYSLOG INFRASTRUCTURE.
Justification	Logs are helpful in report creation, during troubleshooting, and for investigating potential security breaches.
Implication	Requires initial configuration and a highly available logging infrastructure.

Table 100: Including All Nutanix Modules in Logging

SEC-010	INCLUDE ALL NUTANIX MODULES IN LOGGING.
Justification	Ensures data from all modules are included and search- able via a logging system. Refer to Nutanix Syslog documentation for additional information. Modules can be excluded if needed.
Implication	Data will be generated per module meaning the more modules included, the more log data is generated.

Table 101: Error Log Level for Nutanix Components

SEC-011	USE ERROR LOG LEVEL FOR THE NUTANIX COMPONENTS.

© 2021 Nutanix, Inc. All rights reserved

Justification	Ensures data from all modules are included and searchable via a logging system. Refer to Nutanix Syslog documentation for additional information. Modules can be excluded if needed.
Implication	Will not send all logging information to syslog infrastructure. The following levels will be included:ErrorCriticalAlertEmergency

Table 102: ESXi Logging Level, Rotation, and File Size

SEC-012	USE DEFAULT ESXI LOGGING LEVEL, LOG ROTATION, AND LOG FILE SIZES.
Justification	No reason to change size and rotation unless a huge amount of logs are expected.
Implication	In rare situations, logs might rotate very fast.

Table 103: Protocols for Log Transport

SEC-013	IF EXTRA SECURITY AND RELIABILITY ARE REQUIRED, THEN USE TCP FOR LOG TRANSPORT. OTHERWISE, USE THE DEFAULT SYSLOG PROTOCOL, UDP.
Justification	Provides a reliable logging setup and ensures the logs are being received by the logging infrastructure.
Implication	Requires a sender and receiver to establish communication which generates minimal additional network traffic. May require configuration of logging infrastructure to accept TCP communication.

Table 104: Using Port 514 for Logging

SEC-014	USE PORT 514 FOR LOGGING.
Justification	Defined port in syslog RFC. No reason has been identified to change the port.

© 2021 Nutanix, Inc. All rights reserved

Implication	Traffic on well-known ports can be easy to locate.

Network Segmentation

To protect Nutanix CVM and hypervisor traffic, place them together in their own dedicated VLAN, separate from other VM traffic. This applies to all hosts in a cluster.

Nutanix recommends configuring the CVM and hypervisor host VLAN as a native, or untagged, VLAN on the connected switch ports. This native VLAN configuration allows for easy node addition and cluster expansion. By default, new Nutanix nodes send and receive untagged traffic. If you use a tagged VLAN for the CVM and hypervisor hosts instead, you must configure that VLAN while provisioning the new node, before adding that node to the Nutanix cluster.

Do not segment Nutanix storage and replication traffic, or iSCSI Volumes traffic, on separate interfaces (VLAN or physical) unless additional segmentation is required by mandatory security policy or the use of separate physical networks. The added complexity of configuring and maintaining separate networks with additional interfaces cannot be justified unless absolutely required.

Table 105: VLAN for Traffic Separation

SEC-015	USE VLAN FOR TRAFFIC SEPARATION OF MANAGEMENT AND USER WORKLOADS.
Justification	VLAN is a well-known standard for traffic separation when there are no requirements specifying physical separation.
Implication	This provides clear separation between: Management and end user traffic, Different management traffic types, and Different end user traffic types.

Table 106: Place CVM and Hypervisor on Same VLAN and Subnet

SEC-016	PLACE CVM AND HYPERVISOR ON THE SAME VLAN AND SUBNET.

Justification	Required for Nutanix functionality. Both hypervisor and CVM should be classified as core or management service traffic delivering the same service and can therefore be placed on the same VLAN.
Implication	Different components placed on the same network segment. Alternate configurations are not supported.

Table 107: Place Out-of-band Management on a Separate VLAN or Physical Network

SEC-017	PLACE OUT-OF-BAND MANAGEMENT ON A SEPARATE VLAN OR PHYSICAL NETWORK.
Justification	Out-of-band management is not required for core Nutanix functionality. To provide additional security, the management interface should not be on the same network segment as the CVM and hypervisor.
Implication	Network or VLAN management.

Role-based Access Control (RBAC)

Nutanix has built-in RBAC in both Prism Central and Prism Element, with the option to create custom roles in Prism Central. RBAC is used to limit access and control for various individuals and groups of administrative users. Use a least-privilege and separation-of- duties approach when assigning permissions to make sure each group or individual user has just enough permissions to perform their duties. Use pre-defined roles or create new roles as needed.

Configure RBAC at the Prism Central level since it provides the overlying management construct. Via Prism Central, consumers and administrators will be directed to underlying components, such as Prism Element, when needed. This ensures your least-privilege configuration stays in place, avoiding common mistakes that occur when RBAC is configured at multiple different levels.

The following table shows Prism Central default roles:

Table 108: Default Prism Central Roles

Role	Purpose

Super Admin	Highest-level admin with full infrastructure and tenant access. Manages a Nutanix deployment and can set up, configure, and make use of every feature in the platform.
Self-Service Admin	Cloud admin for a Nutanix tenant. Manages virtual infrastructure, oversees self-service, and can delegate end-user management.
Project Admin	Team lead to whom cloud administration is delegated in the context of a project. Manages end-users within a project and has full access to their entities.
Prism Viewer	View-only admin. Has access to all infrastructure and platform features but cannot make any changes.
Prism Admin	Day-to-day admin of a Nutanix deployment. Manages the infrastructure and platform but cannot entitle other users to be admins.
Operator	Lifecycle manager for team applications. Works on existing application deployments, exercises blueprint actions.
Developer	Application developer within a team. Authors blueprints, tests deployments, and publishes applications for other project members.
Consumer	Owner of team applications at runtime. Launches blueprints and controls their lifecycle and actions.

When creating custom roles in Prism Central there are multiple entities available, each with their own set of permission definitions:

- App
- VM
- Blueprint
- Marketplace Item
- Report
- Cluster
- Subnet
- Image

The following table describes Prism Element default roles:

Table 109: Prism Element Default Roles

Role	Purpose
User Admin	Able to view information, perform any administrative task, and create or modify user accounts.
Cluster Admin	Able to view information and perform any administrative task. Cannot create or modify user accounts.
Viewer	Able to view information only. No permission to perform any administrative tasks. Useful for auditing purposes.

In addition, there are pre-defined roles in VMware vCenter Server, with the option to create custom roles.

Table 110: Use Least-Privilege Approach when Determining Access

SEC-018	USE A LEAST-PRIVILEGE ACCESS APPROACH WHEN DECIDING WHO HAS ACCESS. ALIGN RBAC STRUCTURE AND USAGE OF DEFAULT PLUS CUSTOM ROLES ACCORDING TO COMPANY REQUIREMENTS.
Justification	This approach helps protect the environment from having users performing actions they are not authorized to perform.
Implication	Access requirements need to be defined, created, and implemented for each access group.

Table 111: Align RBAC to Company Structure

SEC-019	ALIGN RBAC STRUCTURE AND USAGE OF DEFAULT PLUS CUSTOM ROLES ACCORDING TO THE COMPANY REQUIREMENTS DEFINED VIA SEC-018.
Justification	There are numerous ways of applying RBAC. This design document can't cover all possibilities.
Implication	If no company structure exists, fall back on description in SEC-018 and create a structure that meets your needs.

© 2021 Nutanix, Inc. All rights reserved

Data-at-Rest Encryption

Data-at-Rest Encryption (DaRE) is a security measure to prevent data from being stolen. DaRE offers protection against data exposure in the event of:

- Activities that require a full data copy that will be used outside the platform.
- Failed drives leaving the datacenter.
- Drive or node theft.

Keeping management traffic, including storage traffic, on a separate network is often an adequate data security practice. Nutanix supports four different options for Data at Rest Encryption (DaRE):

- Self-Encrypting Drives (SEDs) with an External Key Manager (EKM).
- Software-based encryption with an EKM.
- Software-based encryption and SED with an EKM (aka Dual Encryption).
- Software-based encryption with the Nutanix Native Key Manager (KMS).

With all software-based encryption options, encryption is performed in the software layer and data is encrypted over the wire, between CVMs. All or part of the user VM data can be encrypted.

- AHV encrypts the entire Nutanix cluster.
- ESXi gives you the option to define encryption on a per-Nutanix-container basis if required.

Self-Encrypting Drives (SEDs) provide FIPS 140-2 Level 2 compliance and can be used without any performance impact. Nutanix Software Based Encryption and Native Key Manager are FIPS 140-2 Level 1 Evaluated.

You can use both software-based encryption and SEDs. This requires an external key management server.

> Note: All methods of DaRE are FIPS 140-2 compliant however if levels 2, 3, or 4 are required a hardware component is necessary. Software-based encryption with the Nutanix Native Key Manager (KMS) can encrypt storage containers (ESXi or Hyper-V) or the entire Cluster (AHV).

© 2021 Nutanix, Inc. All rights reserved

Table 112: Do Not Use Storage Encryption

SEC-020	DO NOT USE STORAGE ENCRYPTION.
Justification	There are no technical requirements to justify implementing storage encryption in this design.
Implication	Data is never encrypted unless encrypted at the virtual machine level or via the ESXi VM encryption feature.

Key Management Server

A key management server is required when using Nutanix storage encryption.

When using an external key management service, you should have a minimum of two key management servers running. At least one server must not run on the infrastructure it is protecting.

Table 113: Do Not Use a Key Management Server

SEC-021	DO NOT USE A KEY MANAGEMENT SERVER.
Justification	No functionality will be used that requires a key management server.
Implication	N/A

Nutanix Flow

During the environment lifecycle, it's important to ensure compliance is met and that your security configuration meets required standards. Nutanix Flow provides a centralized view of the security posture of your Nutanix environment and provides visibility into the security of your on-prem environment based on known compliance standards.

Flow can help you to comply with regulatory and business-specific compliance policies such as PCI-DSS, HIPAA, and NIST. (note: Xi Beam is not available for dark sites)

You gain deep insights into your on-premises Nutanix deployments based on over 300 audit checks and security best practices according to:

- Audit security checks for access, data, networking and virtual machines.

- Compliance checks against PCI-DSS v3.2.1 for AHV, AOS, Flow and Prism Central.
- For more information on Nutanix Flow, see the section Nutanix Flow for Microsegmentation.

Automation Layer Design

Nutanix supports intelligent IT operations and advanced automation that enable you to streamline operations, enable IT as a Service, and support the needs of developers and business teams. This section covers automation and orchestration of virtual infrastructure, focusing on provisioning and maintenance which are important aspects of the overall solution. Nutanix tools reduce the time required to perform initial setup plus software and firmware upgrades.

Upgrades

Upgrades can occur across a variety of components. The process to upgrade many of these components is primarily executed using LCM, which is able to understand dependencies without operator intervention.

It's important to understand the impact of upgrading certain software components, as their upgrade may affect the outside world.

- AOSWhen upgrading, each CVM is individually upgraded in a rolling fashion. While a CVM is rebooting, the host it is running on is redirected to a remote CVM to deliver storage IO. This is invisible to the VM. However, it does result in a loss of locality, which can potentially impact IO throughput and latency if the load on the system is high.
- HypervisorWhen upgrading the hypervisor on a node, each VM must be migrated off the host for the update to be performed so that a reboot of the host can occur. For vSphere, this requires vCenter integration to be configured to allow for a host to be put into maintenance mode.AHV live migration and vSphere vMotion are normally non-disruptive, however certain applications that utilize polling-style device drivers or that have near-real-time kernel operations cannot tolerate being migrated during the final

cutover step. Hypervisor updates will require downtime for these apps if they don't offer native failover functionality.

Firmware

Firmware can be updated for a variety of devices including:

- Board Management Controller (BMC)
- Motherboard (BIOS)
- Host Bus Adapter (HBA)
- Host Boot Device (SATA DOM)
- SSD or HDD
- NIC

Software

Software updates can be performed for multiple components, including:

- Acropolis Operating System (Core AOS in CVM)
- Hypervisor (vSphere or AHV)
- Nutanix Cluster Check (NCC)
- Life Cycle Manager (LCM)
- Acropolis Operating System (AOS in Prism Central)
- MicroServices Platform (MSP in Prism Central)
- Services Functionality
  - Nutanix Files (Prism Element)
  - Calm/Epsilon (Prism Central)
  - Objects Manager (Prism Central)
  - Karbon (Prism Central)
  - Era (independent)
  - Move (independent)
  - X-Ray (independent)

© 2021 Nutanix, Inc. All rights reserved

Life Cycle Manager (LCM)

Nutanix Lifecycle Manager (LCM) improves the efficiency and reliability of IT infrastructure upgrades in modern datacenters. Nutanix LCM determines any software and firmware dependencies, intelligently prioritizes updates, and automates the entire upgrade process across all clustered hosts, without any impact to applications or data availability. It supports one-click upgrades across multiple qualified hardware manufacturers and configurations, so IT teams have the flexibility to deploy the best hardware for each use case – and still benefit from centralized upgrade capabilities. LCM supports AHV, ESXi and Hyper-V hypervisors, plus hardware support for Nutanix, Hewlett Packard Enterprise, Dell/EMC, Lenovo, Fujitsu, Inspur and Intel. LCM supports environments with internet connectivity as well as dark site deployments.

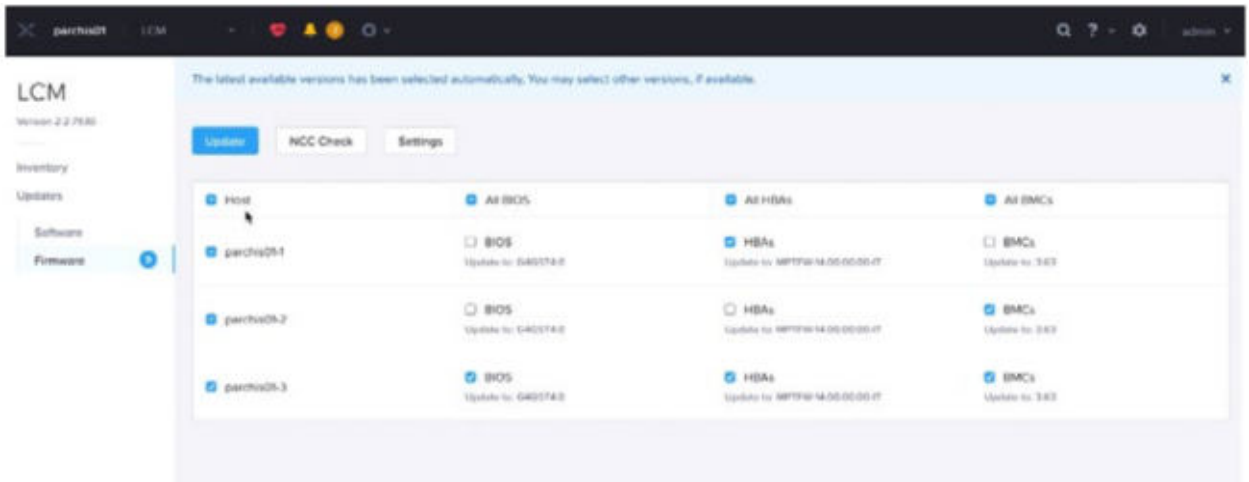


Figure 51: Life Cycle Manager

Foundation

Foundation manages the initial setup and configuration of a cluster. Nutanix nodes may come pre-installed with AHV and the Controller Virtual Machine (CVM) and you can:

- Add the Nodes to an existing Nutanix cluster.
- Create a new Nutanix cluster.
- Re-image the nodes with different AHV and or AOS version or different hypervisor and create a Nutanix cluster.

There are five ways of invoking a Foundation process:

- Connect to a factory deployed Nutanix node's CVM via http://CVM_IP:8000/
- Use portable Foundation, Mac or Windows executable.
- Standalone Foundation VM.
- When adding a node to an existing Nutanix cluster.
- Call Foundation API via third party solution which provides a way to orchestrate the deployment.

Note: The Foundation process may vary for different hardware vendors.

Nutanix provides a Foundation pre-configuration service which is accessible via install. nutanix.com. Via the service, you can define and download the Nutanix cluster configuration to be used during the Foundation process. The downloaded file contains all configuration required to perform the foundation operation and comes in json format. This makes it easy to keep track of configurations, document them, and share with your peers.

© 2021 Nutanix, Inc. All rights reserved

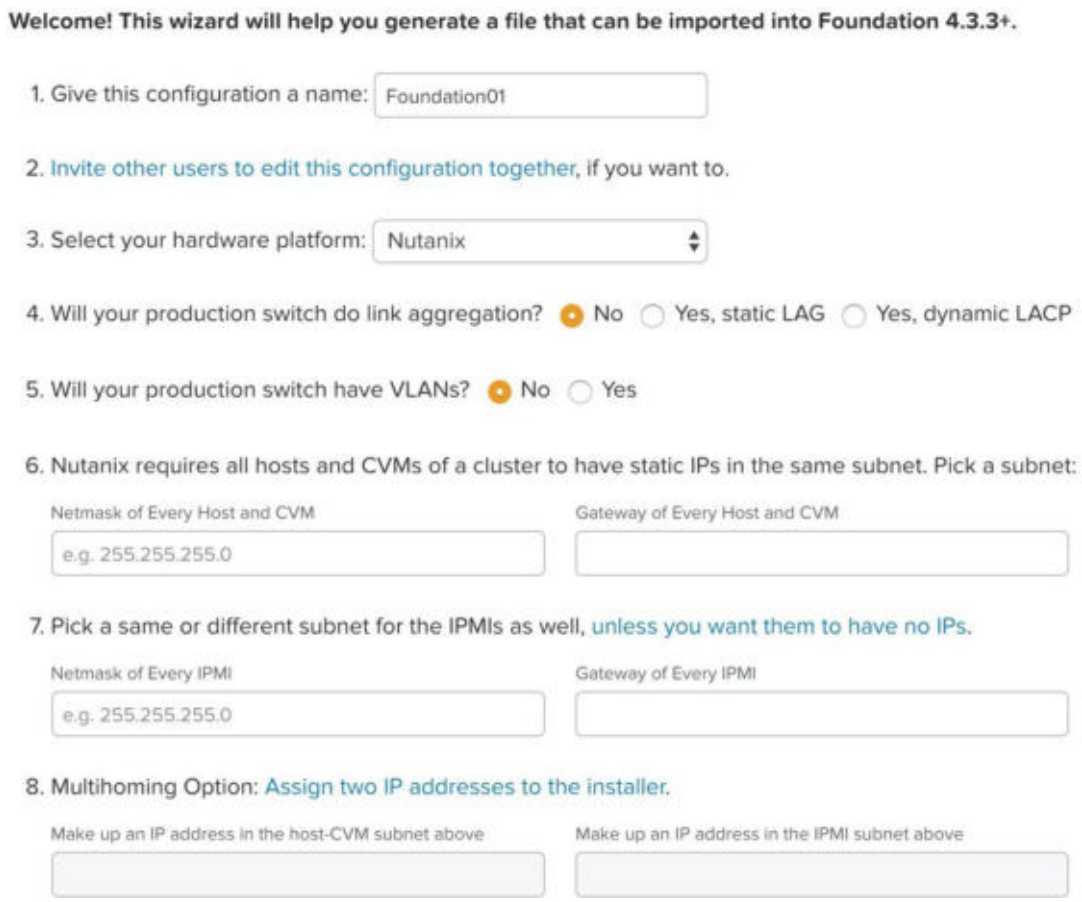


Figure 52: Foundation Pre-Configuration Service

The following table provides an overview of the functions available for the different Foundation software options.

Table 114: Overview of Different Foundation Software Options

	CVM Foundation	Portable Foundation (Windows, Mac)	Standalone Foundation VM
Function	Factory-imaged nodes.	Factory-imaged nodes.	Maintains high availability in the event of the loss of one switch.

Bare-metal nodes	Factory-imaged nodes.		Requires at least two TOR switches, which will impact cost and rack space. Increases total bandwidth to each host to 20 Gbps.
Bare-metal nodes			
Hardware	Any	Nutanix (NX)	
Dell (XC)			
HPE (DX)	Any		
If IPV6 is disabled	Cannot image nodes.	IPMI IPv4	
required on the nodes.	IPMI IPv4 required on the nodes.		
VLAN Support	No	No	Yes
LACP Support	Yes	Yes	Yes
Multi-homing Support	N/A	Yes	Yes

Note: When available, Foundation Central will provide capabilities to perform Foundation operations via Prism Central.

Operations Design

Capacity and Resource Planning

After initial installation and migration of workloads to the platform, long term capacity planning should be enabled to avoid running out of resources. Many of the features discussed are integrated into Prism Pro.

Table 115: Deploying Prism Pro for Enhanced Cluster Management

OPS-001	DEPLOY PRISM PRO FOR ENHANCED CLUSTER MANAGEMENT

Justification	Analytics, Capacity planning, Custom Dashboards, and Playbooks require the presence of Prism Pro.
Implication	Requires an additional license, which can increase cost.

Cluster capacity

Native runway calculations built into Prism Central will automatically calculate the remaining capacity of the system as soon as the cluster Prism Element is brought under management by Prism Central.

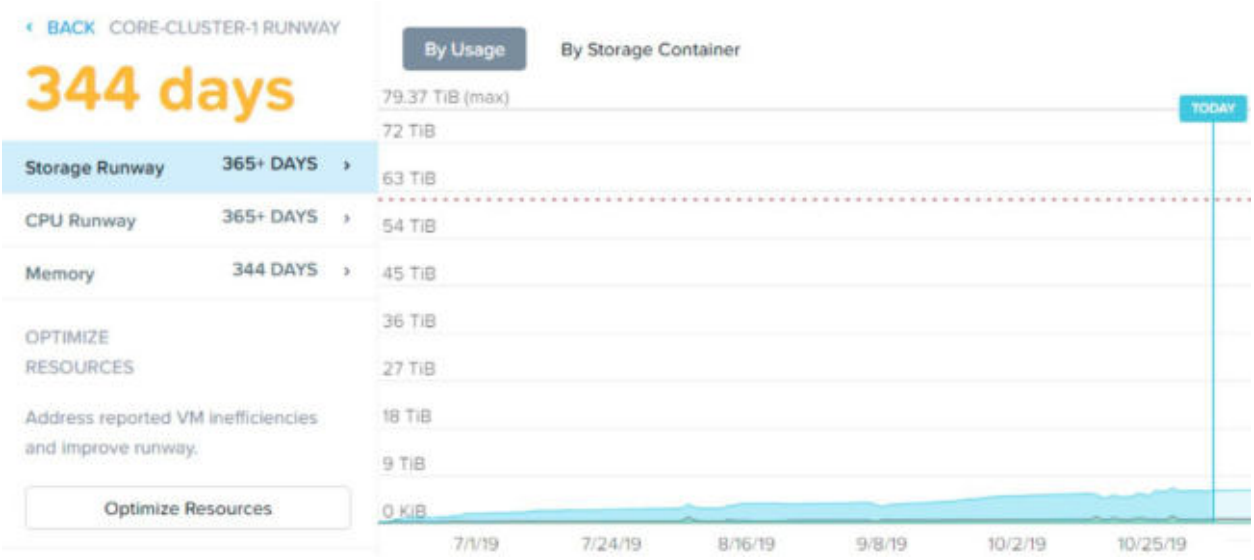


Figure 53: Capacity Runway

These runway calculations should be configured to run as part of periodic reports and reviewed on a regular basis to ensure sufficient capacity exists. This is especially important for organizations that have a significant lag between the date a commitment to purchase additional gear occurs and when its online and available to use.

Expansion planning

When a new workload is identified to be onboarded, planning needs to occur if the new workload size or requirements are outside of established patterns. Prism Central offers a scenario simulation function that shows how the available

capacity runway would change if this new workload was accommodated. Utilizing this planning functionality helps avoid unplanned capacity constraints.

Figure 54: Expansion Planning

Table 116: Reviewing Monthly Capacity Planning

OPS-002	REVIEW MONTHLY CAPACITY PLANNING
Justification	Analytics, capacity planning, custom dashboards, and playbooks require Prism Pro.
Implication	Requires an additional license, which can impact cost.

Right-Sizing VMs

While general system capacity planning is useful, accurate and efficient system level planning requires accurate sizing for individual workloads. Machine learning in Prism Central provides anomaly detection for VMs when the workload crosses learned thresholds.

In addition, based on a number of thresholds, the system will categorize VMs based on their behavior. These categories include:

- Bully: A VM with potential to degrade the overall cluster performance by impacting the capability of the node they reside on.
- Constrained: A VM that is experiencing very high CPU or RAM utilization and potentially is unable to meet the needs of the application its running and is likely a candidate to increase the resources provided.

© 2021 Nutanix, Inc. All rights reserved

- Over-provisioned: A VM that is significantly underutilizing the resources provided to it and is likely a candidate to reduce in size.
- Inactive: A VM that has not powered on recently (dead) or is virtually idle (zombie) that is a candidate for reclamation.

In addition, custom alert policies can be created that match VMs by conditions.

Upgrade Methodology

This section describes the design decisions associated with upgrading the Nutanix infrastructure. Whether to perform upgrades during business hours or outside business hours comes down to a number of possible factors, including:

- Is the environment sized to tolerate failures during upgrades?
- Will performance be adequate during upgrades?
- Past experiences.
- Size of Nutanix cluster.

Table 117: Performing Updates

OPS-003	PERFORM UPDATES AFTER HOURS FOR PERFORMANCE OR MIGRATION SENSITIVE APPLICATIONS
Justification	Reduces load on the system to be handled during rolling CVM reboot from code upgrade.
Implication	Potential staffing impact.

When upgrading, AOS offers the choice of release trains to apply. Each of these is denoted with a major.minor.maintenance.patch numbering scheme, for example: 5.10.7.1 or 5.11.1.1.

A release train is based on the major and minor components. There are two types of release trains:

- Short Term Support (STS) releases which include new features and provide a regular and frequent upgrade path. These releases are maintained for shorter durations.

- Long Term Support (LTS) releases which provide bug fixes for features that have been generally available for a longer period of time. After features have been generally available in an STS for some time, they are included in and LTS, which are maintained for a longer duration.

KB 5505 on the Nutanix Support Portal covers the differences in greater detail.

Table 118: Utilizing the Current LTS Branch

OPS-004	UTILIZE THE CURRENT LTS BRANCH
Justification	Unless a new feature is required for a design, enterprise customers typically prefer a less frequent major version change of software components.
Implication	Lack of access to features present in or requiring new branch.

Updates to an existing train are released on a regular basis and should be applied on a standard cadence.

Table 119: Upgrading to Maintenance and Current Patch Versions

OPS-005	UPDATE TO THE NEXT MAINTENANCE VERSION 4 WEEKS AFTER RELEASE. UPDATE TO THE CURRENT PATCH VERSION 2 WEEKS AFTER RELEASE.
Justification	Unless a new feature is required for a design, enterprise customers typically prefer a less frequent major version change of software components.
Implication	Frequent smaller updates keep the amount of changes per update relatively small, minimizing the amount to troubleshoot in the event something goes wrong.

Testing

Environments which seek to achieve high overall uptime greater than 99.9% should build a pre-production environment to mimic production so that configuration changes can be tested before being pushed into production.

Table 120: Maintaining a Pre-Production Environment

OPS-006	MAINTAIN A PRE-PRODUCTION ENVIRONMENT FOR TESTING ANY CHANGES NEEDED (FIRMWARE, SOFTWARE, HARDWARE) PRIOR TO EXECUTING THE CHANGE IN PRODUCTION.
Justification	Environments typically have stringent uptime due to associated financial penalties or losses that occur from unavailable services. The cost of an outage with a realistic duration often outweighs the pre-production environment cost.
Implication	A pre-production environment will increase the overall cost of the solution.

Monitoring

Nutanix includes a variety of built-in, system-level monitoring functions. The relevant metrics for built-in monitoring are automatically gathered and stored without user intervention required.

Native cluster alerts can be sent from either the individual cluster or Prism Central.

Table 121: Configuring Alerts and Alert Policies

OPS-007	CONFIGURE ALERTS AND ALERT POLICIES IN PRISM CENTRAL, NOT PRISM ELEMENT.
Justification	Creates consistency across multiple clusters and reduces effort when making multiple changes. Also allows for anomaly detection.
Implication	Requires Prism Central to receive alerts.

© 2021 Nutanix, Inc. All rights reserved

When alerts are generated, in addition to raising the alert in Prism, the system can generate an outbound message. The two options available for sending alerts are SNMP and SMTP.

Table 122: Utilizing SMTP for Alert Transmission

OPS-008	UTILIZE SMTP FOR ALERT TRANSMISSION
Justification	Creates consistency across multiple clusters and reduces effort to make multiple changes. Also allows for anomaly detection. Offers more options to customize delivery.
Implication	Requires access to SMTP systems.

The following screen shows how to configure Prism to send alerts to a specified email account via SMTP.

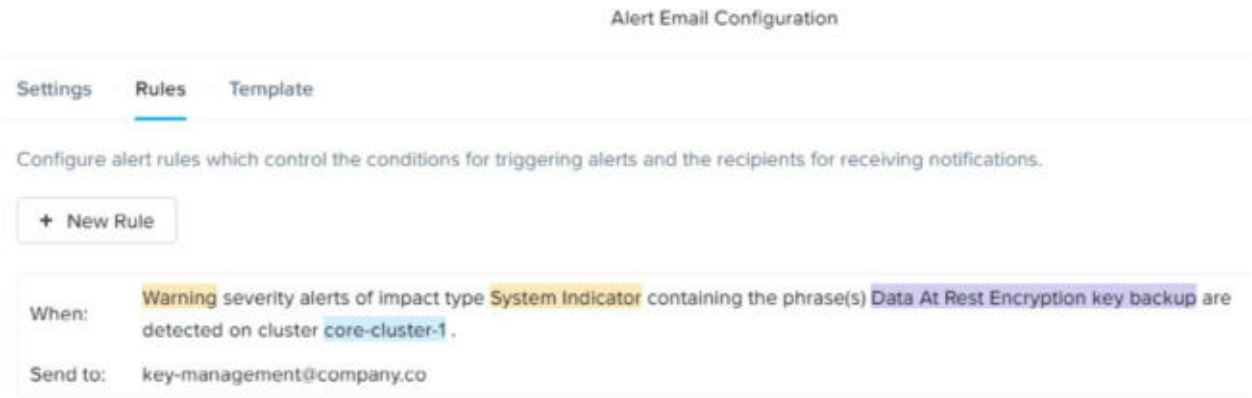


Figure 55: Alert Email Configuration

5. Multi-Datacenter Design for Nutanix Core and BC/DR

Choosing the Right BC/DR Solution

General Guidance

Choosing the right solution for business continuity and disaster recovery is one of the key decisions every architect must make. A mistake can cause significant pain to the business and millions of dollars of losses.

The two key business requirements when designing BC/DR solutions are Recovery Point Objective (RPO) and Recovery Time Objective (RTO). In addition to RPO and RTO – Work Recovery Time (WRT) and Maximum Tolerable Downtime (MTD) should be taken into account.

- RPO: The maximum acceptable amount of data loss measured in time
- RTO: The maximum tolerable amount of time needed to bring a critical system back online
- WRT: The maximum tolerable amount of time that is needed to verify the system and/or data integrity
- MTD: The total amount of time that a business process can be disrupted without causing any unacceptable consequences

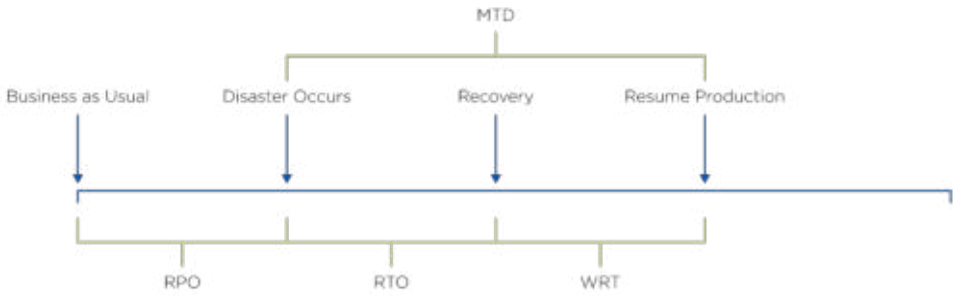


Figure 56: General Guidance for Business Continuity

When designing new BCDR solution make sure to understand your business requirements and technical constraints. Nutanix provides a range of protection mechanisms to help with data protection. The general advice when designing BC\DR is:

If an application has native data protection mechanisms (e.g. Exchange DAG, MS SQL AlwaysOn, Oracle RAC) leverage those to provide application data protection. If an application does not provide a data protection mechanism, use Nutanix built-in protection mechanisms.

The flowchart below can help guide your decision making.

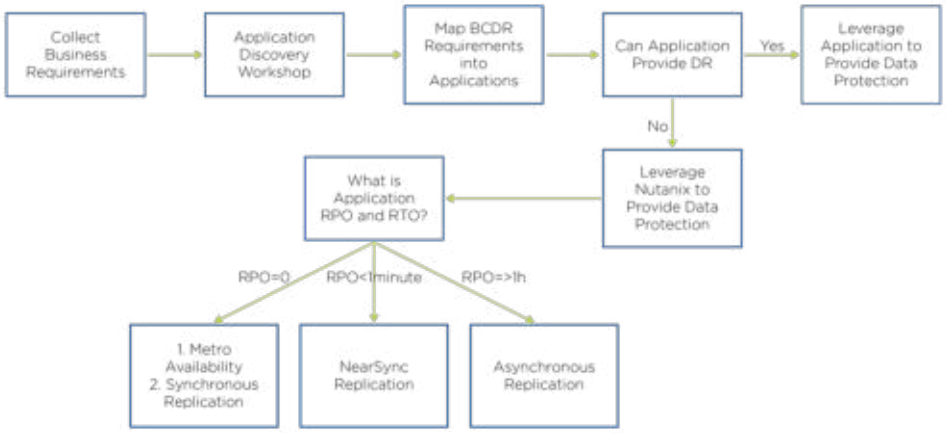


Figure 57: BCDR Workflow

DR Target Considerations

Disaster recovery requires a secondary location that is far enough away from the primary location to ensure that it won't be affected by the disasters you need to protect against but close enough that latency and bandwidth don't become a limiting factor. You can use multi-site DR, as described later to protect workloads using a combination of sites at different distances.

Another important consideration is whether DR will utilize on-premises locations or an option in the cloud:

- On-premises: DR using on-premises target locations gives you complete control over (and responsibility for) your DR operations. However, it may require significant CapEx and ongoing operating expense to maintain adequate resources to support failover if it becomes necessary.

- Nutanix Xi Leap: DR-as-a-Service (DRaaS) offloads much of the complexity and cost of DR. Xi Leap is a simple-to-configure DR service available to Nutanix customers. To learn more visit nutanix.com/leap
- Nutanix Clusters: Nutanix Clusters allow you to operate a full Nutanix environment in the public cloud. You can utilize Nutanix Clusters as a DR target and take advantage of cloud resources while maintaining full control and complete operational continuity with on-premises Nutanix AOS environments. (See the section Nutanix Clusters for Hybrid and Multicloud Operations for more information.)

Multi-Datacenter Design

Datacenter Resiliency Patterns

When operating applications across multiple physical locations for the purposes of improved resiliency, there are two common sets of operating patterns.

Active-Active

In an active/active pattern, two (or more) datacenters provide resources to actively support of a service. For this definition, the application is expected to be fully capable of continuing to service requests in the event one of the active AZs goes down. While it is up to the specific application architecture to determine exactly how the components are split up and their fault/partition tolerance, active sites are typically operated with a low latency constraint of <= 5ms RTT.

In the event of a localized disaster affecting only one availability zone, the application maximum throughput/capability is reduced proportional to the amount of resources lost in the down AZ.

© 2021 Nutanix, Inc. All rights reserved

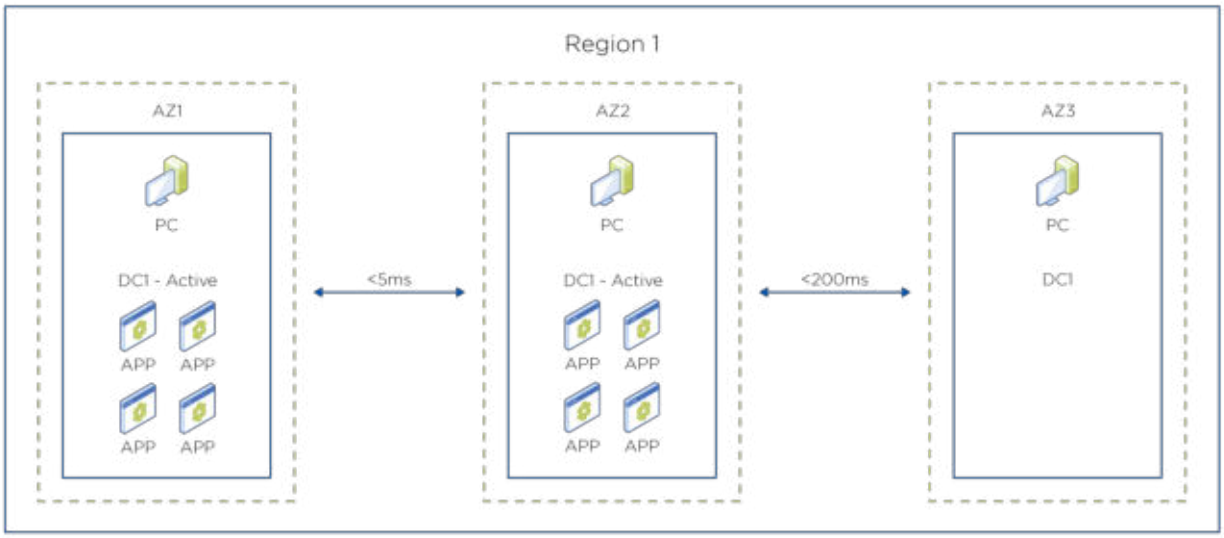


Figure 58: Active – Active

Active-Passive

In an active/passive scenario, only one datacenter has active resources supporting a service. An additional datacenter is available to provide services in the event that the primary datacenter goes down. For some application architectures (sometimes referred to as warm standby) the passive facility has resources configured and ready, however under normal conditions those resources are not servicing user requests. These warm standby applications are delineated from true passive systems, where the data is replicated, but the compute is not ready and waiting.

A hybrid model is also possible, where a portion of the environment in each datacenter is active and a portion is a standby for another site. For example, business unit 1 (BU1) may run actively out of AZ1 and passively out of AZ2 due to the application design, while BU2 is actively running out of AZ1 and AZ2.

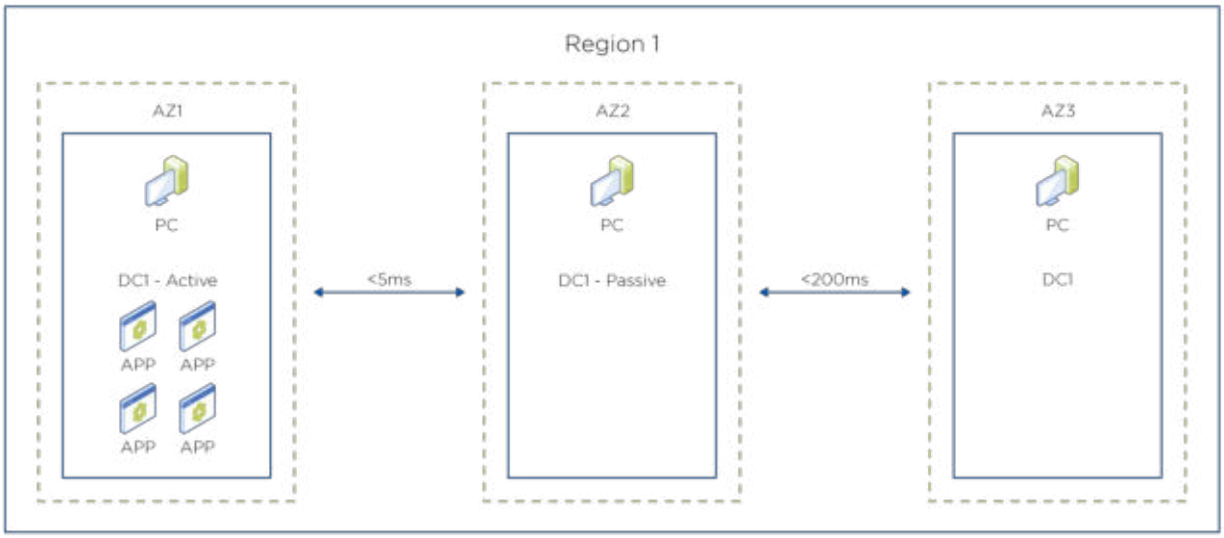


Figure 59: Active - Passive

Multi-Site DR

In either of the above patterns, application components may be further protected via an N+1 site providing DR capacity. Typically, the DR site is located either geographically far away from the primary system inside of the same region, or in another region entirely. This allows for better isolation from a widespread disaster (hurricane, large -scale flooding, tier 1 ISP issues), however the physical distances often create greater latency.

Replication Options

Snapshots & Clones

To provide full coverage data protection, Nutanix includes a number of native features, as well as integration with third-party backup vendors.

As a first level of protection (local recovery), an individual cluster can create a snapshot of a vDisk or an entire VM. Nutanix snapshots are redirect-on-write, where the pointers that indicate which physical disk extent maps to a virtual disk are updated only as data changes. This technique creates little to no performance overhead for a running workload when a snapshot exists. Cloned

entities use the same principle and enjoy the same benefit. Additional data capacity is not used by the snapshot or clone unless changes are made to it.

The minimal overhead makes it low risk to protect as many workloads as necessary with snapshots. Protected VMs can have differing levels of retention and snapshot frequency, allowing for as much flexibility as needed to work around application scheduling requirements.

When choosing a retention schedule, choose the minimum number of snapshots that meet RPO requirements. Each snapshot requires additional metadata space to track the related extents on disk, and excessively deep snapshot chains can increase the background load on the CVM as part of its maintenance processing.

There are two options available for snapshots. Snapshots that invoke a method for quiescence inside the guest VM operating system are termed application-consistent, and those that do not inform the guest VM operating system are termed crash-consistent. AOS can trigger an application-consistent snapshot via Nutanix Guest Tools and/or VMware tools, depending on the hypervisor in use.

Guest limitations may prevent proper quiescence for very large machines (for example Windows VSS cannot handle more than 64 devices/drives concurrently).

Table 123: Utilizing Application-consistent Snapshots

BCN-001	UTILIZE APPLICATION CONSISTENT SNAPSHOTS WHEN NEEDED BY THE APPLICATION
Justification	VM/application quiescence allows for a “clean” snapshot from the application’s viewpoint.
Implication	Application quiescence can potentially take a significant amount of time, and the throughput of the application may be adversely affected. Consistency groups should have less than 20 members

For ESXi hypervisor triggered clones, AOS supports offloading the clone via VAAI. The VM must meet vSphere’s requirements to trigger offloading (VM

must be powered off and the clone must reside on the same datastore as the primary VM).

Local snapshots and clones provide an easy restoration point on the local cluster in the event something goes wrong with the VM itself (OS bluescreen after patches, recover to a point before ransomware encryption, etc.). These snapshots can be instantiated in seconds, allowing for a very short RTO until VM boot. Via API integration, 3rd party backup vendors can utilize Prism REST API endpoints to minimize backup load and space required.

Nutanix also provides the ability to replicate data between clusters, where the data point referenced in the snapshot is stored on one or more remote clusters. As part of the snapshot schedule, these remote clusters can have a different retention period than the primary. An example of this would be keeping three daily copies locally on a performance-oriented all flash cluster, and 14 daily copies on a remote hybrid cluster.

There are two different forms of snapshots to support different modes of replication:

- Full snapshots for asynchronous replication (with RPO of 60 minutes or greater).
- Lightweight snapshots (LWS) for NearSync replication (when the RPO is between 15 minutes and 1 minute).

Full snapshots are efficient for keeping system resource usage low when you are creating many snapshots over an extended period of time. LWS reduces metadata management overhead and increases storage performance by decreasing the number of storage I/O operations that long snapshot chains can cause.

Nutanix snapshots are a vital first element of an overall data protection strategy, however they are not a substitute for a full backup methodology. Many organizations subscribe to what's referred to as the 3-2-1 rule, which mandates 3 copies of data, on 2 backup mediums, with at least 1 copy offsite.

Snapshots can be replicated to multiple destination clusters to provide multiple redundant copies if needed. These arrangements are configured as part of the snapshot schedule by specifying each remote site that should receive a copy, and how many to retain at that site. One to many relationships are allowed (Site

© 2021 Nutanix, Inc. All rights reserved

A-> Site B, Site A -> Site B, Site A –> Site N), but cascading relationships (Site A -> Site B -> Site N) are not supported.

Direct Cluster to Cluster Backup and Recovery

When creating the configuration and schedules for snapshots, there are multiple options. Prism Element on each cluster offers a construct called a Protection Domain (PD), which is a grouping of VMs. During a failover event, all the entities in the Protection Domain are activated at the remote site as a group. Each VM must be placed manually in an applicable PD, either via Prism or by an API call.

The RPO (time between successive snapshots) chosen for these entities will determine whether the system uses traditional asynchronous snapshots, or lightweight snapshots.

Recommendations: Protection Domains

- No more than 200 VMs per PD.
- No more than 10 VMs per PD with NearSync.
- Group VMs with similar RPO requirements in the same PD.

Note that size and capacity of a node can limit the RPO achievable for VMs running on it. See AOS Snapshot Frequency for Nutanix Nodes on the Nutanix Support portal for details.

Consistency Groups

Administrators can create a Consistency Group for VMs and volume groups that are part of a Protection Domain where you want to snapshot all members of the group in a crash-consistent manner. All entities in the same Consistency Group will have their snapshot execute at exactly the same time. Entities that are not in a Consistency Group, but part of the same schedule will typically snapshot close together, but not at the exact same moment (potentially a few seconds apart).

Recommendations: Consistency Groups

- Keep Consistency Groups as small as possible. Limit Consistency Groups to fewer than 10 VMs.

© 2021 Nutanix, Inc. All rights reserved

- Consistency Groups should only be used for applications with a shared state, such as database replicas.
- A Consistency Groups using application-consistent snapshots can contain only 1 VM.

Table 124: Place Nearsync VMs in their Own Protection Domain

BCN-002	PLACE NEARSYNC VM'S IN THEIR OWN PROTECTION DOMAIN.
Justification	NearSync can only have one schedule so place NearSync VMs in their own PD.
Implication	Multiple PDs will need to be created for different NearSync VMs depending on their schedule requirements.

Snapshot Schedule and Retention Policy

Full Snapshots and Async Replication

Your RPO determines how much data you will lose in the event of a failure. You can create multiple schedules for a Protection Domain using full snapshots at various frequencies with different retention policies.

The snapshot interval should be shorter than the desired RPO to allow for failure and recovery, without manual intervention. This is ideally at least twice as fast as your desired RPO, including data transmission timing. For example:

- RPO objective: data should be no more than 4 hours old
- Snapshot schedule: every 4 hours, starting at 12 AM, with retention on a remote cluster
- Average replication of changes takes approximately 30 minutes.
- The source site experiences an outage at 4:20 AM

In the scenario above because replication of the 4AM snapshot has not completed, the most recent snapshot available is older than 4 hours, violating the RPO goal. A 2-hour (or more frequent) snapshot schedule would avoid this RPO violation.

© 2021 Nutanix, Inc. All rights reserved

Table 125: Configure Snapshot Schedules to be More Frequent

BCN-003	CONFIGURE SNAPSHOT SCHEDULES TO BE MORE FREQUENT THAN THE DESIRED RPO
Justification	Cross cluster replication prevents the instant transfer of snapshot contents, increasing the time before snapshot contents are available.
Implication	Desired RPO may be achievable given technology limitations, or the RPO may need to be adjusted to accommodate replication time.

Table 126: Configure Snapshot Schedules to Retain Lowest Number of Snapshots

BCN-004	CONFIGURE SNAPSHOT SCHEDULES TO RETAIN THE LOWEST NUMBER OF SNAPSHOTS WHILE STILL MEETING THE RETENTION POLICY.
Justification	Metadata space management on a cluster is more efficient with a lower number of snapshots.
Implication	Multiple schedules should be created for the same Protection Domain at different levels rather than a simple daily schedule.

LWS and NearSync Replication

Nutanix offers NearSync with a telescopic schedule (time-based retention). When the RPO is set to be #15 minutes and #one minute, you have the option to specify the maximum retention. The system will automatically roll up intermediate recovery points within the retention window specified, based on the table below. Multiple schedules cannot be created with NearSync.

The following table represents the schedule to save recovery points for 1 month:

Table 127: Schedule to Save Recovery Points

TYPE	FREQUENCY	RETENTION
Minute	Every minute	15 minutes
Hourly	Every hour	6 hours

© 2021 Nutanix, Inc. All rights reserved

Daily	Every 24 hours	7 days
Weekly	Every week	4 weeks
Monthly	Every month	1 month

See Nearsync Requirements and Limitations on the Nutanix Support Portal for more information.

Synchronous Data Transfer

For applications that need to meet an RPO goal of zero (simultaneous data updates at multiple sites), Nutanix supports a synchronous replication option. Configuration of synchronous replication also utilizes Protection Domains as described above.

Configuration of synchronous replication requires a container on the source and destination with the same name. The two clusters participating in synchronous replication are required to be within 5ms round trip time (RTT) of each other.

By default, systems protected synchronously will also automatically create a periodic snapshot schedule. These snapshot references are used to reduce the recovery time in the event that the data must be resynchronized, for example after a WAN outage.

Intermixing Protection Levels

Depending on the version of AOS in use, hypervisor, number of sites, and whether you are using Protection Domains or Protection Policies, the supported configuration may differ. For Protection Domains (Prism Element), refer to the following two tables for supported intermixing of protection levels.

Table 128: Protection Domain 2-Site Support Matrix

REPLICATION TYPE	2 SITES
Async	AHV, ESX
NEAR Sync	AHV, ESX
Synchronous	ESX

Table 129: Protection Domain Multi-site Support Matrix

REPLICATION COMBINATION	DIFFERENT ENTITIES	SAME ENTITIES
Multiple Async only combination (1:N)	AHV, ESXI	AHV, ESX
Multiple Near-Sync only combination (1:2)	AHV, ESX	AHV, ESX
Multiple Sync only combination (1:2)	ESX	Not Supported
Async + Near Sync (1:2) *	AHV, ESX	AHV, ESX
Async + Sync (1:2)	ESX	ESX
Sync + Near Sync (1:2)	ESX	ESX
Async + Near Sync + Sync (1:3) *	ESX	ESX

* STS only, 5.19 and later

If you are using Prism Central Protection Policies, the following two tables show the two site and multi-site intermixing support matrix.

Table 130: Protection Policy 2 Site Support

REPLICATION TYPE	2 SITES
Async	AHV, ESX
Near Sync	AHV, ESX
Synchronous	AHV*

* STS only, 5.17 and later

Table 131: Protection Policy Multi-site Support Matrix

REPLICATION COMBINATION	DIFFERENT/SAME ENTITIES
Multiple Async only combination (1:N)*	AHV, ESXI
Multiple Near-Sync only combination (1:2)	AHV, ESX
Multiple Sync only combination (1:2)	Not Supported

© 2021 Nutanix, Inc. All rights reserved

Async + Near Sync (1:2) *	AHV, ESX
Async + Sync (1:2)	Not Supported
Sync + Near Sync (1:2)	Not Supported
Async + Near Sync + Sync (1:3)	Not Supported

* STS only, 5.19 and later

For additional details, see Data Protection and Recovery with Prism Element on the Nutanix Support Portal.

In all cases, a failover action causes all items within the identified PD to be made available at the remote site as a collective unit. Only related entities should be grouped into a PD.

Table 132: Group Applications Together in Unique Protection Domains

BCN-005	GROUP APPLICATIONS TOGETHER IN UNIQUE PROTECTION DOMAINS. KEEP NUMBER OF VM'S PER PROTECTION DOMAIN AS SMALL AS REASONABLY POSSIBLE
Justification	Small groupings of related entities help minimize accidental collateral damage during a partial failover and reduce the need to track complex intermingled scheduling.
Implication	Entire site failover will require activating a larger number of Protection Domains, potentially increasing the time required to bring all entities online.

Leap

Protection Domains are a construct within Prism Element. One of their limitations however is that they don't provide additional controls around VMs associated with a large-scale failover, like power-on order control, or changing the VM IP address due to site change.

Prism Central introduced a feature called Leap, which allows for more advanced constructs relating to replication and failover.

© 2021 Nutanix, Inc. All rights reserved

Protection policies offer a way to configure replication grouping using categories. VMs that match a category are automatically protected, versus a Protection Domain where they must be manually added.

Leap also provides Recovery Plans. Recovery Plans allow for controlling the failover process. They allow you to specify:

- Power-on order
- Mapping networks when using DHCP
- Changing static IPs for sites that don't have access to stretch L2 networking
- Scripts to perform in guest actions as part of failover (if necessary)
- Perform test failovers and validation of the configuration without downtime
- With Leap, you can perform test failovers and validate your configuration without downtime.

Full coverage of leap is out of scope for this document. Additional details can be found in the Leap Administration Guide found on the Nutanix Support Portal.

Additional Options

Nutanix offers a number of additional options for ensuring business continuity that are beyond the scope of this design:

- Xi Leap: Nutanix (or partner) hosted datacenters provide a disaster recovery as a service option. This utilizes the Leap feature of Prism Central, where data is replicated to a managed cloud service location similar to how it would be replicated to an on-prem DR location.
- Nutanix Mine: As part of the conceptual 3-2-1 rule, Nutanix supports multiple backup vendors that can integrate via API to offload backup. Mine offers a bundle of backup vendor software and Nutanix platform to provide storage for the retention of backup data, independent of the primary cluster(s) in use. (See the section Nutanix Mine for Integrated Backup for more information on Mine.)

© 2021 Nutanix, Inc. All rights reserved

Disaster Recovery Decision Tree

A good DR solution should involve different levels of service for DR because applications will have different levels of criticality to the business and therefore different requirements for their disaster recovery needs. As a result, this document will include multiple architectures specific to meet a variety of scenarios that as a group can be configured to meet the requirements for all application needs. The following flow chart walks through the decision tree for choosing the best Nutanix DR solution to meet your requirements. Each DR solutions is discussed in the sections that follow.

© 2021 Nutanix, Inc. All rights reserved

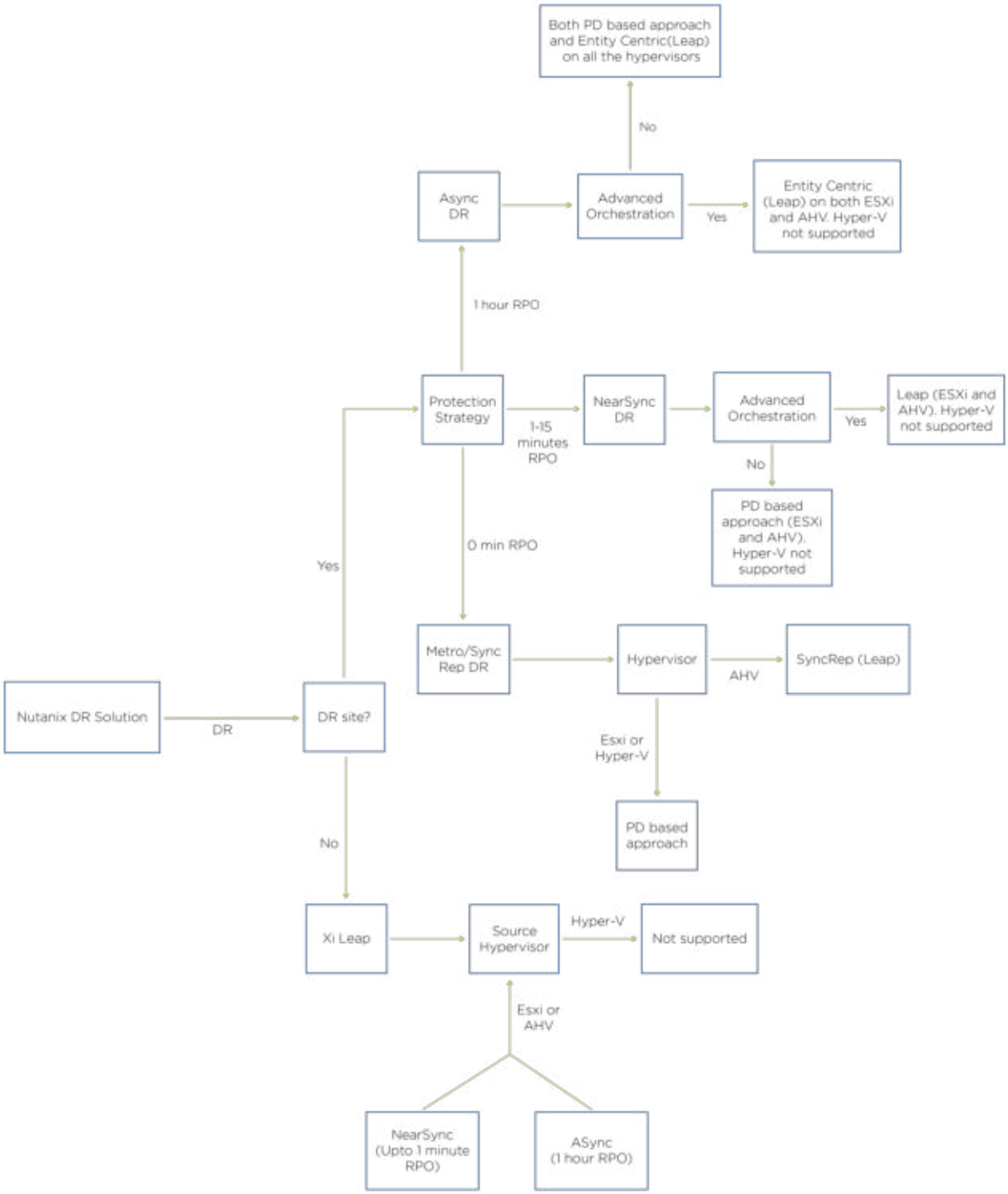


Figure 60: Disaster Recovery Decision Tree

Nutanix DR solutions

Nutanix Metro Availability

Metro Availability (MA) allows vSphere administrators to leverage hypervisor clustering technology across datacenters. This type of configuration presents two distinct Nutanix cluster to vSphere as a single stretched cluster, and it helps to minimize downtime during unplanned outages. In the event an entire site goes down, vSphere HA automatically starts the virtual machines on the secondary site if a witness has been properly configured.

This configuration is the most complicated choice, requiring stretched layer 2 networking of some form between sites. In addition, while this provides the most rapid failover, the configuration is highly dependent on the network connectivity being resilient and providing a high enough throughput to accommodate application change rates. Systems with frequent intermittent network issues may be better served via a NearSync style connection that is less impacted by network drops.

See the Metro Availability Guide on the Nutanix Support Portal for more information.

Table 133: Application Requires RPO

BCN-006	APPLICATION REQUIRES RPO=0 AND RTO NEAR ZERO
Justification	Business critical applications requires zero data loss and minimal to zero recovery time.
Implication	Nutanix MA requires high bandwidth and low latency (under 5ms). To comply with near 0 RTO, stretched layer 2 networks are required. Nutanix MA is supported only with VMware ESXi, which may bring additional management overhead and increase licensing costs.

Table 134: Application Requires Zero Downtime

BCN-007	APPLICATION REQUIRES ZERO DOWNTIME DR AVOIDANCE SOLUTION
Justification	Business critical application requires zero data loss and zero recovery time during DR avoidance event.
Implication	Nutanix MA requires high bandwidth and low latency (under 5ms). To comply with ZERO downtime during DR avoidance, stretched layer2 networks are required. Nutanix MA is supported only with VMware ESXi, which may bring additional management overhead and increase licensing costs. Nutanix AOS enterprise license is required. With Nutanix Metro Availability you can protect entire datastore. You cannot choose VM to protect.

Synchronous replication

Nutanix synchronous replication is available with Nutanix AHV (using Leap and AOS 5.17) or VMware vSphere. With synchronous replication workloads are protected with RPO=0. Synchronous replication is used when you cannot provide spanned L2 across datacenters (required by Nutanix Metro Availability) but still need to provide RPO=0 to applications.

Table 135: Application Requires RPO

BCN-008	APPLICATION REQUIRES RPO=0
Justification	Applications require 0 RPO and minimal RTO.
Implication	Nutanix Sync replication requires high bandwidth and low latency (under 5ms) datacenter connectivity. Nutanix Sync replication with AHV is supported only with Nutanix Leap. Nutanix synchronous replication requires Nutanix AOS enterprise license or Leap Advanced addon license.

© 2021 Nutanix, Inc. All rights reserved

NearSync replication

NearSync replication enables you to protect your data with an RPO as low as 1 minute. You configure a Protection Policy with NearSync replication by defining the VMs or the Categories of VMs. The policy creates a Recovery Point of the VMs in minutes (1–15 minutes) and replicates it to the recovery site.

Table 136: Provide RPO Between 1 and 15 Min

BCN-009	PROVIDE RPO BETWEEN 1 MIN AND 15 MIN TO AN APPLICATION
Justification	Provide low RPO to storage intensive workloads over limited connectivity.
Implication	A cluster is NearSync capable if the capacity of each SSD in the cluster is at least 1.2 TB. All-flash clusters do not have any specific SSD sizing requirements. Nutanix NearSync replication with AHV is supported only with Nutanix Leap. Nutanix NearSync replication requires Nutanix AOS enterprise license or Leap Advanced addon license.

Asynchronous replication

Nutanix asynchronous replication enables you to protect your data with an RPO of 1h or more. You can configure a Protection Policy or Protection Domain with asynchronous replication by defining the VMs or the Categories of VMs. The policy or protection domain scheduler creates a Recovery Point of the VMs in minutes (60 minutes) and replicates it to the recovery site.

Table 137: Application Requires RPO

BCN-010	APPLICATION REQUIRES RPO=>1H
Justification	Provide RPO=>1h to the application utilizing low bandwidth\high latency datacenter links.
Implication	Nutanix asynchronous replication requires Nutanix AOS enterprise license or Leap Advance addon license.

© 2021 Nutanix, Inc. All rights reserved

DR Patterns

1-to-1

For smaller environments with only a pair of locations, protect applications between two availability zones. Availability zones can be the same region or across regions. The same workload or application can be protected with different RPO thresholds. Choosing availability zones locations (same region or across regions) defines RPO value for protected application\workloads. If application or workload requires RPO=0, choose availability zones in the same region (latency requirement for RPO=0 is 5ms RTT or less).

1. Two AZ within the same region

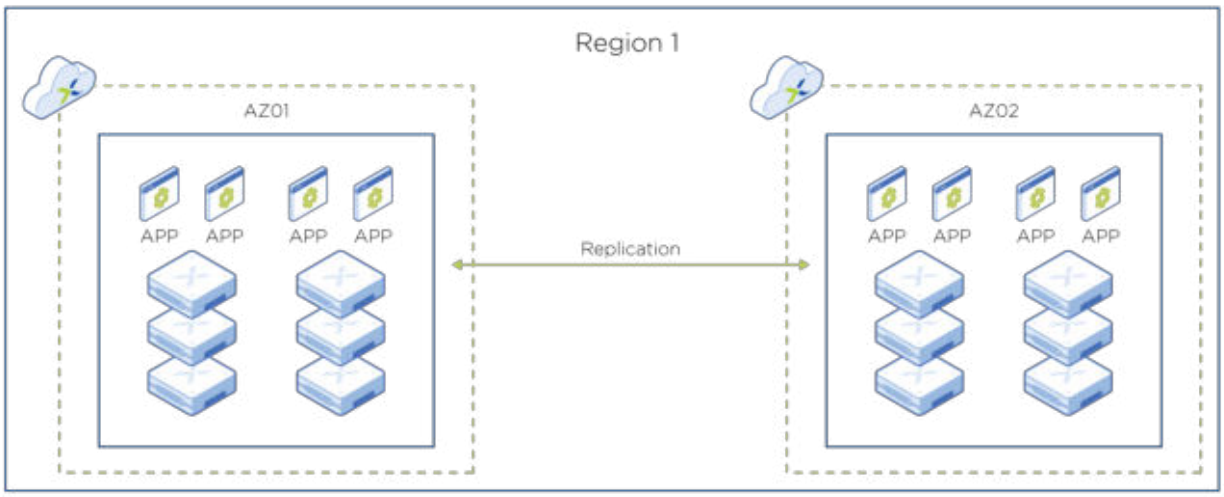


Figure 61: Two AZ within the Same Region

1. Two AZ across multiple regions

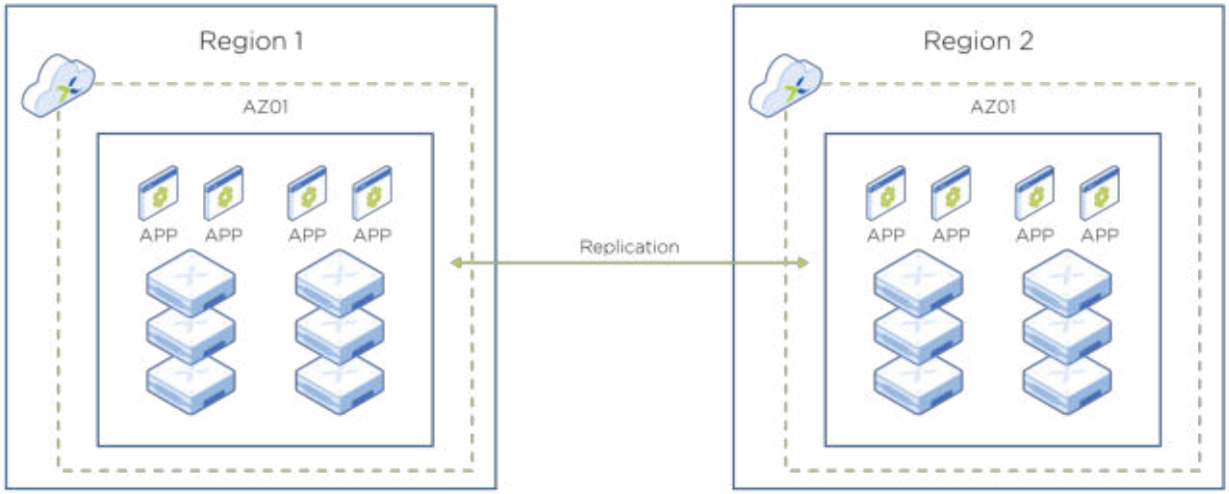


Figure 62: Two AZ across Multiple Regions

1-to-Many

The 1-to-Many DR pattern maps a single source Nutanix cluster to multiple DR target Nutanix clusters. This use case is applicable where an application requires multiple levels of data protection and at the same time providing high uptime and minimal to zero data loss. Availability zones can be located within the same region or across multiple regions. Choosing the availability zone location (same region or across regions) defines minimum RPO value for protected application\workloads.

If an application or workload requires RPO=0 and protection against natural disasters like hurricanes or earth quakes, choose an availability zone in the same region to provide RPO=0 (latency requirement for RPO=0 is 5ms RTT or less) and second availability zone in different region (to provide application resiliency against wide spread natural disasters like hurricanes, floods or earth quakes).

1. Multiple AZ within a single region

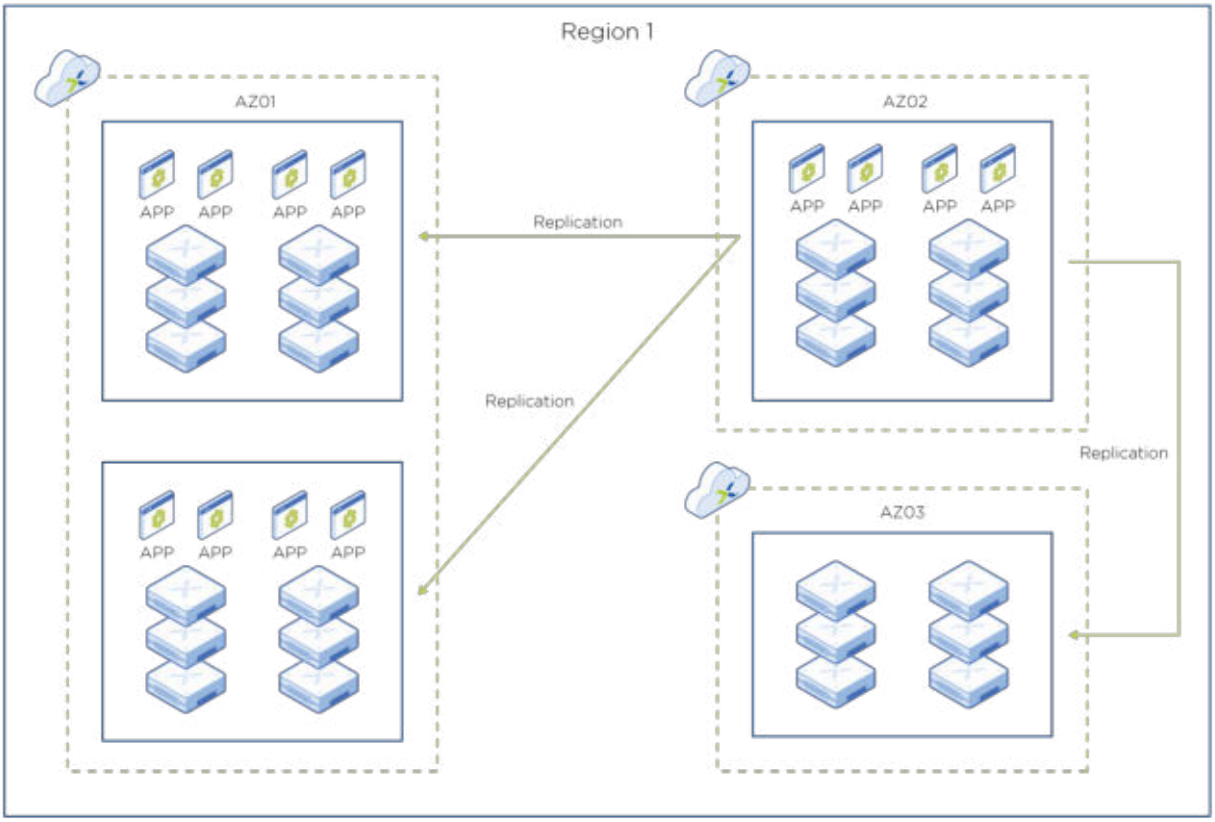


Figure 63: Multiple AZ within a Single Region

1. Multiple AZ across multiple regions

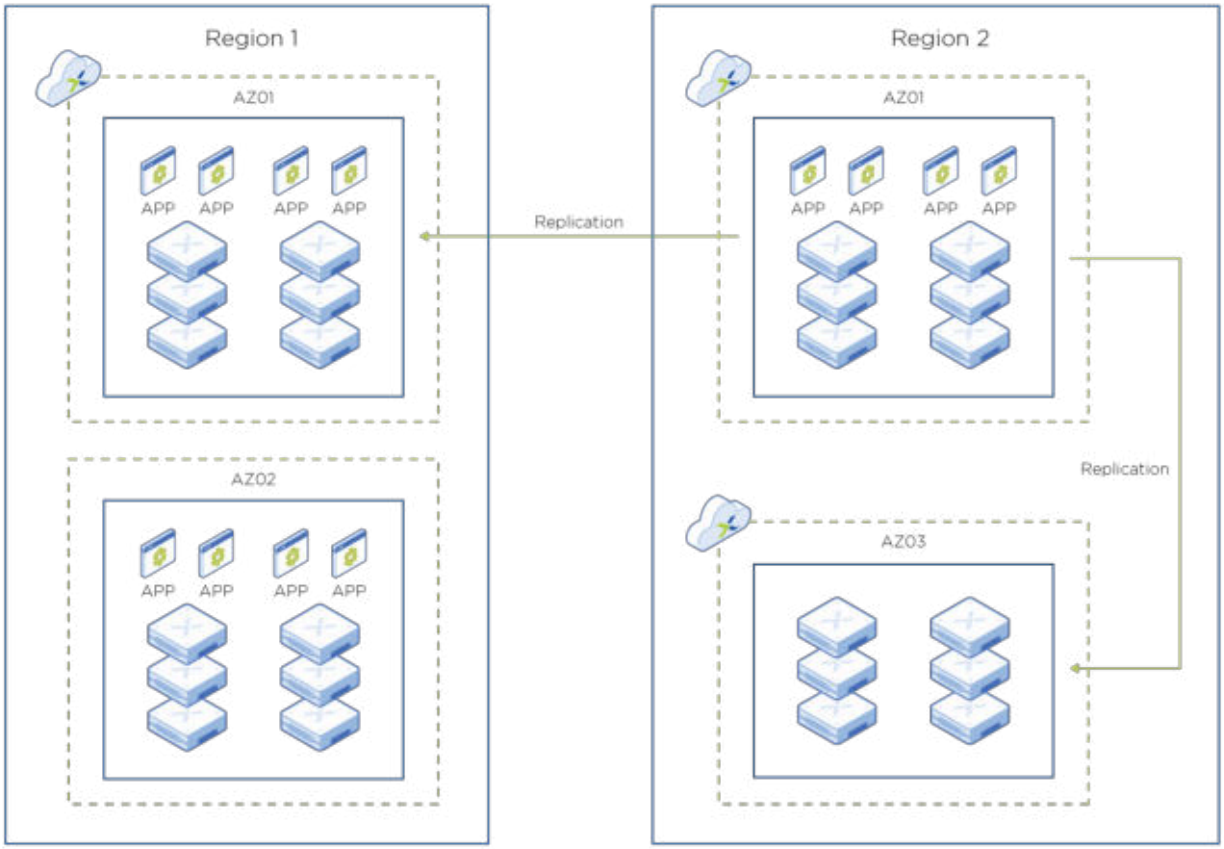


Figure 64: Multiple AZ across Multiple Regions

Many-to-1

The Many-to-1 DR pattern maps multiple source Nutanix clusters to single DR target Nutanix cluster. The best use case for the many-to-1 DR pattern is to protect multiple ROBO locations to a single Nutanix target in a DR hub, which is usually located in a regional or GEO datacenter. Each ROBO location is a separate availability zone which replicates data to the DR hub (which is located in separate availability zone) but the same region.

1. Multiple AZ within a single region

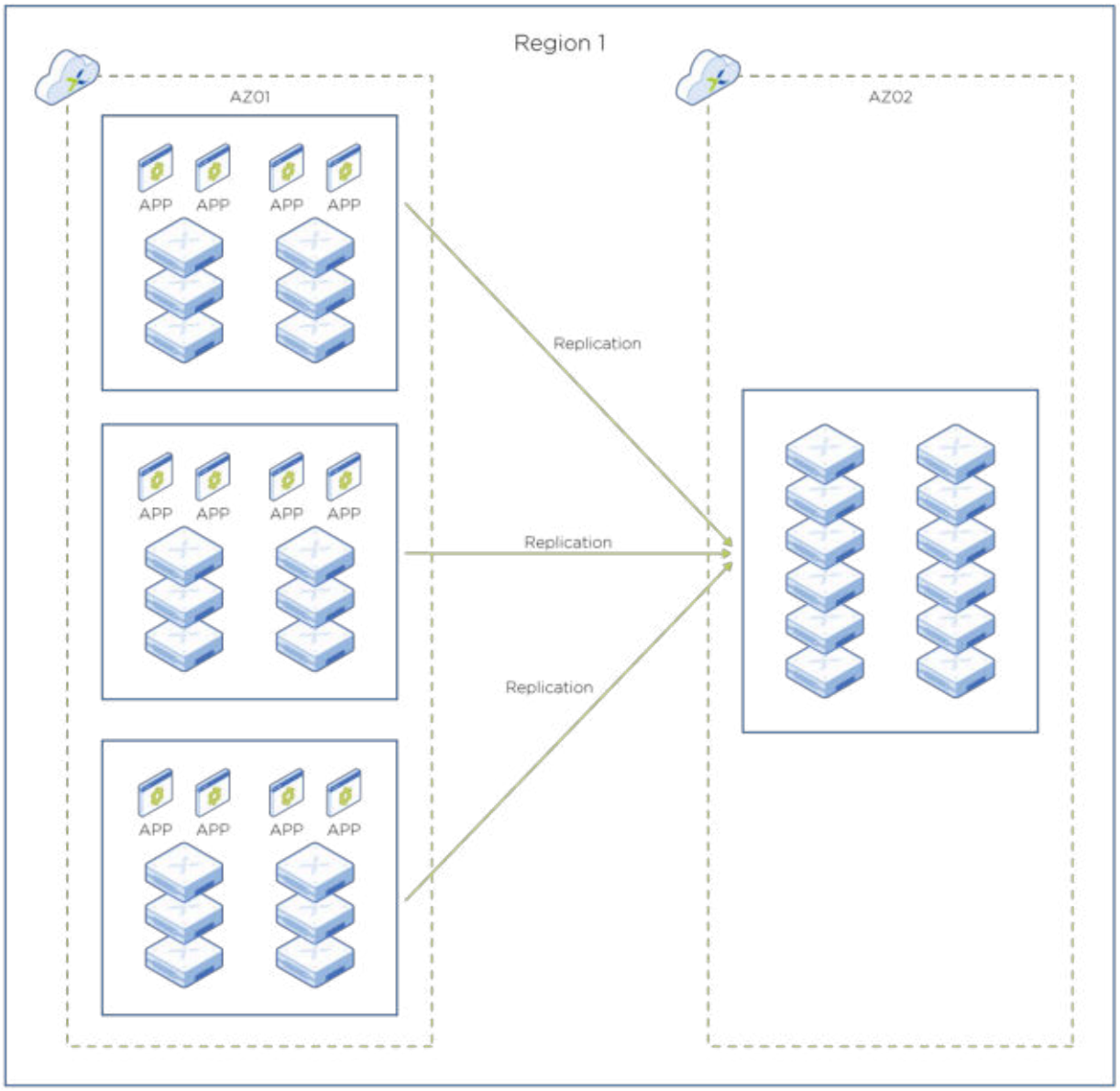


Figure 65: Multiple AZ within a Single Region

1. Multiple AZ across multiple regions

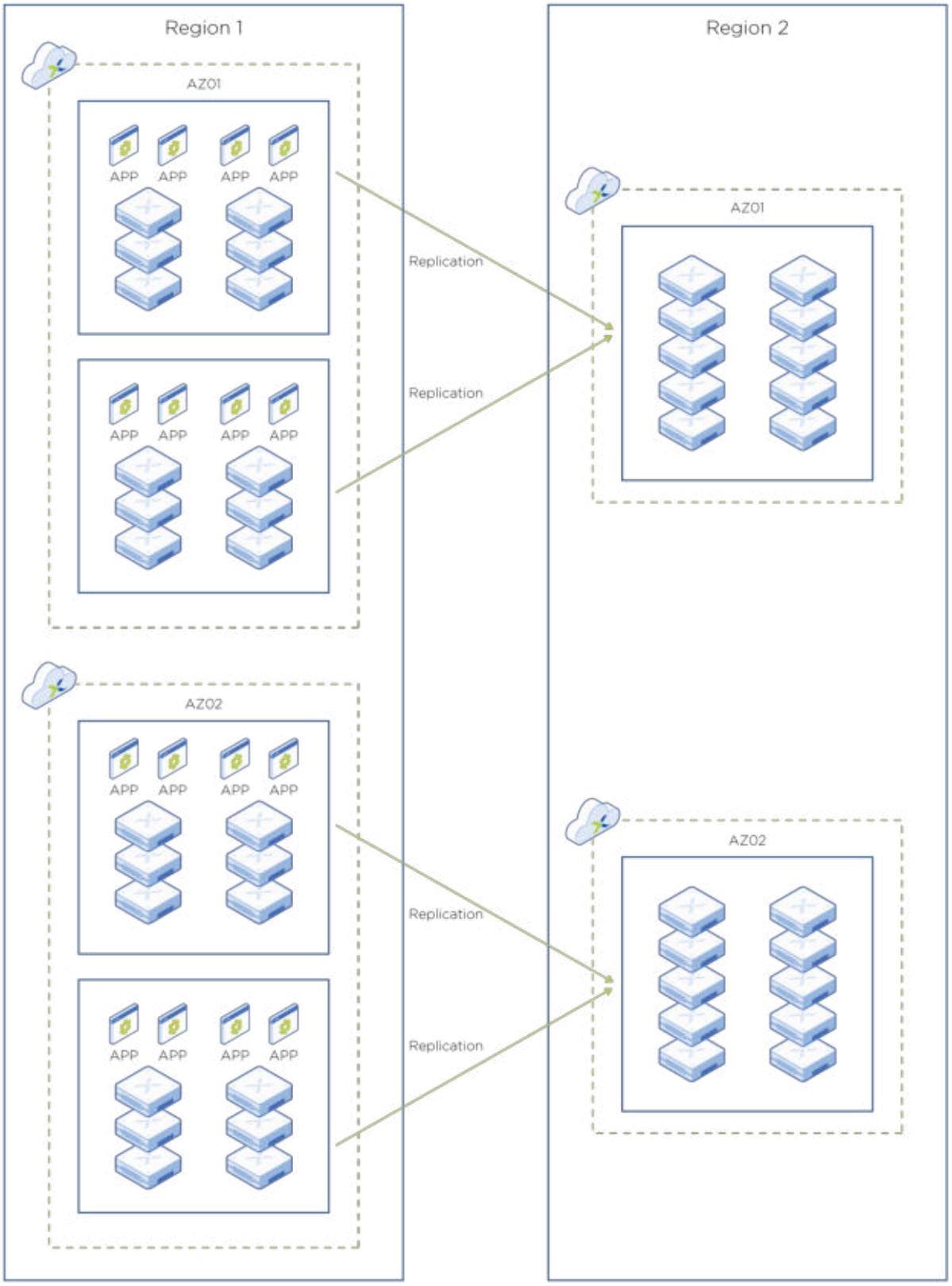


Figure 66: Multiple AZ across Multiple Regions

© 2021 Nutanix, Inc. All rights reserved

Many-to-Many

Description

The Many-to-Many DR pattern maps multiple source DR clusters with multiple target DR Nutanix clusters.

This pattern is suitable to applications which can run concurrently in more than one location. These applications which are require zero data loss and zero to minimal downtime. Leveraging multiple RPO schedules (synchronous, asynchronous, and nearsync) you can protect each application\workload with the most appropriate method to meet availability requirements.

1. Single region

© 2021 Nutanix, Inc. All rights reserved

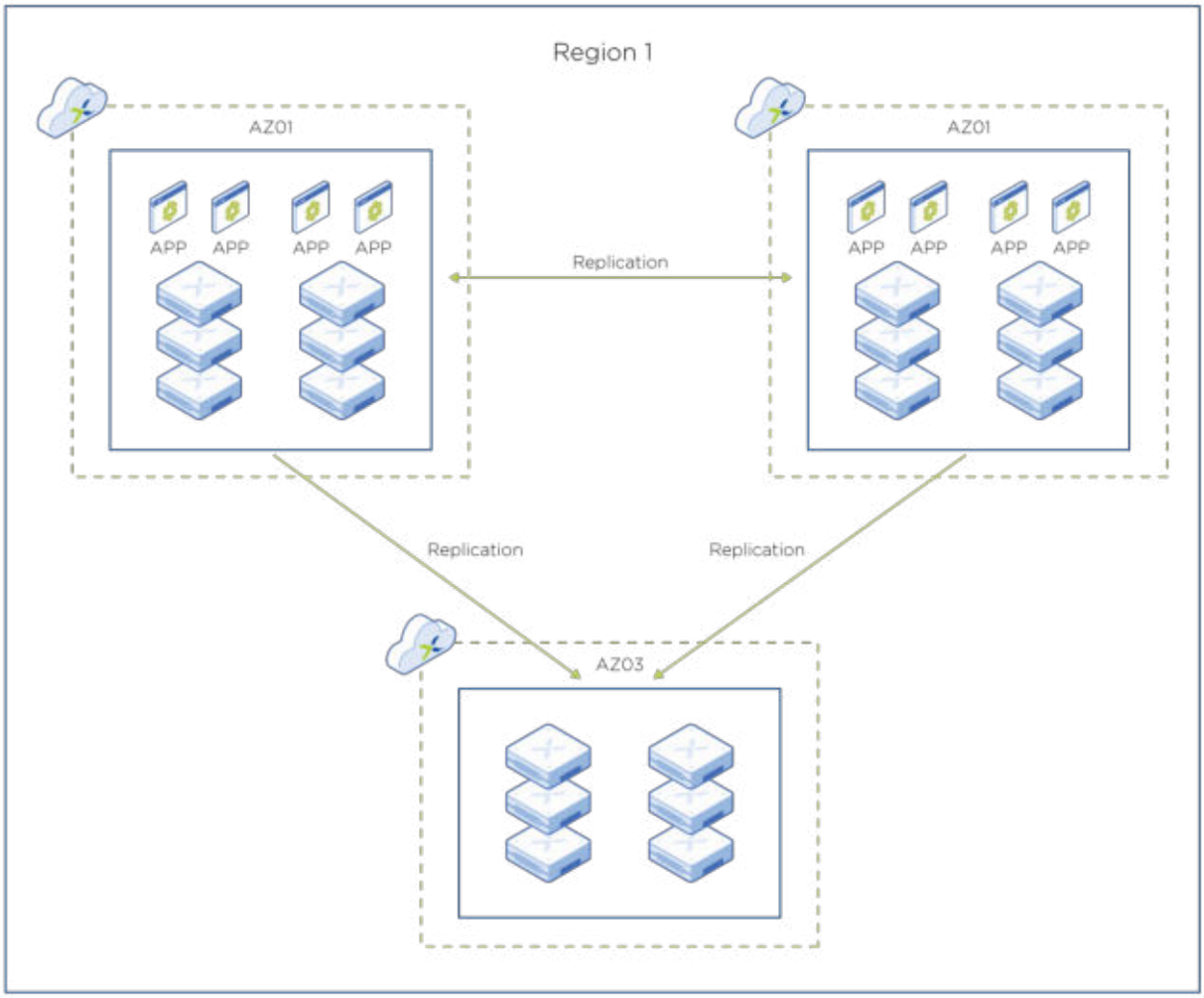


Figure 67: Single Region

1. Multiple regions

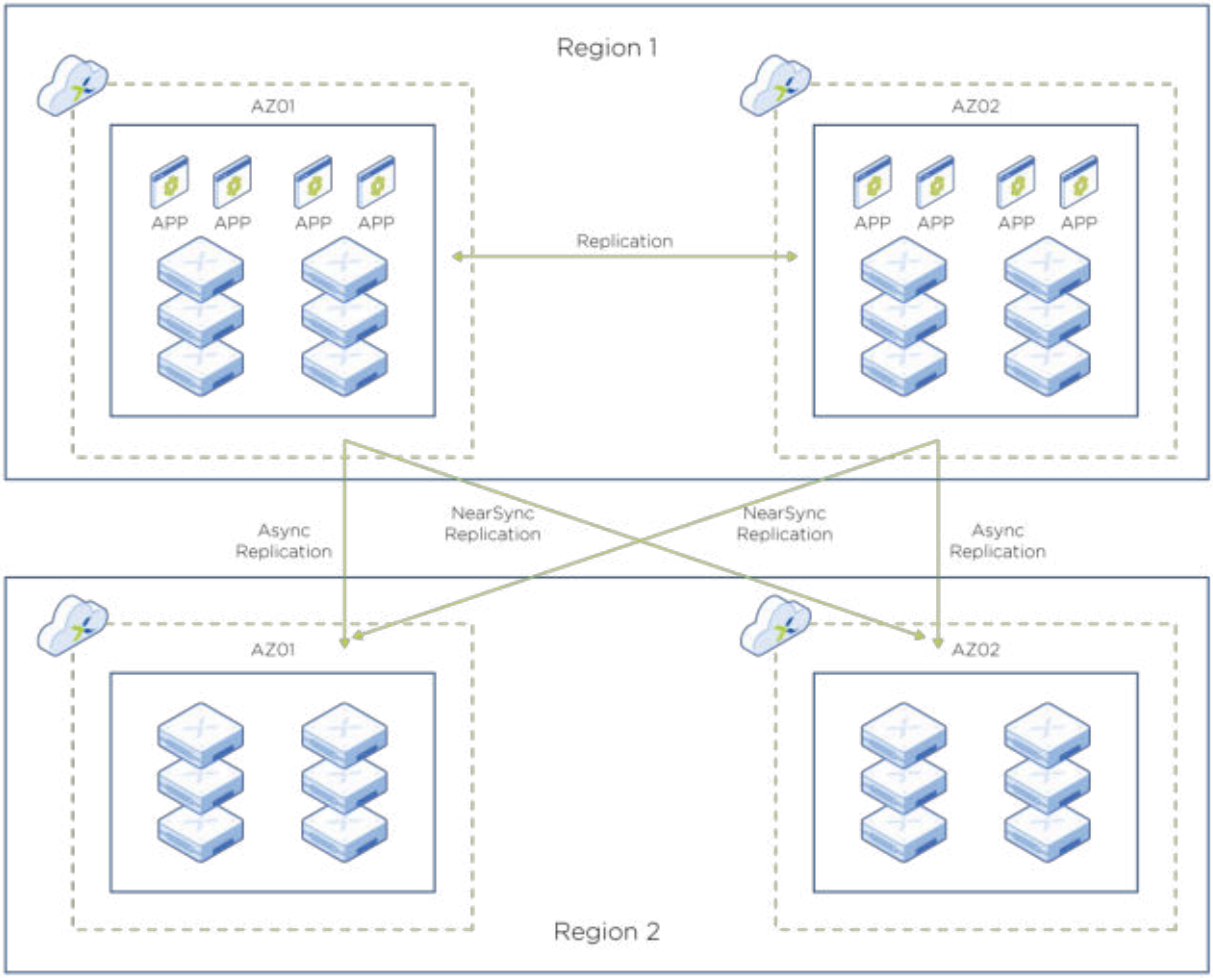


Figure 68: Multiple Regions

ROBO Patterns

For remote and branch offices, there are a few specific considerations. For sites utilizing 3 or more nodes in a cluster, the operation of that cluster is similar to a datacenter, while typically constrained by bandwidth/latency and facilities limitations.

A limited number of models are supported in either a single-node (no HA) or two-node (basic HA) configuration. These configurations are suitable for small footprint sites running a limited number of VMs (5-10, depending on load).

A two-node solution offers the same capacity sizing options as one-node, while allowing for failover in the event one of the nodes goes down. Automatic failover of VMs will require connecting the two-node cluster with a remotely hosted witness VM.

In addition, there are limitations on feature support related to one and two-node clusters. Details on these are maintained in KB 5943 on the Nutanix Support Portal.

Some of the most relevant limitations are:

- Lack of expandability (cannot grow to 3-nodes or larger without re-foundation)
- Limited VM sizing capacity due to limitations of the supported hardware models
- RPO support of 6 hours or greater

For ROBO sites, this creates a decision point. One and two-node solutions are appropriate for systems where a reduced feature set and performance still meet necessary requirements. Two-node systems can have their intra cluster traffic configured using directly connected links to avoid the need for high speed switches. This reduction in equipment is a valid tradeoff to sustain a lower price point compared to traditional 3+ node datacenter clusters.

An additional option is per VM ROBO licensing where the physical hardware is still composed of 3+ hosts. Each VM is explicitly licensed and limited to 32 GB of RAM.

Table 138: Decide on Cluster Licensing Model

DCD-001	DECIDE ON CLUSTER LICENSING MODEL (CBL OR PER VM)
Justification	The two licensing models cannot be mixed on a single cluster.
Implication	Per-VM ROBO licensing model allows for greater hardware capacity and feature set capabilities, while potentially increasing the footprint and hardware cost.

© 2021 Nutanix, Inc. All rights reserved

Table 139: Decide on Cluster Type Used

DCD-002	DECIDE ON CLUSTER TYPE USED (IF NOT ROBO PER VM)
Justification	One and two-node clusters are restricted to specific models, 3+ node models are not restricted.
Implication	Limitations imposed by a one/two-node solution cannot be corrected later (in place) by upgrading to a 3+ node.

Fan Architecture

In a fan-in/out style architecture, multiple ROBO sites are configured to replicate data back to a centralized location. The main location is normally larger, with enough capacity to meet the storage requirements for all the remote sites' expected replication and retention needs.

When building out a central location, at a minimum it needs enough storage capacity to accommodate the data from each ROBO site that requires protection. An important decision is whether to provide compute capacity to actively sustain the workload in the central site in a DR scenario, or only to provide capacity to store the data.

Table 140: Decide Whether to Provide Compute for ROBO Failover

DCD-003	DECIDE WHETHER TO PROVIDE COMPUTE FOR ROBO FAILOVER
Justification	A data replication only ROBO DR strategy will be sustainable by a smaller number of data dense storage nodes. An active failover capable ROBO DR strategy will require additional compute capacity to bring the failed services online in the event of a disaster.
Implication	Failover compute capacity must be reserved, potentially increasing solution cost and footprint.

Table 141: Decide on ROBO Failover Compute

DCD-004	DECIDE ON ROBO FAILOVER COMPUTE OVER COMMIT

Justification	The recovery site will likely only have to provide compute failover capacity for a limited number of concurrent site failures.
Implication	For a centralized location serving many ROBOs, the added cost of failover hardware may be relatively modest and worth the investment.

DR Pairing

For configurations lacking a centralized replication target, an alternative is to create replication partner sites. These sites are typically within a region/availability zone and may have improved network speeds between sites (for example via metro ethernet) versus to a main datacenter. Each of the paired sites should be far enough apart from each other to support recoverability requirements in the event of a disaster.

This configuration can also be used for ROBO in a different country/legal entity than the main datacenter, where data privacy laws make replication back to a larger datacenter impossible.

Control Planes

Control plane design in a multi-site environment is a delicate balance between having a centralized view of everything versus a completely disaggregated view. This design suggests striking a balance, with separate control planes when necessary, to meet availability and product (e.g. VMware SRM or Prism Central) requirements.

vCenter

The guidance for vCenter placement in a multi-datacenter environment is to generally keep the vCenter instance and the clusters its managing within the same availability zone. An exception to this is in ROBO designs where having a vCenter at each remote site is not always desirable since it lacks a single management point for a larger number of sites.

If protecting workloads across availability zones—whether in the same region or across regions—you need to have at least a single vCenter to manage workloads at the source and target sites. This allows for Nutanix DR playbooks

or VMware Site Recovery Manager (SRM) to automate the failover of these workloads.

Prism Central

The guidance for Prism Central (PC) placement is to have at least one PC deployment within each availability zone depending on the number of VMs. This allows the PC instance to manage local clusters and resources while utilizing DR runbook automation to connect to PC instances in other regions for protecting workloads. Having separate PC instances at the source and target locations is a requirement for DR runbook automation available within PC. The only time you would deploy more than one PC per availability zone is when you scale past the limits of a single PC.

There are only two situations where you might consider having a PC instance manage clusters in other availability zones:

- ROBO architecture where each site typically has a small cluster and it's desirable to have a single central PC instance to manage them.
- If you have multiple AZs and are not replicating between them, they can be managed by a single PC instance.

Multi-DC Networking

Networking is critical for disaster recovery, since the network forms the backbone carrying all replicated data. Each component in the path from the physical layer of the source cluster, to the physical layer of the destination cluster must be considered. Especially important are the site-to-site links that connect protected sites.

Each Nutanix replication technology has specific requirements to ensure that the connection between sites has sufficient bandwidth, low enough latency, and meets redundancy requirements. Please note that the formulas below do not take into account link latency, and as such, replication over links with high latency can see significantly reduced performance.

The following table summarizes the required networking characteristics for each replication type:

© 2021 Nutanix, Inc. All rights reserved

Table 142: Required Networking Characteristics

REPLICATION TYPE	LATENCY REQUIREMENTS	BANDWIDTH CALCULATION	NOTES
Async DR	NA	Bandwidth needed = (RPO change rate * (1 - compression on wire savings %)) / RPO Max of 800mbps per node	
Near Sync DR	NA	Bandwidth needed = (RPO change rate * (1 - compression on wire savings %)) / RPO Max - Unbounded	
ESXi Metro Availability	Sites: 5msec RTT Witness: 200msec	Metro Container data change rate Max - unbounded	Redundant site links recommended Witness: Separate links
AHV Synchronous Replication	Sites: 5msec RTT	Data change rate Max - unbounded	Redundant site links recommended

In addition to the characteristics of the network transport between protected sites, careful attention must be paid to the VM networking at each site. When a workload moves from one site to another there are many possible options for maintaining workload connectivity to the required customers or endpoints.

© 2021 Nutanix, Inc. All rights reserved

Technologies like stretched networking, load balancing, NAT, and more should be considered to provide seamless replication and application connectivity.

Bandwidth Requirements

Replicating data between sites requires that enough bandwidth is available to replicate all of the data changed during the RPO window within the RPO interval. Refer to the Data Protection and Disaster Recovery Guide for a complete explanation of bandwidth calculation.

Calculate this data change rate for both asynchronous and NearSync replication. Here is an example of the bandwidth needed for an RPO of 1 hour with 15GB of data changed during that hour. We assume 30% savings for compression.

Table 143: Bandwidth Needed for an RPO

BANDWIDTH NEEDED = (RPO CHANGE RATE * (1 - COMPRESSION ON WIRE SAVINGS %)) / RPO
Example: RPO: 1hr (3,600s) RPO Change Rate: 15GB Compression Savings: 30% (15GB * (1 - 0.3)) / 3,600s (15GB * 0.7) / 3,600s 10.5GB / 3,600s (10.5GB * 1,000 MB/GB * 8 bits/byte) / 3,600s - converting to Mb/s 84,000 Mb / 3,600s = 23.33 Mb/s Bandwidth needed = 23.33 Mb/s

The bandwidth needed is easier to calculate when the RPO is zero, since it is equal to the data change rate. Data will be compressed before being sent to another site, so assume compression in this calculation as well.

Table 144: Calculate Required Storage Replication Bandwidth

MULTI-NET-001	CALCULATE REQUIRED STORAGE REPLICATION BANDWIDTH BASED ON RPO.
Justification	Bandwidth must be available to keep up with data change rate and meet the desired RPO.
Implication	Achieving the desired RPO can require significant bandwidth and requires knowledge of the workload's current usage.

In addition to storage replication between sites, bandwidth must be allocated for workload traffic and application level replication traffic. First determine how much workload traffic and application replication traffic will exist. This is largely dependent on the applications in use. Next determine if there are any priority or latency requirements that are unique to the applications. Applications with latency sensitive traffic may require WAN quality of service, or even dedicated WAN links.

Table 145: Calculate Application Specific Bandwidth

MULTI-NET-002	CALCULATE APPLICATION SPECIFIC BANDWIDTH.
Justification	WAN links must also carry application level replication traffic and application traffic.
Implication	Critical application traffic may require more WAN links or WAN quality of service policies.

Latency Requirements

Latency between protected sites will have an impact on the speed of data replication. Latency is a function of physical distance, network cabling plus equipment, and the speed of light.

Asynchronous DR has no minimum network latency requirements as long as the bandwidth requirements can be met. This makes asynchronous replication a great choice for protecting sites a long distance apart such as between regions.

Near-sync DR also has no fixed requirement for minimum latency between sites, but the bandwidth requirement must be met.

© 2021 Nutanix, Inc. All rights reserved

Metro Availability has two separate requirements for network latency, one between the protected sites, and another between the protected sites and the optional witness. Between protected sites the latency must be less than 5 milliseconds round trip, or 2.5 milliseconds one way. The sites must therefore be located relatively close to each other, but in separate availability zones. Each protected site must also have a connection to the witness that is within 200 milliseconds or less to allow for reliable message exchange during site failures.

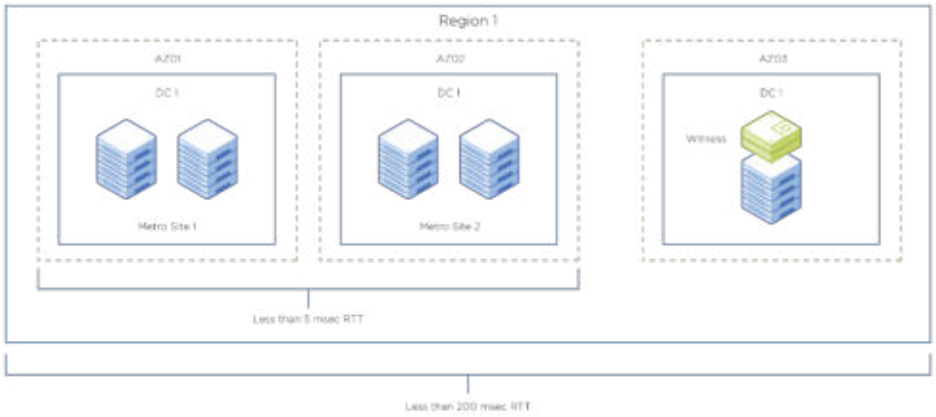


Figure 69: Latency Requirements

Every bit of latency on a link between Metro Availability sites also affects storage write latency for workloads. Keep this latency as low as possible by choosing MA sites that are as close to each other as possible while still providing the necessary physical resilience. In this guide we assume that MA sites exist within the same availability zone, but the witness can be in another AZ, or even in another region.

Table 146: Place Metro and Synchronous Replication Sites within the Same Region

MULTI-NET-003	PLACE METRO AND SYNCHRONOUS REPLICATION SITES WITHIN THE SAME REGION WITHIN 100KM OR LESS THAN 5MSEC RTT
Justification	Latency on metro links is directly added to storage write latency.
Implication	Metro Availability site design must consider physical location carefully.

Table 147: Place Metro Witness within 200msec

MULTI-NET-004	PLACE METRO WITNESS WITHIN 200MSEC
Justification	Sites must receive responses from the witness to determine failure handling.
Implication	Witness can be relatively far from protected sites.

Separating DR Traffic

Nutanix does not recommend isolating DR traffic on separate WAN links because it adds complexity, but if security or business requirements dictate a separated WAN, keep the following in mind.

Consider the scenario where a company has several separate links between protected sites where some are for low latency applications that take high priority. This customer may decide to also purchase an entirely separate physical link between sites that is for higher bandwidth replication traffic that is not latency sensitive and at a lower cost. In essence this is a form of QoS for the application. The customer can ensure that critical application traffic is never stopped by lower priority replication traffic.

It's also possible that the reverse is true for a case like Metro Availability, where certain application and storage replication traffic may take higher priority than other application traffic, and we dedicate an entire WAN link to this storage traffic.

In these cases, the desire is to segment traffic for one purpose over one WAN link between sites, and traffic for another purpose over another link between sites. There are several ways to achieve this goal.

Network Segmentation in the Nutanix CVM

Network Segmentation in the CVM creates another virtual NIC on the CVM and the hypervisor, allowing the separation of user, replication, and DR traffic to different VLANs or physical interfaces. Every CVM and every host needs additional IP addresses in this method, so the configuration for each host becomes slightly more complex.

Network Routing: Manual or Policy Based

Using network routing, the network devices use criteria like source and destination address, and even TCP ports to determine which link is used for routing traffic. This method requires more advanced network configuration but simplifies the CVM and host configuration.

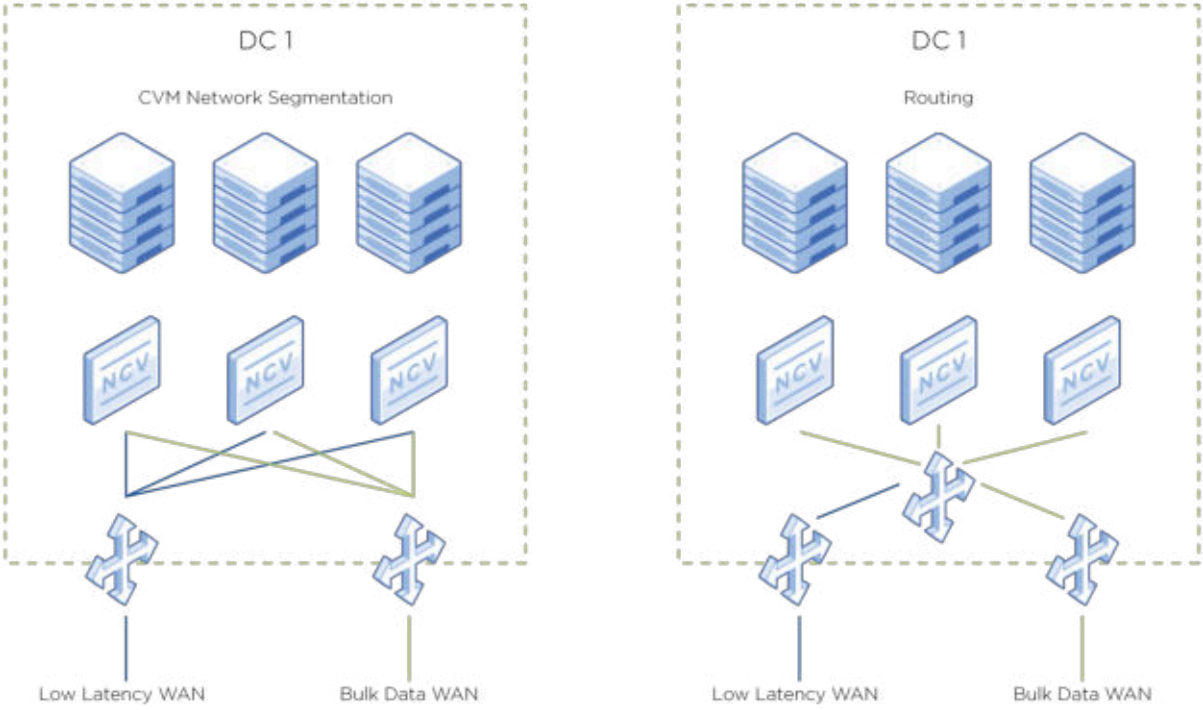


Figure 70: Network Routing

Table 148: Do Not Use Network Segmentation for DR

MULTI-NET-005	DO NOT USE NETWORK SEGMENTATION FOR DR UNLESS REQUIRED. ROUTING SEGMENTATION IS PREFERRED OVER CVM SEGMENTATION
Justification	Network segmentation in the CVM and host adds complexity that must be managed during the system lifecycle.
Implication	Unless specifically required to do otherwise, traffic from applications and DR will share the same network links.

Network Redundancy

Ensure there are redundant paths between protected sites and minimize or eliminate single points of failure. This prevents the failure of a single networking component from stopping replication of storage and application traffic. At each network hop between two protected sites, evaluate whether failure of that component would be gracefully handled, or if would lead to an outage.

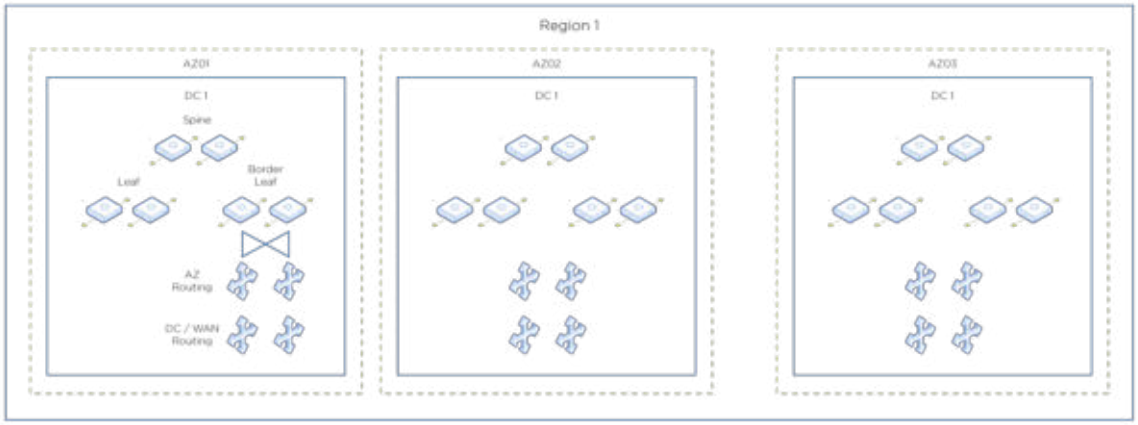


Figure 71: Network Redundancy

For example, in the path between two sites connected over a WAN we have the following components:

- Site 1: Top-of-rack switch (leaf)
- Site 1: Spine switches
- Site 1: Border leaf
- Site 1: DC router
- Site 1: WAN router
- Site 2: WAN router
- Site 2: DC router
- Site 2: Border leaf
- Site 2: Spine switches
- Site 2: Top-of-rack switch (leaf)

© 2021 Nutanix, Inc. All rights reserved

This means we need to ensure redundancy in the following components:

Top-of-rack switch

Ensure that each Nutanix host in every cluster is connected to at least two top-of-rack switches.

Spine switch

Ensure that there are at least two spine switches, and that all leaf switches are connected to all spine switches.

Border leaf

The border leaf is where external services such as routing connect. Ensure that there are at least two border leaf switches, and that they connect to all spines. The border leaf switches should connect to at least two routers, to allow redundant external routing.

DC router

Depending on the size of the deployment, there will be one or more levels of routing between the rack networks and the WAN edge. Each routing tier should provide high availability to allow external routing from the racks to the rest of the DC.

WAN router

The WAN router provides connectivity external to the datacenter. These should be deployed in a redundant fashion, allowing for the failure of any single routing component.

The WAN also presents a special case for WAN Link redundancy. In addition to providing multiple WAN routers, designs should use multiple WAN links, potentially from different network providers.

The WAN provider may be an external entity, or in the case of connection between sites within a campus, they may be entirely within the control of one entity. In these cases where the links are under your control, it is still desirable to use multiple separate physical connections, such as separate fiber links between sites. It's desirable to route these links to the WAN through separate physical paths to protect against ground or aerial utility faults (such as backhoes). It may not always be possible to use different physical paths based

© 2021 Nutanix, Inc. All rights reserved

on datacenter location, design, or infrastructure provider constraints. Where a single point of failure exists, it should be well documented and understood. One example could be a single piece of conduit carrying multiple WAN links inside it between datacenters.

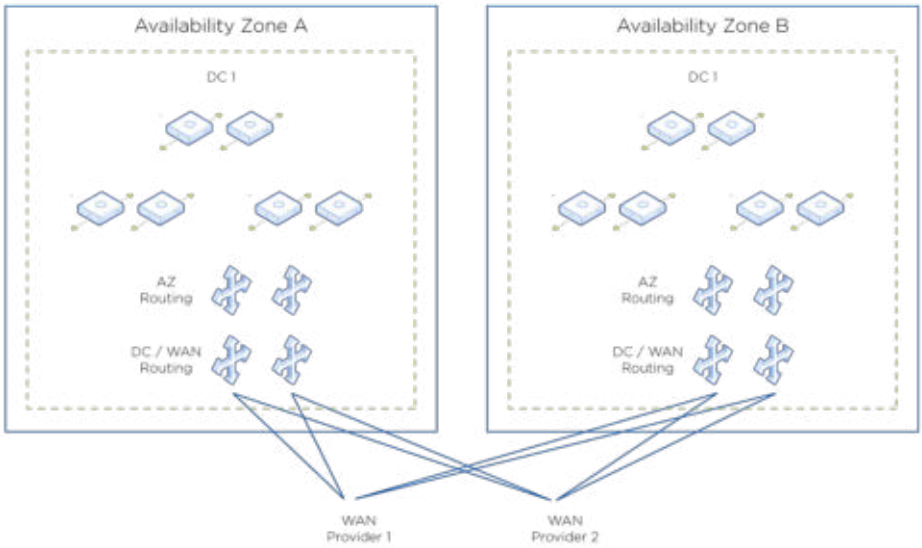


Figure 72: Single Piece of Conduit with Multiple WAN Links

Testing is an important part of the redundancy in any system and networking is no exception. Create and maintain a regular schedule to test the redundancy of each individual component such as routers and switches. Maintenance such as vendor software upgrades and patches as well as regular power maintenance provide great opportunities to verify the resilience during planned single component outages.

Table 149: Ensure Redundancy of Each Network Device

MULTI-NET-006	ENSURE REDUNDANCY OF EACH NETWORK DEVICE. TRACK THE COMPLETE NETWORK PATH BETWEEN PROTECTED SITES
Justification	Failure of a single component must not lead to connectivity failure between protected sites.
Implication	Redundant components must be purchased and configured for high availability and failover.

Table 150: Create and Follow a Redundancy Test Plan

MULTI-NET-007	CREATE AND FOLLOW A REDUNDANCY TEST PLAN. CAPTURE NETWORK COMPONENTS IN PLAN
Justification	High availability configurations are often complex and must be tested to ensure they work as expected.
Implication	Regularly scheduled maintenance must include additional protocols to test network components during this time.

Witness Path Requirements

Using Metro Availability with a witness creates resiliency requirements between the protected sites and the witness site.

First, the witness should be at a separate site from either of the protected sites. This allows the witness to monitor the availability of both sites independently and means that the witness should survive the failure of one of the protected sites. The witness provides no extra benefit if it goes down during a site failure.

In addition, the network path between the protected sites should be a separate path from the path to reach the witness. That means each protected site should have at least two routing paths, one to reach the witness, and another to reach the other protected site.

Finally, the connection between the protected site and the witness site should have no more than 200msec of latency. This helps ensure timely event handling during a failure scenario.

© 2021 Nutanix, Inc. All rights reserved

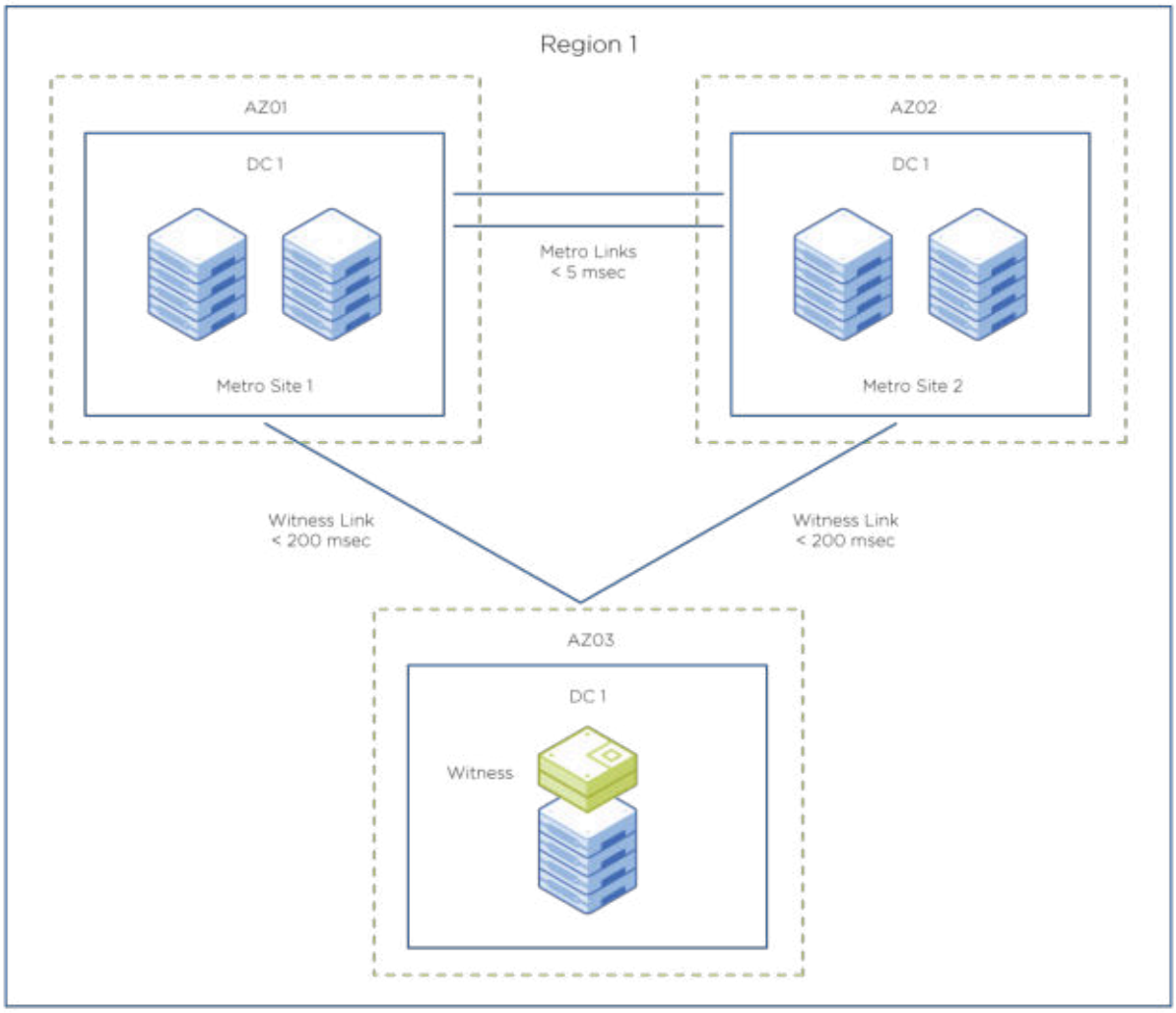


Figure 73: Metro Witness Link Requirements

Table 151: Ensure that the Link Between Protected Sites and a Witness does not Use the Link Between Sites

MULTI-NET-008	ENSURE THAT THE LINK BETWEEN PROTECTED SITES AND A WITNESS DOES NOT USE THE LINK BETWEEN SITES.

Justification	The path to the witness should be a separate, protected link. The witness should be reachable during failures of the link between the protected sites.
Implication	Additional WAN connectivity may be required to maintain a separate external connection to the witness.

Workload Networking

In addition to transporting critical storage traffic, the network also provides access to running applications. During site failures it's important to consider how the application or workload will provide constant access for clients, even when the workload may be running in a secondary location.

IP Address Assignment

There are two approaches for managing workload connectivity during site failures.In the first approach, workloads keep their existing IP addresses and the network provides connectivity in the backup site even when using previous IP addresses. In the second approach, workloads fail over to the second site and receive new IP addresses belonging to the backup site.

The network tools used are outside the scope of this document but consider the following when moving workloads between sites using these tools. The workload and the network capabilities will determine the chosen method.

Case 1: Workloads keep the same IP address

Preserving the IP addresses of workloads during failover is helpful for several reasons. No change is required inside the workloads, which may sometimes not even support IP address changes. This also means any scripts or infrastructure that rely on hardcoded IP addresses will continue to function. The network infrastructure seamlessly handles the transition by migrating the network to the new site, or by maintaining a stretched network at all times. The responsibility of failover reachability is shifted away from the application and to the network.

The following approaches are frequently used to preserve IP addresses:

- Stretched layer 2 networks
- NAT

- Layer 3 route advertisement updates

Case 2: Workloads receive a new IP address

Provisioning workloads with new IP addresses allows more network flexibility, for instances where moving a network between sites is not feasible. When the workload easily supports IP address changes, or if the network does not support network mobility, changing workload addresses may be the best option. Using a tool to automate these address changes ensures a smooth recovery.

The following tools are often used to provision new addresses at the DR site.

- DHCP / IPAM
- Site Recovery Manager
- Nutanix Runbook Automation
- Manual or scripted steps
- Vendor tools such as Zerto

Table 152: For Each Workload Decide Whether to Maintain or Change

MULTI-NET-009	FOR EACH WORKLOAD, DECIDE WHETHER TO MAINTAIN OR CHANGE IP ADDRESSES DURING FAILOVER.
Justification	
Implication	

Service Advertisement

Service availability is the advertisement of the workload and its availability to external clients at the DR site. There are a number of network and application technologies that can provide service advertisement, and this will be specific to your network and workload. For each application, consider how clients will reach this application at the DR site. This may involve application multi-site awareness, DNS updates, routing updates, load balancers, or some combination of these.

6. Incorporating Optional Nutanix Products and Services

Nutanix offers a number of optional products and services that you may choose to incorporate when implementing private, hybrid, or multicloud solutions on the Nutanix platform. These options can help you quickly satisfy unique organizational requirements.

The subsections that follow introduce the following solutions and provide operational guidelines consistent with earlier sections of this guide:

- Prism Operations: Prism Pro and Prism Ultimate add to the base capabilities of Nutanix Prism to provide monitoring, analytics, and automation capabilities from a single management interface.
- Nutanix Clusters: Clusters extends the Nutanix stack to the public cloud, creating a single management domain that spans private and public clouds and removes the friction from multicloud operations.
- Flow: Microsegmentation and other capabilities offered by Flow further enhance the security of your Nutanix operations.
- Files: Files enables a Nutanix cluster to incorporate full-featured file services capabilities, eliminating the need for standalone NAS appliances or file servers.
- Objects: Objects enables a Nutanix cluster to provide S3-compatible, software-defined, scale-out object stores for a variety of use cases, eliminating the need for standalone object storage while supporting applications that rely on object stores with an on-premises alternative.
- Mine: Nutanix Mine is an integrated backup solution, combining the benefits of the Nutanix HCI architecture with the capabilities of proven backup vendors.
- Calm: Calm provides application-level orchestration and lifecycle management, simplifying application management and enabling self-service.

© 2021 Nutanix, Inc. All rights reserved

- Karbon: Karbon integrates certified Kubernetes with the Nutanix operating environment, enabling Nutanix clusters to support both VM- and container-based applications while simplifying Kubernetes deployment and management.
- Era: Era automates and simplifies database administration, provisioning, and lifecycle management, facilitating database as a service (DBaaS).

Each subsection includes: Key Design Objectives, Architecture Overview, and Detailed Technical Design Considerations.

Prism Operations

Key Design Objectives

Table 153: Prism Ops Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	NEW / UPDATE / REASONING
Enable proactive operations management	Prism Pro and Prism Ultimate licenses provide: - Operational automation in response to common triggers - Monitoring for non-Nutanix VMware environments - Application discovery and application-level monitoring	New

Architecture Overview

The operations tier in Nutanix Prism (provided by Prism Pro & Ultimate) focus on the entire IT Ops triangle—monitoring, analytics, and automation. Prism offers a single pane of glass that combines monitoring and automation to isolate pain points and ease the deployment of operational automation. This allows you to harness the full power of HCI. The solution aims to remove silos

between different monitoring and automation tools and avoid the noise created by dynamic virtual infrastructure environments.

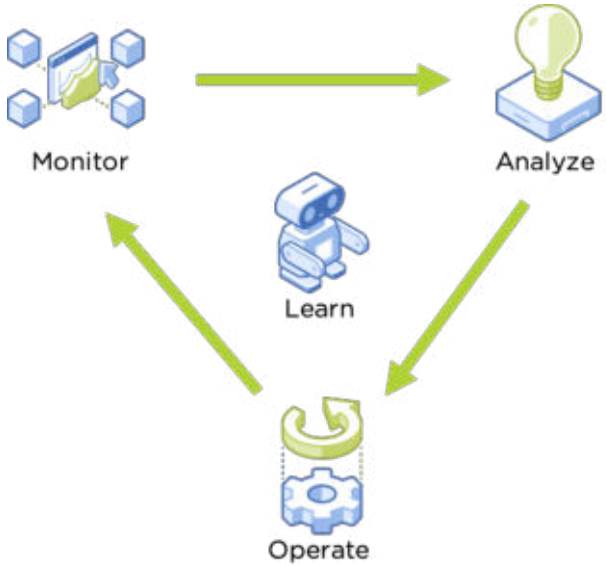


Figure 74: IT Ops Triangle

The goal is to provide broad observability and actionable signals powered by machine learning and automation for a seamless operations experience.

Prism Operations (Prism Ops) is powered by X-FIT, a distributed time series analysis and forecasting system that uses an ensemble of models. X-FIT is an enabler for building a management fabric; it can autonomously optimize datacenter performance and intelligently manage application resource demands using operational data.

Virtualization

- Prism Pro and Ultimate are hypervisor agnostic, working with both AHV and ESXi.
- With the release of Prism Central 2020.9, it is possible to monitor non-AOS vCenter VMs with Prism Pro. Similar capabilities available for an AOS environment, such as runway forecasting and playbook execution, are extended to VMware when it is monitored with Prism.

Management

All Prism Operations features are accessed through Prism Central, which removes the need for separate management tools, providing one tool for monitoring, operational analysis, and automation. With Prism Ultimate, this visibility and functionality has been extended to the application layer.

Automation

- X-Play is a simple-to-use automation engine geared toward the average IT admin managing Nutanix and non-Nutanix infrastructure (as of PC 2020.9). It is based on various triggers, including actionable alerts generated by machine learning insights. As X-Play runs through a Playbook, it initiates a sequence of actions that can be tracked through a corresponding Play.
- Six triggers are currently available: alert-based, manual, time-based, alerts matching multiple criteria, event-based, and webhook triggers. There are 27 out-of-box actions provided that can be customized according to your needs and integrated with third-party tools. Playbooks can also be imported and exported across different Prism Central instances.

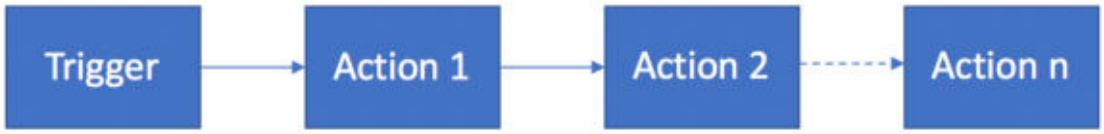


Figure 75: With Prism X-Play, a single trigger can precipitate multiple actions

Prism Licenses

Prism has three licensing tiers:

- Starter: Included free with AOS and deployed automatically with Prism Central. Features include:
  - Monitoring and troubleshooting
  - RBAC
  - LCM upgrades
  - Comprehensive search

- Pro: Separate add-on license. Enabled via Prism Central. Features include:
  - Customizable dashboards and reporting
  - Capacity forecast and runway planning
  - VM inefficiency detection and right-sizing
  - Anomaly detection and advanced troubleshooting
  - Monitoring non-AOS infrastructure
  - Smart operations automation
- Ultimate. Separate add-on license. Enabled via Prism Central. Features include:
  - Application discovery
  - Application monitoring – SQL Server content pack
  - Cost-metering, budgeting, and chargeback for resources

Prism Ultimate

Administrators may have zero visibility into the applications running inside VMs deployed on Nutanix infrastructure. Troubleshooting application issues becomes onerous without this application data and often requires separate software for application monitoring.

Prism Ultimate collects application metrics using third-party/Nutanix collectors (agentless), providing a single pane of glass to view both application and infrastructure data. With collected performance metrics, Prism Central can now provide a holistic view of IT operations that allows admins to easily monitor and diagnose problems caused by app performance issues without finger pointing.

High-Level Design Considerations

Currently, Prism's operations features are tied to a single instance of Prism Central. This means all nodes registered to a Prism Central instance must be licensed for Prism Pro & Ultimate in order for those nodes to be license compliant.

© 2021 Nutanix, Inc. All rights reserved

- To enable Prism Ultimate features, you must increase Prism Central resources to 14 vCPUs and 30 GiB memory on a small Prism Central instance or 18 vCPUs and 48 GiB memory on a large Prism Central instance.
- Prism Ultimate is available beginning with release PC 2020.8. Non-AOS VM monitoring is available beginning with PC 2020.9.

Operational Automation

Table 154: Leverage X-Play Playbooks

OPS-001	LEVERAGE X-PLAY FOR POWERFUL LOW-CODE/NO-CODE OPERATIONAL AUTOMATION
Justification	Playbooks require the presence of Prism Pro.
Implication	Requires an additional license that can increase cost.

The webhook trigger gives a third-party tool the ability to execute actions within Prism Central. Conditional execution based on a string that is read and parsed allows for if-this-then-that (IFTTT) workflows, giving complete control to the admin for building powerful automations. For example, you can create a service ticket in your preferred tool (ServiceNow, etc.) based on alerts or events and trigger further actions in Prism Central based on approval or denial.

© 2021 Nutanix, Inc. All rights reserved

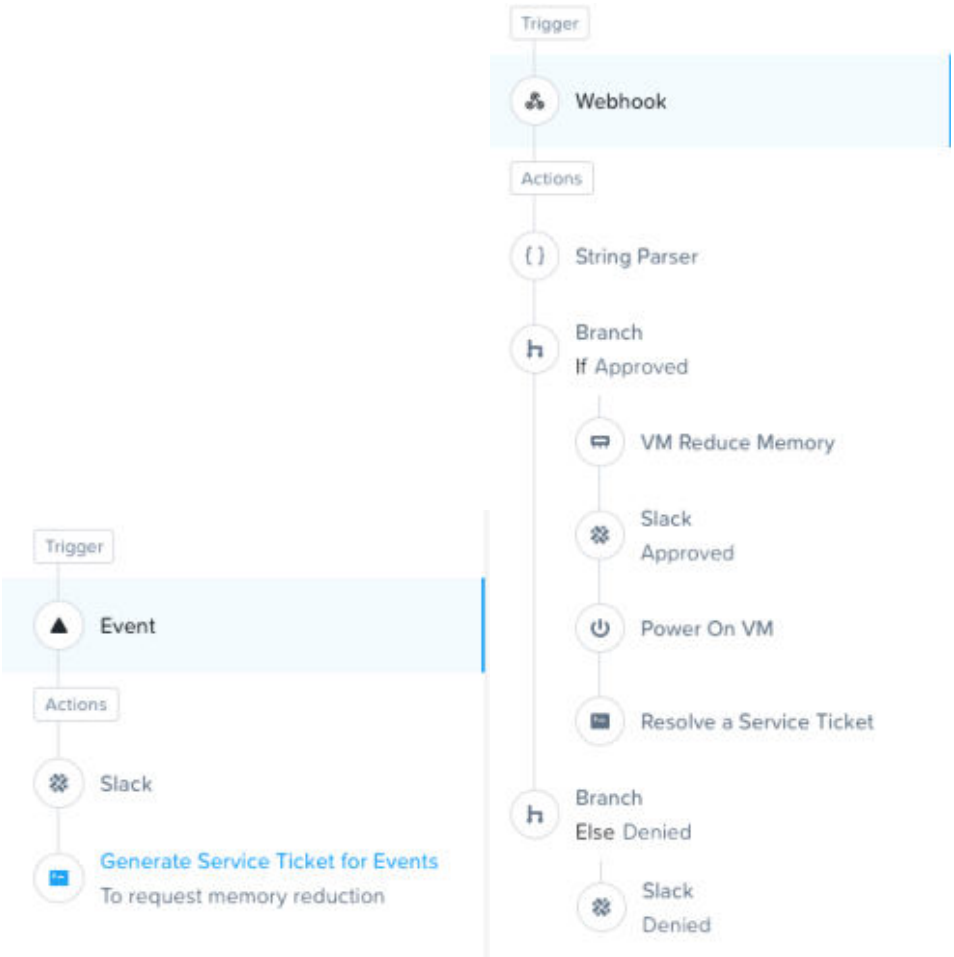


Figure 76: Example Workflows Based on Various Triggers

Non-AOS VM Monitoring

Table 155: Monitor Non-AOS VMs

OPS-002	MONITOR NON-AOS VMs
Justification	VMware virtualized environments not running under AOS can be monitored with Prism Pro.
Implication	Requires an additional license that may increase cost.

A containerized Nutanix collector is deployed on the Prism Central VM and gathers data for each configured VMware cluster, using the X-Stream interface to store the data. X-FIT generates the same intelligence as for a Nutanix cluster, and X-Play can be used to connect to automation.

Figure 77: Monitoring non-AOS VMs with Prism Ultimate

Application Discovery

Table 156: Discover Applications with Prism Ultimate

OPS-003	DISCOVER APPLICATIONS RUNNING ON YOUR INFRASTRUCTURE.
Justification	You can discover and export applications running on your Nutanix cluster with Prism Ultimate.
Implication	Requires an additional license that may increase cost.

Prism Ultimate's agentless discovery allows you to identify application-to-VM dependencies using an API for CMDB integration. (A cloud connection is required.) Internet Protocol Flow Information Export (IPFIX) data is collected from VMs and sent to the Nutanix cloud which analyzes and syncs the findings back to Prism.

App Monitoring

Table 157: Monitor Microsoft SQL Server

OPS-004	MONITOR SQL SERVER

Justification	Prism Ultimate can provide greater insights into SQL Server operation, enabling you to understand how infrastructure impacts applications and to isolate problems between the two layers.
Implication	Requires an additional license that may increase cost.

Prism Ultimate provides visibility into databases, queries, and SQL-specific metrics plus behavior learning and anomaly detection. The containerized Nutanix collector is deployed on the Prism Central VM and gathers data for each configured SQL instance, using the X-Stream interface. X-FIT then analyzes the data to generate insights and alerts for troubleshooting and automation.

SQL Server is the first app monitoring content pack to be released, providing deeper insights into application performance metrics. It currently supports 100 SQL instances or 94,000 metrics.

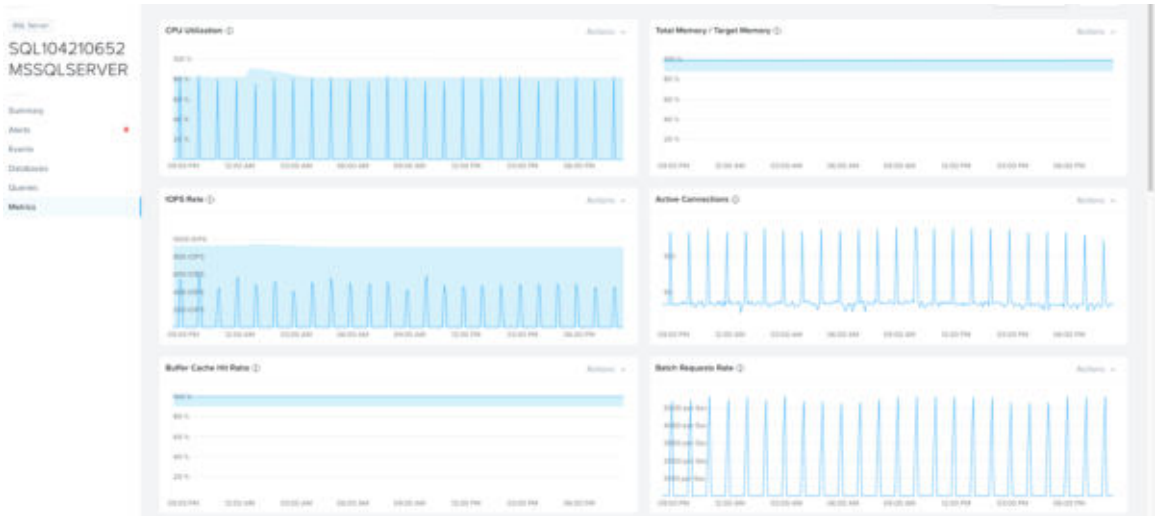


Figure 78: Monitoring SQL Server with Prism Ultimate

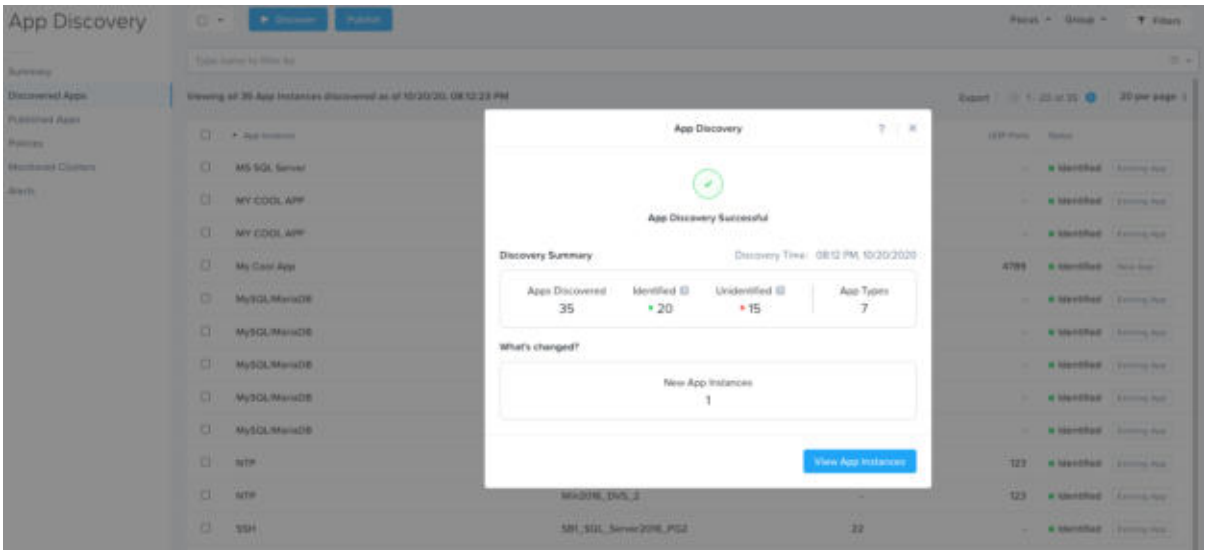


Figure 79: App Discovery

Nutanix Clusters for Hybrid and Multicloud Operations

Key Design Objectives

Table 158: Nutanix Clusters Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	NEW / UPDATE / REASONING
Extend Nutanix operations to encompass public cloud	Utilizing Nutanix Clusters extends the Nutanix operational domain to encompass AWS in addition to on-premises locations	New

Architecture Overview

Nutanix enables customers running Nutanix software on-premises to deploy the same software stack on cloud computing providers such as Amazon Web Services (AWS), resulting in a more uniform hybrid cloud environment. Because Nutanix Clusters on AWS (NCA) runs Nutanix AOS and AHV —with the same

CLI, UI, and APIs—existing IT processes and third-party integrations that work on-premises operate in the cloud without modification.

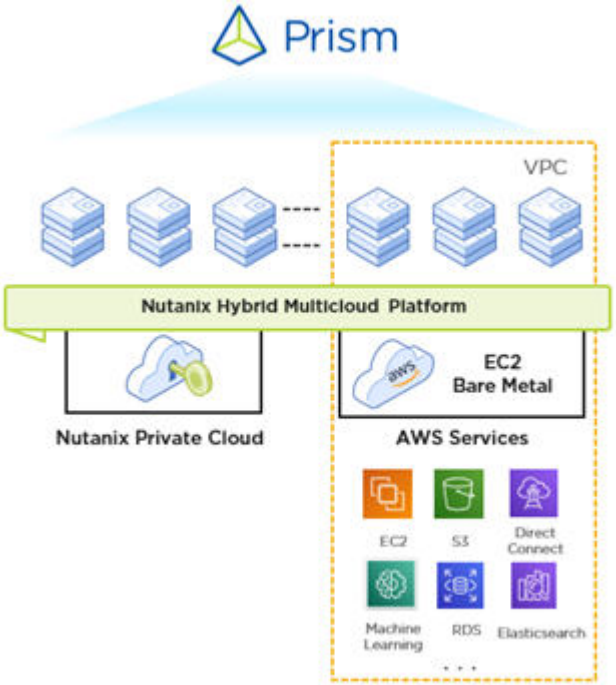


Figure 80: Nutanix Clusters on AWS

NCA implements the complete Nutanix hyperconverged infrastructure (HCI) stack using Amazon Elastic Compute Cloud (EC2) bare-metal instances. Each bare-metal instance runs the same Controller VM (CVM) and the Nutanix AHV hypervisor as used on-premises, using the AWS elastic network interface (ENI) to connect to networking. AHV user VMs do not require any additional configuration to access AWS services or other EC2 instances.

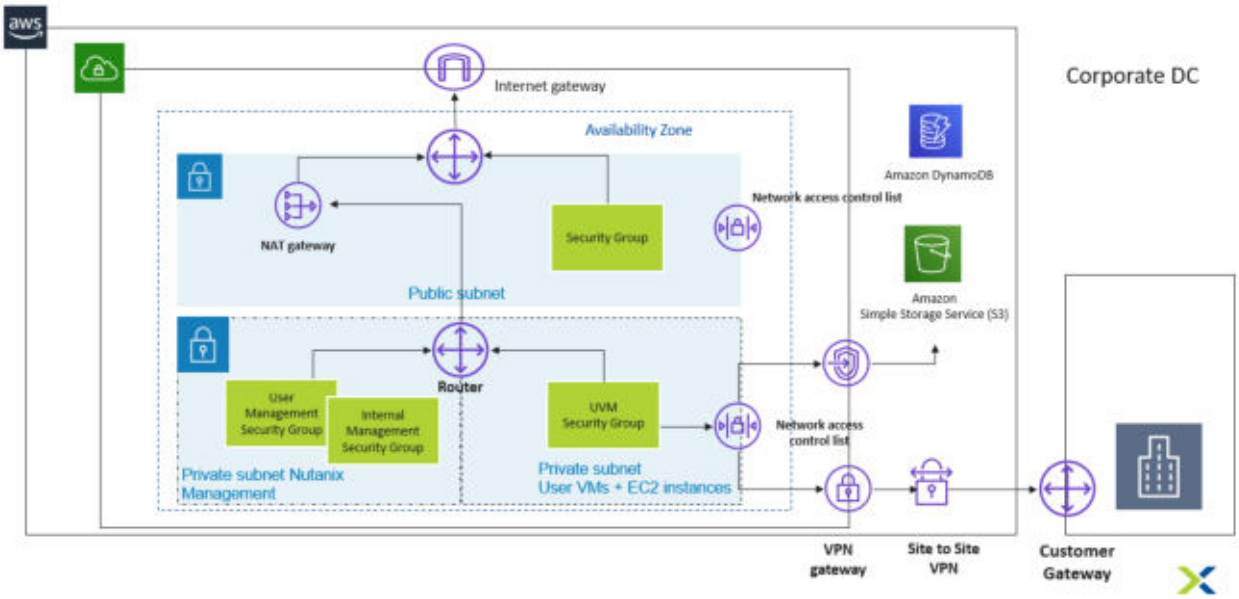


Figure 81: Nutanix Clusters on AWS Architecture

In most ways, Nutanix Clusters on AWS is similar to running Nutanix AOS on any hardware OEM. Almost everything you know about AOS is the same. This reduces the need for different systems administrators, tooling, and scripting between private and public cloud. The following description focuses on the differences related to deployment in AWS.

Networking

AHV runs an embedded distributed network controller that efficiently integrates user VM networking with AWS networking. AHV assigns all user VM IPs to the bare-metal host where VMs are running. Instead of creating an overlay network, the AHV embedded network controller simply provides the networking information for VMs running on NCA, even as VMs move between AHV hosts. Because NCA IP address management is integrated with AWS Virtual Private Cloud (VPC), AWS allocates all user VM IPs from the AWS subnets in your existing VPCs.

NCA utilizes AWS Elastic Network Interfaces (ENI) to provide IP addresses for virtual machines that get created on EC2 bare-metal nodes using AHV's built-in IPAM. One EC2 bare-metal host can have a maximum of 15 ENIs; NCA uses the first ENI for host management (CVMs and AHV). The remaining 14 ENIs can be

used as VM IPs. An ENI on the EC2 instance will only be consumed if there is a VM present on the subnet.

Networking limits and restrictions

The following limits and restrictions apply to NCA deployments:

- One subnet per ENI.
- An ENI can have up to 50 IP addresses, one of which must be the primary IP.
- The primary IP of an ENI cannot be changed or migrated to another ENI, leaving 49 IP addresses per ENI.
- The private management subnet cannot be used for deploying VMs.
- The user VM subnets cannot be shared between clusters.
- The user VMs should use a separate subnet from other AWS EC2 instances.

Currently, the maximum number of VMs supported is:

14 ENI * 49 IPs = 686 VMs per node

© 2021 Nutanix, Inc. All rights reserved

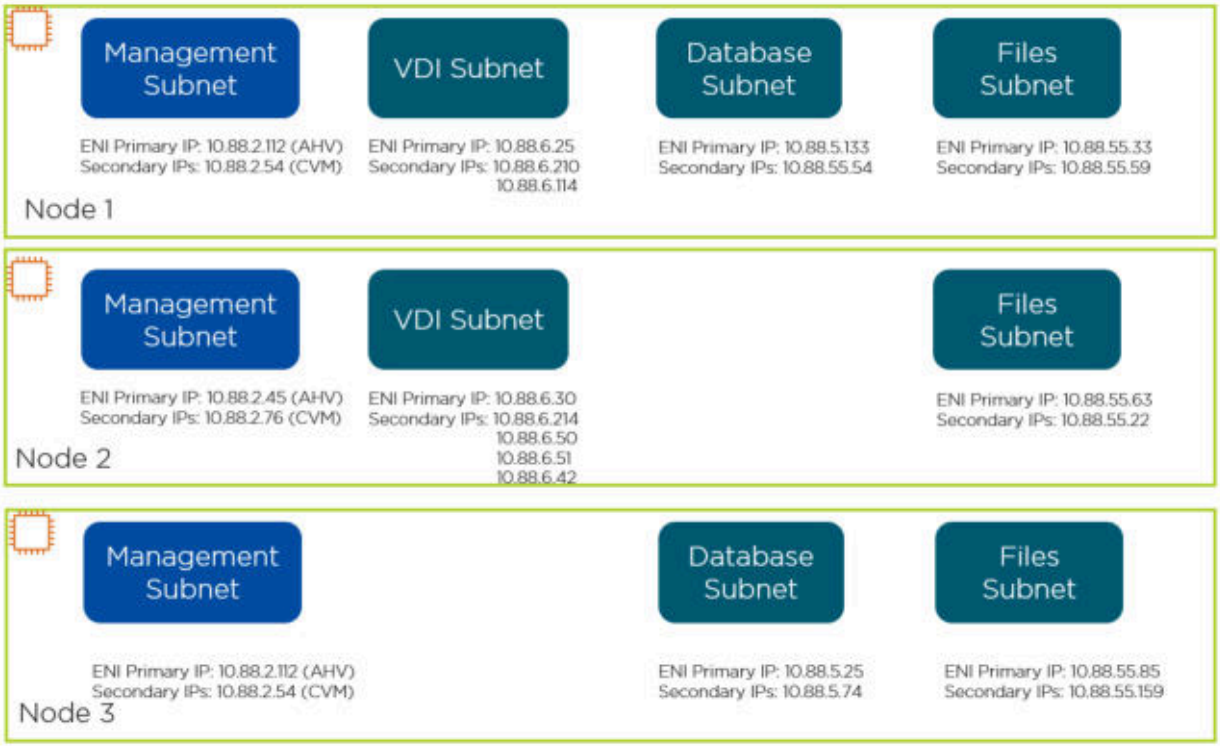


Figure 82: Example IP Address Assignment for an NCA Deployment

The figure above illustrates that nothing is running on the third node of the cluster on the VDI subnet. We can tell this because no ENI/IPs are being used on that node. Likewise, we only have two database VMs deployed on the cluster on nodes 1 and 3.

IP Mobility

Nutanix does not currently provide a stretched layer-2 network connection to AWS. The Nutanix Leap Disaster Recovery (DR) service preserves IP and MAC addresses and can be paired with third-party software to stretch the network. Leap can automate assigning new IPs to your VMs, and it uses recovery plans that provide network mappings for remote subnets.

Compute

NCA only supports homogeneous clusters today, so you must use the same EC2 bare-metal instance type for all nodes within a cluster. However, you can

have multiple clusters with different instance types that are optimally suited to the workload(s) in each cluster.

Hardware Choice

The instance types available will depend on the AWS region you select. NCA supports four possible EC2 bare-metal instances:

Table 159: Supported EC2 Bare-Metal Instances

INSTANCE	vCPU	MEMORY (GiB)	NETWORK PERFORMANCE	SSD STORAGE (GB)
z1d.metal	48	384	25 Gigabit	2 * 900 NVMe
m5d.metal	96	384	25 Gigabit	4 * 900 NVMe
i3.metal	72	512	25 Gigabit	8 x 1,900 NVMe
i3en.metal	96	768	100 Gigabit	8 x 7,500 NVMe

The i3.metal instance is a great choice for virtual desktop deployments, offering the best price/performance. The i3en.metal instance is good for database workloads and can provide lower TCO when high storage capacity is needed. The z1d.metal and m5d.metal instances are well positioned for general compute and dev/test workloads that are temporary in nature.

Minimum cluster size is 3 nodes, up to a max of 16 nodes. Each instance type consumes a certain amount of vCPU that goes against your configured limits in AWS. You need to ensure that you have sufficient headroom so that you are not prevented from deploying additional nodes in a cluster. If you are deploying a cluster for N+1 availability, your vCPU limits should account for both the additional node and any future growth.

Business Continuity Layer

Protection against site failures

- A Nutanix AWS cluster with three or more nodes is deployed with a Replication Factor (RF) of 2 by default, with nodes explicitly distributed across racks to protect against failure of a single device, node, or rack. You can also configure RF=3 to provide higher levels of redundancy.

- Similar to Nutanix on-premises operation, a single Nutanix cluster on AWS cannot protect against infrastructure failures that affect more than one rack or an entire AWS availability zone. To learn more about Rack Awareness see KB 9723 .

To protect data against large-scale infrastructure failures, Nutanix recommends you regularly back up and/or replicate data to another Nutanix cluster running on premises or in a different AWS Availability Zone (AZ). Consult Data Protection and Recovery with Prism Element for more information. You can also leverage Nutanix Xi Cloud services such as Xi Leap for DRaaS. See Nutanix Clusters on AWS for more documentation.

Protection against disruptive actions

- A simple reboot (intentional or not) of an EC2 bare metal instance only temporarily affects a Nutanix cluster running on AWS. The node automatically rejoins the cluster.
- Stopping or terminating an EC2 bare metal instance that is part of a Nutanix cluster on AWS effectively removes that node from the cluster. With the default RF=2 configuration, removing a single node does not affect availability. However, stopping or terminating more than one EC2 instance can result in data loss.
- Shutting down or powering off an AHV host running on AWS from the AHV shell results in the permanent removal of that node from the cluster. The cluster will add a node back when it self-heals. The default RF=2 configuration can tolerate unavailability of one node. Shutting down or powering off more than one AHV host from the AHV shell can result in data loss.

Security

Securing traffic to and from your AWS cluster can be accomplished using a combination of third-party firewalls, Nutanix built-in security groups, and Nutanix Flow (see the section Nutanix Flow for Microsegmentation for details).

The Nutanix Clusters console creates the following security groups at the time a cluster launches in AWS:

- Internal Management

- User Management
- UVM

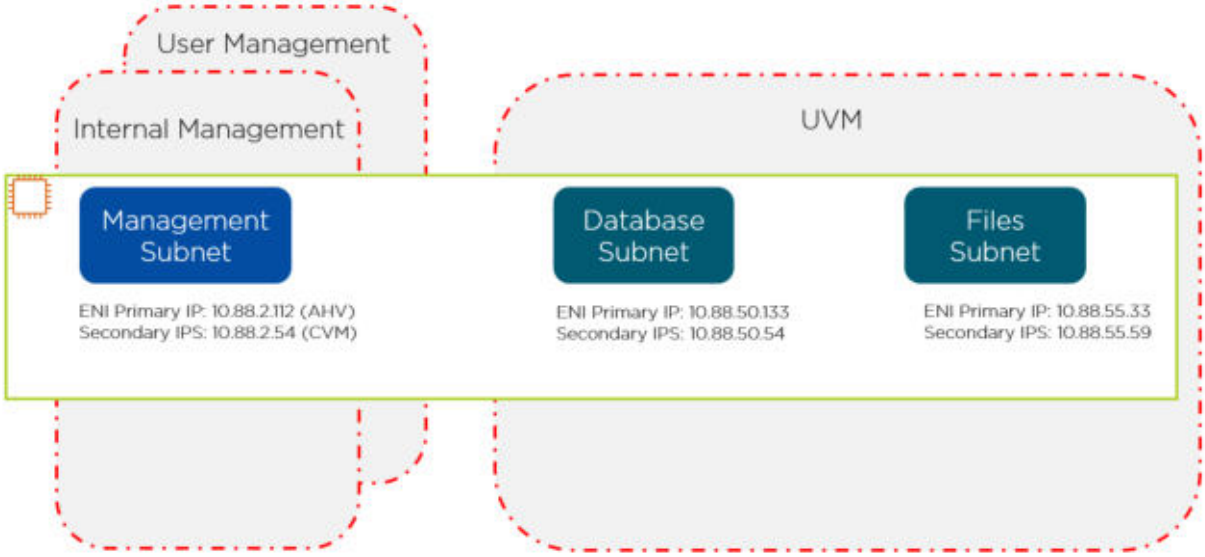


Figure 83: Security groups used by NCA

These security groups have the recommended default rules. Nutanix recommends that you do not modify internal management.

All management network ENIs created have the following security groups attached, even after the cluster is running:

- internal_management: Allows all communication between AHV and CVM within a cluster
- user_management: Allows specific ports for UVM to CVM communication

ENIs created for user VM subnets have the following security group attached:

- UVM: Allows communication between user VMs

Detailed information on allowed port groups can be found in the Nutanix Clusters on AWS Deployment and User Guide. Nutanix Flow works both in the cloud and on-prem and can be used to segment operations and to block east/west traffic.

High-Level Design Considerations

Networking

Internet Access

Portal Access: Determine how you will enable outbound access from your NCA cluster to the Cluster Portal. You can route all Internet traffic through your corporate firewall or set up an Internet Gateway in AWS. If your company is blocking outbound traffic, you also need a firewall that can block domain traffic versus subnet ranges since some of the Cluster Portal services are dynamic.

On-Premises Connectivity: There are three native ways to connect to on-premises operations from AWS:

- VPC VPN: Works well for connecting to a single site.
- Transit Gateway: Preferred when you plan to connect multiple sites to your NCA cluster.
- Direct Connect: Typically, more expensive but provides a very stable connection at your desired speed.

Table 160: Separate Subnets

AWS-01	CREATE SEPARATE SUBNETS FOR USER VMS
Justification	AWS clusters cannot deploy user VMs on the same subnet as the CVMs and AHV hosts. New subnets are also needed if you are using Leap to failover VMs into AWS. You should also use different subnets for native AWS EC2 instances.
Implication	Unless using third-party software to create a stretched layer 2 network, you will have to change your VMs' IP addresses when they move to AWS.

Business Continuity

Determine your availability needs. If you are running dev/test, you probably don't need to replicate data to another region or to an on-premises datacenter.

If you are running production workloads, you need to replicate a copy of your data on-prem, to another AWS region, or use an AHV backup product.

Table 161: Create Availability Zones

AWS-02	USE AWS DIRECT CONNECT OR CREATE A 2ND AWS CLUSTER IN A DIFFERENT AVAILABILITY ZONE.
Justification	To achieve an RPO of zero, you will need to ensure you meet the latency requirements for synchronous replication. If your latency is over 5ms round trip from your primary datacenter, you will have to create a secondary AWS cluster in a different AZ.
Implication	A disaster that affects an entire geography or a system wide event might affect more than one AZ in a given region.

Table 162: Use Replication

AWS-03	PROTECT YOUR AWS CLUSTER BY REPLICATING BACK TO ON-PREM.
Justification	You need to protect yourself from an outage in AWS. You can back up your AWS cluster using AHV-based backup software, but the region where you place the backups has to be able to be restored in a timely fashion. Also, the region you restore service to has to be close enough to meet any on-prem SLAs you might have.
Justification	You need to protect yourself from an outage in AWS. You can back up your AWS cluster using AHV-based backup software, but the region where you place the backups has to be able to be restored in a timely fashion. Also, the region you restore service to has to be close enough to meet any on-prem SLAs you might have.

© 2021 Nutanix, Inc. All rights reserved

Table 163: Don't Overlap Subnets

AWS-04	PLAN YOUR AWS VIRTUAL PRIVATE CLOUD NOT TO OVERLAP WITH YOUR ON-PREM SUBNETS.
Justification	Nutanix Clusters does not provide stretched layer 2 networking natively. If you want to failover or migrate your on-prem applications and also keep a subnet active on-prem, you will need a subnet that does not overlap. If your goal is only to achieve a complete failover into AWS, you can create the same subnet in AWS, since your on-premises environment won't be active after a failover occurs.
Implication	Failing over an entire subnet without re-IPing requires you to shut down those subnets on-prem if you intend to keep the connection alive.

Security

AWS Accounts: Determine if your level of risk is acceptable with the privileges delegated to the Cluster Portal. While the Cluster Portal is built with the utmost attention to security, some customers may still want a greater level of separation. To achieve this, customers can use a separate AWS account with VPC pairing to their current AWS account.

Table 164: Allow Outbound Internet Access

AWS-05	ALLOW OUTBOUND INTERNET ACCESS THROUGH AN EXISTING CONNECTION TO THE INTERNET TO ALLOW THE CLUSTER PORTAL TO MANAGE YOUR CLUSTER.
Justification	You should limit the number of attack vectors into your environment. In a hybrid cloud environment, you most likely already have an existing outbound connection. If you only have a public cloud environment, you will have to setup NAT and an Internet gateway.

© 2021 Nutanix, Inc. All rights reserved

Implication	You will have to pay for egress networking costs to your on-prem environment(s).

VPC Limits

Table 165: Plan VPC Limits for Growth and Availability

AWS-06	You will have to pay for egress networking costs to your on-prem environment(s).
Justification	Your vCPU limits are set at the regional level and cover many instance types. You should ensure that you have enough headroom for expansion and to accommodate a node failure. If your cluster is using RF=3, then you must ensure that you have enough spare vCPU capacity to accommodate the vCPU needs for two failed nodes.
Implication	Increasing limits affects a variety of EC2 services. You should monitor for any suspicious activity since the limit will be higher.

Nutanix Flow for Microsegmentation

Key Design Objectives

Table 166: Flow Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	NEW / UPDATE / REASONING
Implement microsegmentation for greater security	Provide stateful firewalls to segment applications running in VMs on Nutanix clusters	New

Architecture Overview

Flow provides microsegmentation functionality using a distributed stateful firewall that enables granular network monitoring and enforcement between VMs running on the AHV platform as well as external entities they communicate with. Flow provides security rules, which enforce what kind of traffic is allowed between VMs. Flow configuration is done via Prism Central (PC) by defining security rules based on categories. Doing the configuration from PC allows centralized management and distributed enforcement, where the enforcement is done on multiple Nutanix clusters and multiple Nutanix AHV nodes. Flow configures OpenFlow rules in each AHV host in order to implement the security rules.

Nutanix Flow uses categories and security rules that are configurable based on defined categories.

Categories

Categories are used to define groups of entities. The category schema consists of keys and values:

- Example Key: Department
- Example Value: Engineering

A {Key, Value} pair together forms a unique tuple in the PC management domain and this tuple is used to tag VMs.

For example, a VM providing production database services may have the following categories assigned to it:

- Key: Value
- AppTier: Database
- AppType: MySQL
- Environment: Production

These categories can then be utilized by Flow to determine what security rules to apply.

© 2021 Nutanix, Inc. All rights reserved

Security Rules

Security rules specify the Flow policies which determine how VMs can talk with each other. These rules govern firewalls for applications and many other use cases, including isolating environments or quarantining VMs that are suspected to be infected. Based on the most common use cases, Flow provides four types of policies as shown in Table 164.

Table 167: Flow Policy Types

POLICY TYPE	USE CASE	EXAMPLE DEPLOYMENT
Quarantine	Quarantine Apps/VMs	Fault isolation and control
Isolation	Environment separation	Geographical isolations
App-Type	Application segmentation	App policy or inter-tier
AD-Type	AD/VDI segmentation	ID firewalls

Quarantine Policy

A Quarantine policy denies all traffic to/from specified VM(s)/categories.

Example: VMs A, B, and C are infected with a virus; isolate them to stop the virus from further infecting the network

Figure 84: Effects of a Quarantine Policy

Isolation Policy

An Isolation policy denies traffic between two categories, while allowing traffic within a category.

Example: Separate tenant A from tenant B. Each environment is isolated and runs in parallel without affecting normal network communication.

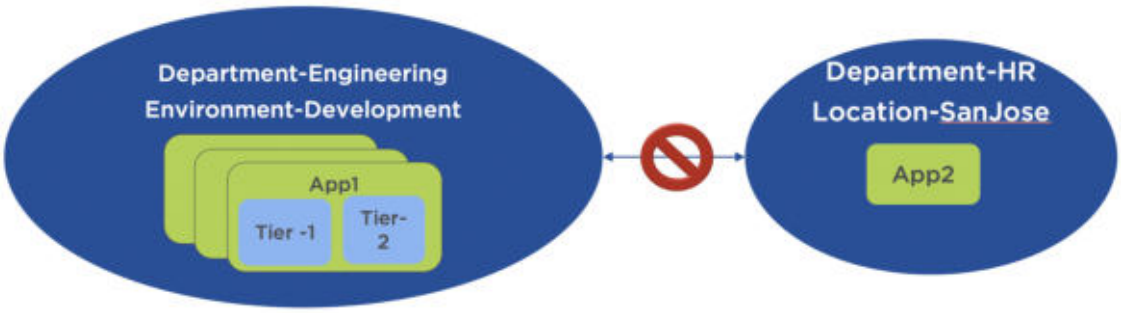


Figure 85: Effects of an Isolation Policy

App-Type Policy

An App-Type policy is used for configuring firewalls for applications running inside VMs. This rule allows you to define what transport (TCP/UDP), Port, and source/destination is allowed/denied.

- [Allow|Deny] Transport: Port(s) [To|From]

Example: Allow transport on TCP port 8080 from Category:Tier:Web to Category:Tier:App

The policy schema consists of Inbounds, Target Group, and Outbounds. Inbounds identify who can talk with the Target Group. Outbounds identify who the Target Group can talk with. Target Group identifies the VMs or applications that are being protected.

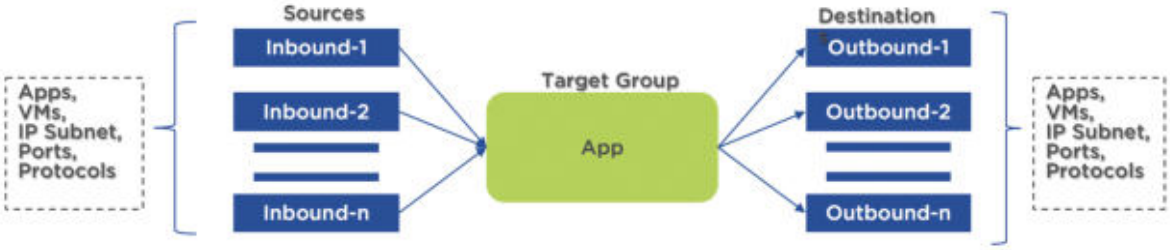


Figure 86: Illustration of an App-Type Policy

AD-Type Policy

An AD-Type policy is used for configuring firewalls for AD/LDAP users. This rule allows you to define what transport (TCP/UDP), Port, and source/destination is allowed/denied.

- [Allow|Deny] Transport: Port(s) [To|From]

Example: Allow subnets (example: clients) to talk with AD groups and allow AD groups to access payroll data

The policy schema consists of Inbounds, AD Groups, and Outbounds. Inbounds identify who can talk with the AD Groups. Outbounds identify who the AD Groups can talk with. AD Groups identify the users (based on VMs that the users have logged into) that are being protected.

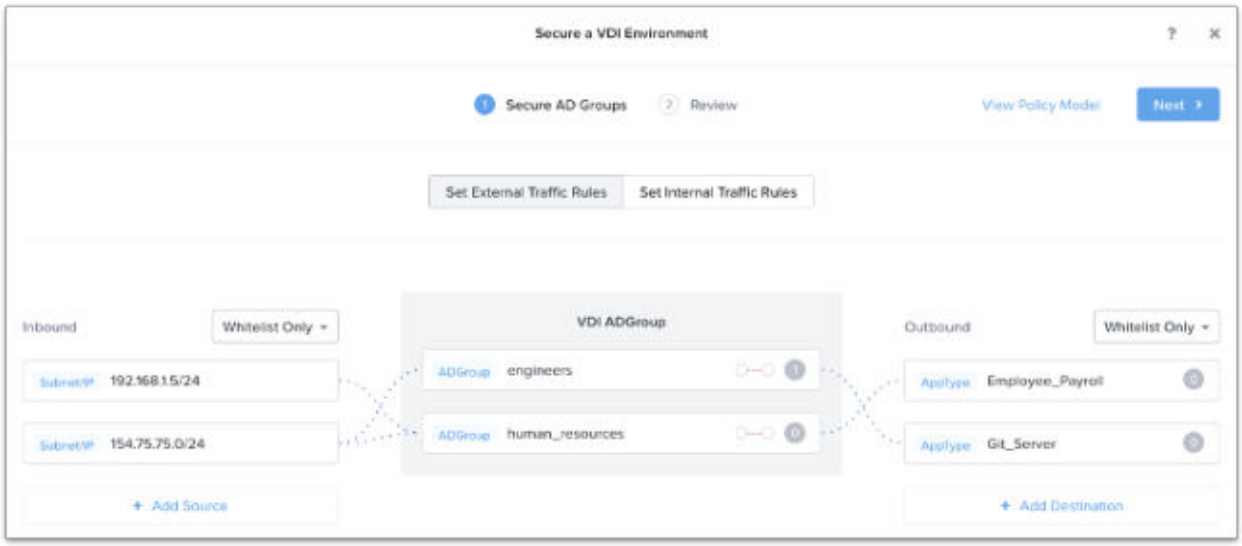


Figure 87: Using an AD-Type Policy to Secure a VDI Environment

The following figure shows an example of using Flow microsegmentation to control traffic for an application:

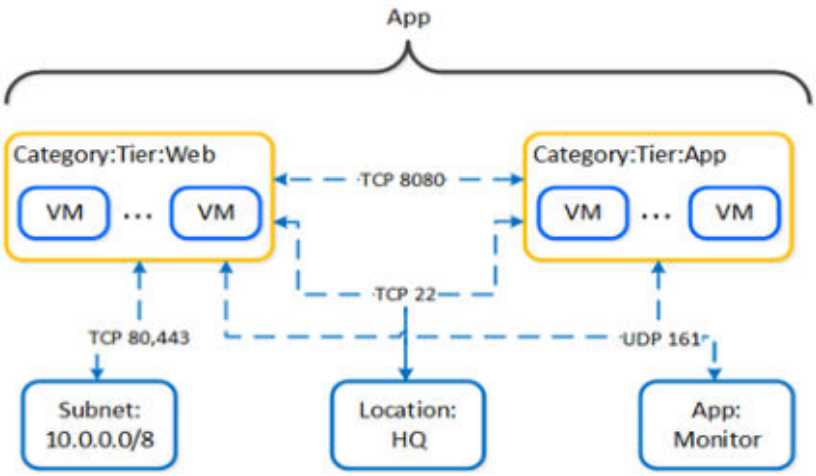


Figure 88: Using an AD-Type Policy to Secure a VDI Environment

Security Rule Actions

A security rule has an action field which determines what action is taken when a rule is matched. With AHV and Flow microsegmentation, there are two types of actions:

- Apply: Enforce the policy by permitting defined flows and blocking all others.
- Monitor: Allow all flows but highlight any packets that violate the policy on the policy visualization page.

Other Flow Features

While microsegmentation—including policy management, policy distribution, policy enforcement, and monitoring— is a core feature of Flow, the product offers many other capabilities such as:

- ID firewalls for VDI security
- Flow policy export/import
- Policy hit logs
- IPV6 enable/disable
- Service groups
- Address groups

- Audits
- Logging
- NCC alerts
- Service chains for advanced firewalls and integration with third-party devices
- Flow security central for advanced auditing, policy automation, governance
- L2 isolation
- External IPAM support
- Live migration connection tracking
- Nutanix Clusters integration

High-Level Design Considerations

Nutanix recommends the use of microsegmentation to isolate applications that don't interact or talk to each other directly.

Table 168: Use Microsegmentation to Isolate Applications

FLOW-001	USE MICROSEGMENTATION WITH APP-TYPE POLICY TO SEGMENT APPLICATIONS
Justification	Microsegmentation between applications can decrease the attack surface if a datacenter is penetrated and also prevent ransomware and other malware from spreading.
Implication	Nutanix AHV is required to utilize Flow microsegmentation.

If you are running VDI in your environment, using flow to isolate traffic based on AD groups is also extremely useful.

Table 169: Use AD-Type Policy to Isolate VDI

FLOW-002	USE MICROSEGMENTATION WITH AD-TYPE POLICY FOR VDI DEPLOYMENTS

© 2021 Nutanix, Inc. All rights reserved

Justification	Microsegmentation based on AD groups can isolate VDI environments, prevent users from accessing restricted resources, and help prevent ransomware and other malware from spreading.
Implication	Nutanix AHV is required to utilize microsegmentation.

Nutanix Files for Integrated File Services

Key Design Objectives

Table 170: Nutanix Files Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	NEW / UPDATE / REASONING
Enable file services from Nutanix clusters	Nutanix Files provides flexible file services implementations including: - Mixed or dedicated clusters - Flexible storage options - Multiple, separate file servers per cluster - Simple data protection and DR	New

Architecture Overview

Nutanix Files is a scale-out, software-defined file server that provides Server Message Block (SMB) and Network File System (NFS) file services to clients. Nutanix Files leverages Nutanix HCI for its core storage, networking, virtualization (AHV and ESXi) and compute requirements. Management for Nutanix Files is performed from Prism Element.

Nutanix Files instances are composed of a set of VMs (called FSVMs). For a given file server instance, Files requires at least three FSVMs running on three physical nodes to satisfy a quorum for high availability. You can scale out a given file server at any time, by adding FSVMs, up to the number of physical

nodes in the HCI cluster or a maximum of 16 FSVMs. You can also scale the CPU and memory of the set of FSVMs at any time.

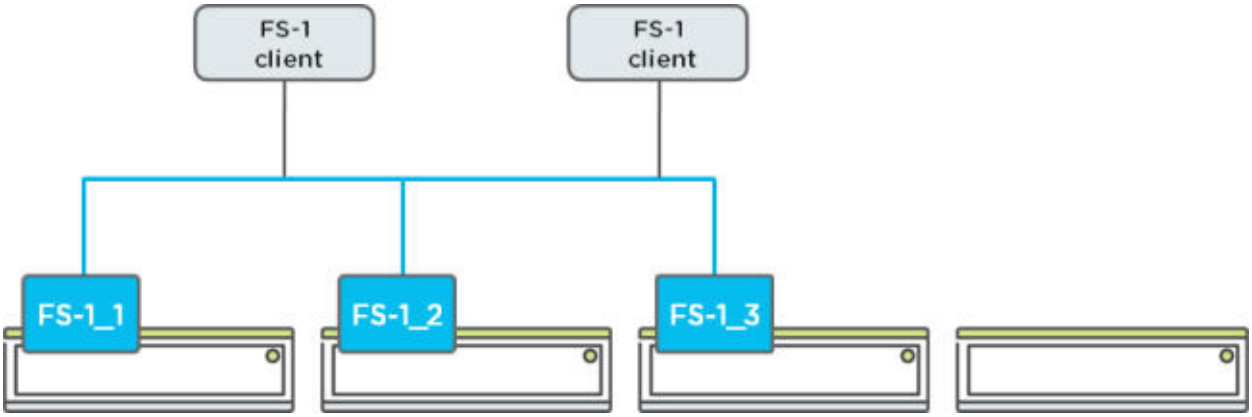


Figure 89: Example of a Nutanix Files Instance

Multiple Files instances can run on the same HCI cluster. Each instance represents an individual namespace. Each instance can scale from three FSVMs up to 16 FSVMs, assuming there are enough resources in the HCI cluster.

You can also deploy single FSVM file servers. Note that single FSVM file servers cannot scale out.

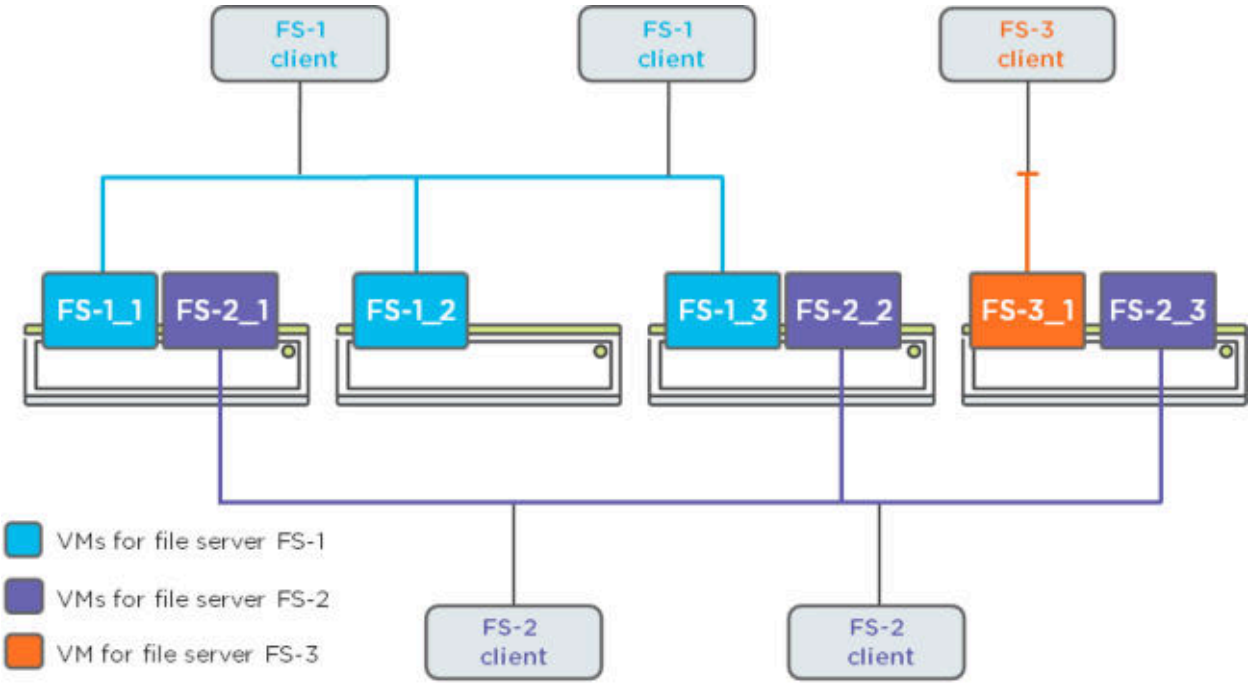


Figure 90: Three Separate Nutanix Files Instances Deployed on a Single Cluster

For file servers with three or more FSVMs, high availability during node failure and Files upgrades is achieved at the file-server level. Resources controlled by a failed or upgrading FSVM are moved temporarily to another FSVM. For single FSVM instances, hypervisor HA provides protection against node failure events.

Each file server has two networks: a client network and a storage network. The client network is used for SMB and NFS traffic and other required services such as active directory, DNS, and NTP. The storage network provides shared storage services for file shares and exports. Nutanix Volumes is used on the backend to provide shared storage, leveraging the iSCSI protocol over the storage network.

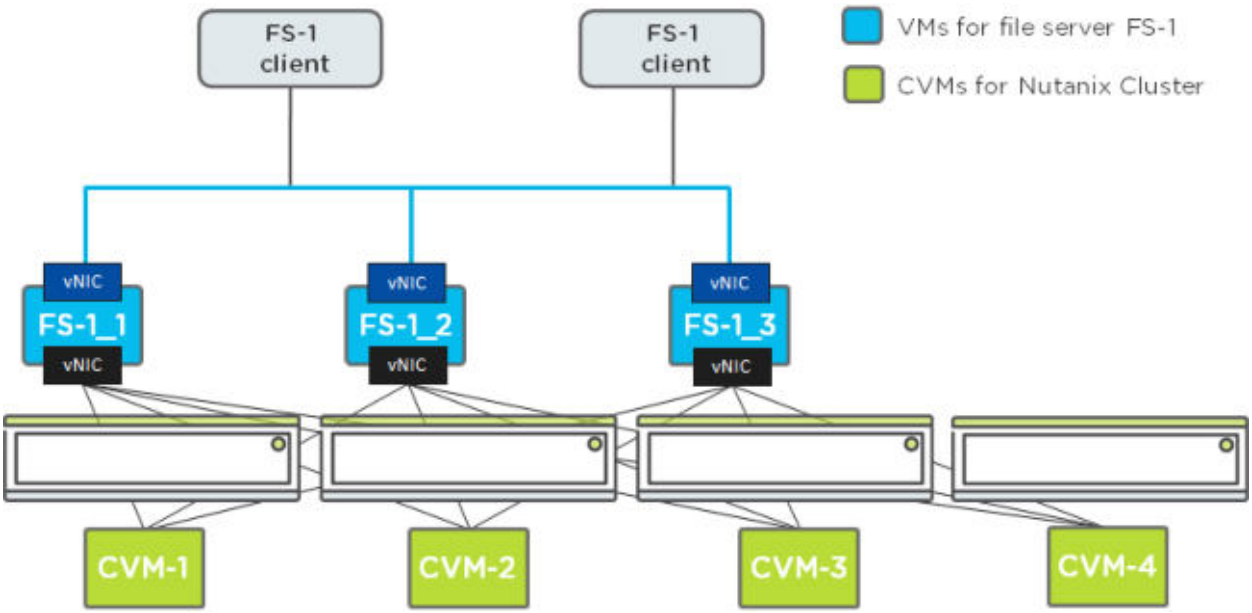


Figure 91: Relationship Between File Server VMs, CVMs, and Cluster Nodes

File shares are load balanced across FSVMs based on share type. Nutanix Files has two types of shares, distributed shares and standard shares.

Distributed shares have distributed ownership across all FSVMs in a Files cluster. Distribution occurs from the top-level directory (TLD) (\\ClusterName \DistributedShare\TLD). Each TLD can belong to a different FSVM and reside in its own volume group. Several volume groups support a distributed share; the total number of volume groups depends on the size of the Files cluster. Each FSVM has five volume groups to support distributed shares. For example, a three-node Files cluster has 15 volume groups to support shares.

© 2021 Nutanix, Inc. All rights reserved

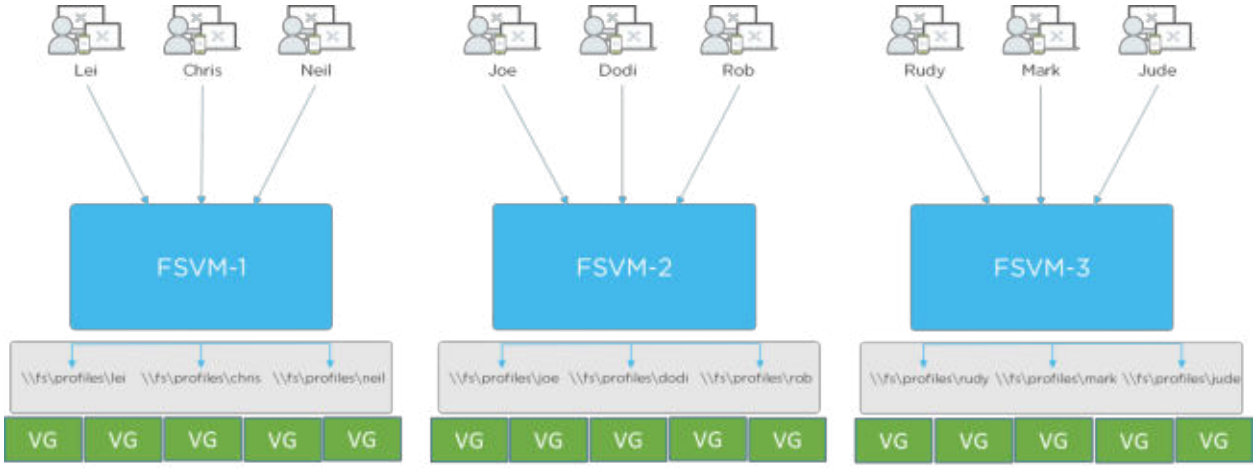


Figure 92: Distributed Shares are Distributed Across Multiple Volume Groups and FSVMs

Standard shares are owned by one file server VM (FSVM) at a time and have a single volume group to store data.

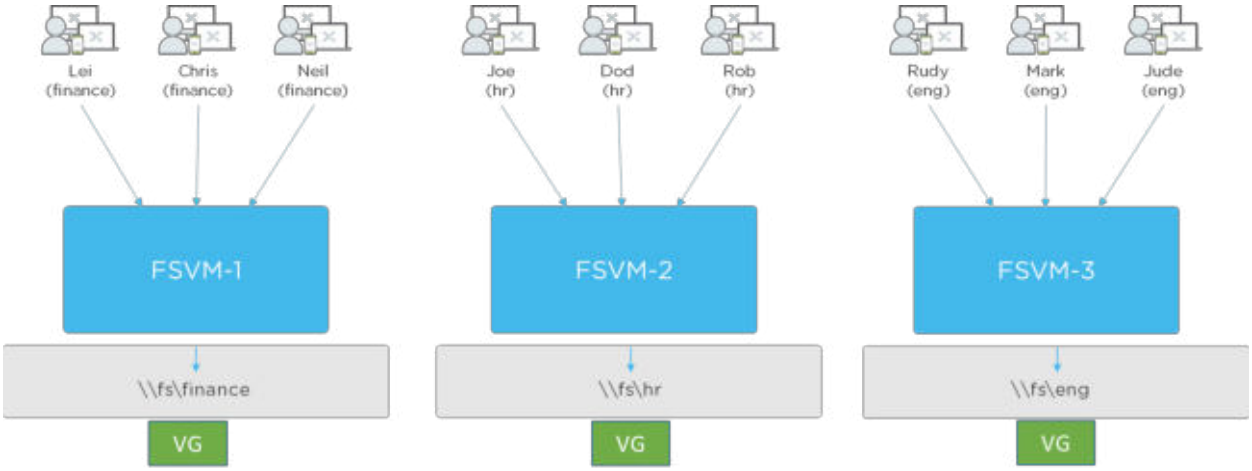


Figure 93: Standard Shares are Assigned to a Single FSVM and a Single Volume Group

For business continuity, Nutanix Files leverages the native data protection of Nutanix HCI. A protection domain is configured for each file server and native Nutanix HCI async and nearsync replication is used to protect configuration and user data.

Nutanix Files can run in a mixed environment along with other workloads, or on a dedicated cluster just for Files. Mixed environments require a Nutanix Files for AOS license, which is an add-on to your AOS license. Dedicated environments

require a Nutanix Files Dedicated license. In both cases Nutanix Files is licensed based on usable TiB from the perspective of the SMB or NFS protocol, plus any space consumed by snapshots.

High-Level Design Considerations

Mixed or Dedicated Clusters

There is no hard and fast rule regarding whether to deploy Nutanix Files in a dedicated cluster or a mixed cluster with other workloads. Like any other workload, the compute and storage density of your applications, performance requirements, failure domains, management domains, and other considerations need to be weighed.

Table 171: Mixed or Dedicated Environments

FILES-001	USE NUTANIX CLUSTERS DEDICATED TO FILE SERVICES WORKLOADS
Justification	File services tend to need storage dense configurations, with a higher ratio of HDD to SSD. Unstructured data workloads typically have limited data overwrites and large amounts of cold data, which benefit from erasure coding (EC-X).
Implication	Dedicating clusters to specific workloads can potentially lead to inefficient compute and storage utilization.

Hybrid Storage or All Flash clusters?

Every environment is different, but unstructured data has a tendency to have a small amount of active data and a large amount of inactive or cold data and also tends to be storage dense. High storage density with a majority of cold data leads many Nutanix Files deployments to utilize clusters with hybrid storage (a combination of SSD and HDD in each node). Considerations for sizing hybrid clusters is discussed in the Storage Capacity Utilization and Sizing section below. The Nutanix Files Sizing Guide (Nutanix Portal account required) can also help.

Table 172: Choose Hybrid Clusters

FILES-002	USE HYBRID STORAGE NUTANIX CLUSTERS
Justification	Most unstructured data use cases work well with hybrid storage solutions where hot data resides on SSD and cold data resides on HDD. The Nutanix Distributed Storage Fabric will tier data as required between local or remote nodes and SSD or HDD based on the workload.
Implication	End user performance will be negatively impacted if an insufficient SSD tier is sized for the use case.

Number of FSVMs per File Server Instance

Each file server instance consists of some number of FSVMs. It's recommended to deploy at least 3 FSVMs for high availability at the file server level. The amount of compute you assign to each FSVM, and the number of FSVMs beyond the minimum of three, depend on your performance requirements. You need to ensure there is enough compute assigned to FSVMs to support the expected number of concurrent user connections. The release notes (System Limits section) for each version of Files contains a table that provides guidance. For example, the number of supported SMB connections per FSVM in the Nutanix Files 3.8 release is:

- 4 vCPU/12 GB memory = (supports up to) 500
- 4 vCPU/16 GB memory = 1000
- 6 vCPU/24 GB memory = 1500
- 6 vCPU/32 GB memory = 2000
- 8 vCPU/40 GB memory = 2750
- 12 vCPU/48 GB memory = 3250
- 2 vCPU/64 - 96 GB memory = 4000

You also need to ensure that enough storage IOPs and throughput are available for your specific workload. The number of FSVMs also plays a role in performance. Recommendations for sustained read and write performance per FSVM can be found in the Nutanix Files Sizing Guide (Nutanix Portal account

required) and the Nutanix Files Performance Guide (Nutanix Portal account required).

Once the minimum number of FSVMs and compute resources are determined, it is recommended that you take a balanced approach to the number of FSVMs, allowing you to scale-up or scale-out if needed. Having a 16-node AOS cluster does not mean you need a 16 FSVM file server.

Table 173: Minimum FSVMs

FILES-003	USE A MINIMUM OF 3 FSVMs
Justification	Three FSVMs or more provides improved high availability with less-disruptive maintenance operations versus single FSVM configurations.
Implication	Using a minimum of three FSVMs may lead to excessive resource utilization in environments with multiple file servers. Also, ROBO configurations require a minimum of three physical nodes which could lead to the over allocation of resources.

Table 174: N+1 Configuration

FILES-004	USE AN N+1 CONFIGURATION OF FSVMs TO PHYSICAL NODES
Justification	Having one physical node more than the number of FSVMs reduces performance impact during HA events.
Implication	Maintaining an additional node leads to unused compute resources during normal operations.

Table 175: CPU and Memory Allocation

FILES-005	ALLOCATE MORE THAN THE MINIMUM FSVM CPU AND MEMORY RECOMMENDED FOR YOUR PLANNED WORKLOAD
Justification	The recommended FSVM configuration for SMB environments is the minimum to support the specified number of users. Allocating more than the minimum will help ensure better performance and avoid under sizing.
Implication	Allocating more than the minimum memory and CPU for FSVMs may lead to over-allocated resources. You can always scale-up FSVM resources to improve performance or support more users online when needed.

Number of File Server Instances

Many environments will need to deploy multiple file server instances on the same hardware. Some of the considerations for deploying multiple file servers include:

- Replacing an existing file server environment with multiple namespaces. Each file server instance represents a namespace.
- Different remote replication (RPO) requirements.
- Multiple untrusted Active Directory domains. Nutanix Files supports one registered Active Directory per file server instance. Domain trusts are required for accessing a common file server.
- File server administration between different teams or tenants may require multiple file servers with separate management domains.
- Nutanix Files supports one external VLAN per file server instance. Non-routed VLANs may require their own file server instances.

Once the number of file server instances is determined, you can then determine the total compute requirements per physical node.

© 2021 Nutanix, Inc. All rights reserved

Table 176: Minimum Number of File Servers

FILES-006	USE THE MINIMUM NUMBER OF FILE SERVER INSTANCES FOR YOUR NEEDS.
Justification	Using the minimum number of file server instances helps simplify management and reduce wasted resources.
Implication	Single, large file server instances can result in longer maintenance operations, such as scaling-up or upgrades.

File Server Networking

It's recommended that the file server client network be in close relation to the clients accessing the shares for best performance. It's also recommended that you place the file server storage network on the same VLAN as eth0 of the CVM. A flat non-routed storage network for iSCSI traffic will ensure the best performance.

Table 177: File Server Storage Network

FILES-007	PLACE THE FILE SERVER STORAGE NETWORK ON THE SAME VLAN AS THE eth0 INTERFACE OF THE CONTROLLER VMS.
Justification	Using the same network VLAN for the FSVM storage network and the CVM iSCSI network ensures the best performance.
Implication	Files does not currently support the iSCSI network segmentation feature of AOS. For security purposes, some environments may require complete separation of the iSCSI network from any management network.

Table 178: Shortest Route Between Client and Server

FILES-008	PLACE THE FSVM CLIENT NETWORK SO IT USES THE SHORTEST NETWORK ROUTE FOR THE MAJORITY OF FILE SERVER CLIENTS.

© 2021 Nutanix, Inc. All rights reserved

Justification	Ensuring the shortest route between client and server will give the best performance for your end users and applications.
Implication	Creating file servers with networks local to your end users can improve performance but may lead to more complex environments, increasing administrative burden.

Standard Shares or Distributed Shares?

Distributed shares are recommended for the majority of use cases. Distributed shares will help to ensure data is hosted by all FSVMs. Data must be spread across multiple top-level directories within a distributed share to ensure load balancing. Standard shares use an individual volume group and can only scale to 280TB (with the Files 3.7 release). Standard shares are also only hosted by one FSVM at a time. Multiple standard shares would be needed to load balance an environment across FSVMs.

Table 179: Utilize Distributed Shares

FILES-009	USE DISTRIBUTED SHARES
Justification	Distributed shares provide automatic load balancing of top-level directories across FSVMs. Distributed shares help to limit the number of shares needed to load balance an environment, simplifying administration.
Implication	Distributed shares do not support files in the root of SMB shares today (Files 3.7 has support for NFS). Using the Nutanix Files MMC to manage top level directories can add some complexity to administration.

Storage Capacity Utilization and Sizing

Nutanix Files leverages the core data services of the Nutanix platform, including fault tolerance and replication factor options. For data reduction, inline LZ4-based compression per file share (performed by Nutanix Files) and erasure coding at the storage container level are recommended.

© 2021 Nutanix, Inc. All rights reserved

Deduplication is not recommended. Any storage sizing exercise for Files should take into account the potential for savings based on both compression and erasure coding. File types should also be taken into account. For example, file types like video and image data will not benefit from storage compression. Cold data benefits from erasure coding.

Storage capacity consumed by self-service snapshots and/or data protection snapshots should also be factored into storage sizing. Nutanix Sizer helps to determine potential capacity when considering data reduction, snapshot overheads and change rates for Nutanix Files workloads.

Table 180: Compression

FILES-010	USE FILE SERVER SHARE-LEVEL INLINE COMPRESSION
Justification	Compression provides considerable space savings with minimal performance impact in Files environments.
Implication	Small potential for performance impact.

Table 181: Erasure Coding (EC-X)

FILES-011	USE AOS CONTAINER-LEVEL ERASURE CODING
Justification	Considerable storage space savings will be obtained by ensuring EC-X is enabled in Files environments.
Implication	Files environments with heavy overwrite activity could see a performance impact with EC-X enabled.

Table 182: Deduplication

FILES-012	DO NOT ENABLE AOS CONTAINER-LEVEL DEDUPLICATION
Justification	Testing has shown minimal space savings benefits when using storage container deduplication with Nutanix Files.
Implication	A small amount of additional space utilization may occur.

© 2021 Nutanix, Inc. All rights reserved

Table 183: Scheduling Snapshots

FILES-013	USE TELESCOPING SNAPSHOTS SCHEDULES
Justification	Telescoping schedules (x hourly, y daily, z weekly) ensures efficient space utilization while minimizing performance impact.
Implication	Keeping fewer long-term snapshots limits fine-grained recoverability over time.

Disaster Recovery

Nutanix Files supports both async and nearsync replication. Files deployments commonly use storage-dense nodes which historically have had snapshot frequency limitations when using AOS based data protection.

Nutanix Files 3.8 introduces Smart DR. With Files Smart DR the replication engine is no longer managed by the AOS clusters using native Nutanix protection domains. Nutanix Files now directly manages the replication. Because Files manages replication, node density limits specific to AOS snapshots and replication no longer apply. You can now use our most storage-dense nodes, which support up to 350TB of hybrid storage, with the benefits of native remote replication.

If you choose to continue to use AOS based snapshots and replication, follow the node density, SSD, and CVM compute requirements for your expected RPO. Read more in the Data Protection Guide.

Table 184: Disaster Recovery

FILES-014	USE THE MINIMUM NUMBER OF DISASTER RECOVERY SNAPSHOTS NEEDED FOR YOUR BUSINESS RPO REQUIREMENTS.
Justification	Keeping a smaller number of short duration snapshots ensures the most efficient storage utilization.
Implication	Retaining fewer long-term snapshots limits fine-grained recoverability over time.

Table 185: Disaster Recovery

FILES-015	USE FILE ANALYTICS
Justification	File Analytics provides native auditing, anomaly detection, data age categorization, reporting and ransomware protection. These features help you better understand your unstructured data and can protect you from unwanted activity.
Implication	File Analytics consumes additional compute and storage resources on your cluster. The File Analytics release notes provides guidance on sizing your File Analytics VM.

Nutanix Objects for Integrated Object Storage

Key Design Objectives

Table 186: Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	NEW / UPDATE / REASONING

Enable S3-compatible object store within Nutanix clusters	Nutanix Objects lets you create flexible object stores suitable for a variety of use cases. Features include: - Mixed or dedicated clusters - Flexible storage options - Advanced storage efficiency - Multiple object stores per cluster - Object stores spanning multiple clusters - Versioning - WORM - Object lock - Data protection and DR	New

Architecture Overview

Nutanix Objects is an S3-compatible, software-defined, scale-out object storage solution. Nutanix Objects leverages Nutanix AOS for its core storage, network, virtualization (AHV and ESXi) and compute requirements. Management of Nutanix Objects is performed from Prism Central.

Nutanix Objects is deployed as a set of Kubernetes Pods running on top of the Nutanix Microservices Platform (MSP). MSP is a fully managed Kubernetes deployment, where the entire lifecycle is managed by the Nutanix platform. There are two types of services which are deployed on MSP, Objects Worker Nodes (Workers) and Objects Load Balancers.

Nutanix Objects deployments can start as small as three nodes and scale-out across multiple AOS clusters. This multi-cluster scalability ensures that you can build large-scale Objects deployments with petabytes of data and billions

of objects. Objects deployments can live on dedicated AOS clusters (Objects Dedicated) or be deployed alongside existing workloads.

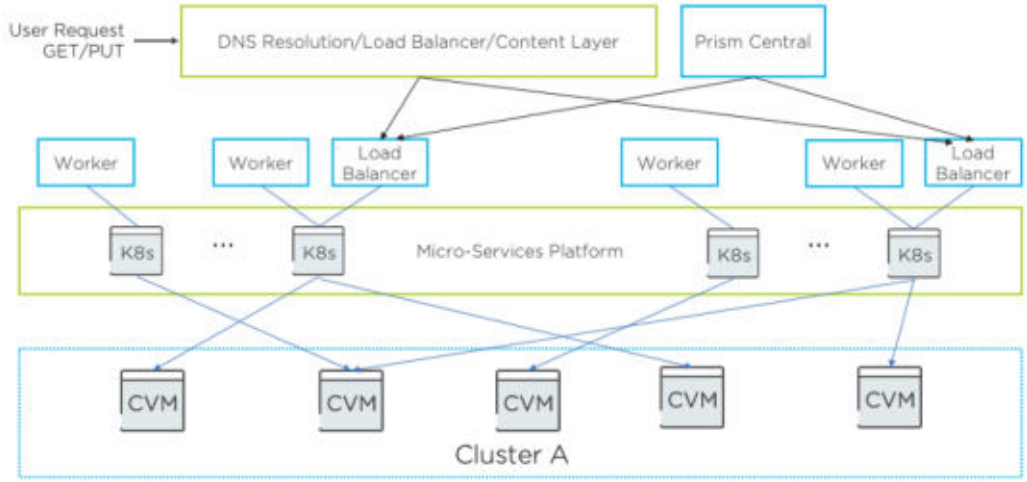


Figure 94: Nutanix Objects Deployments

Multiple Objects instances can run on the same HCI cluster. These instances can represent different namespaces (ex. customer-a.domain.tld, customer-b.domain.tld) or different use-cases (ex. backups.domain.tld, splunk.domain.tld). Each instance can scale from three Workers to 31 Workers in a single cluster.

For Object Stores with three or more Workers, high availability during node failure and Objects upgrades are performed at the object server level. All API requests for objects are redirected to other Workers by the Load Balancers during an HA or Upgrade event. For single-node Objects instances (Dev/Test use cases), hypervisor HA is utilized during HA or Upgrade events.

Each Worker has two networks: An Internal network and an External network. Each Load Balancer (has two networks as well: An Internal network and an External network, used for client access. The Worker utilizes the Internal network interface to communicate to the Load Balancer, while the Load Balancer utilizes its External network interface to communicate with clients utilizing the S3 API protocol.

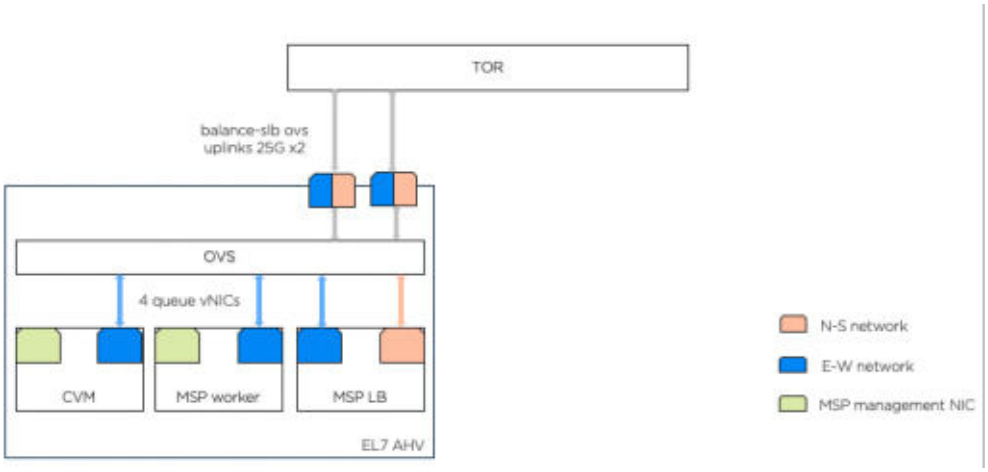


Figure 95: Internal and External Networks

Hints about the physical location of the metadata are stored at the Objects layer, enabling faster reads. This is because the hint can be passed for the lookup at the CVM layer, and we can skip the vBlock lookup. This keeps the latency for time-to-first-byte low.

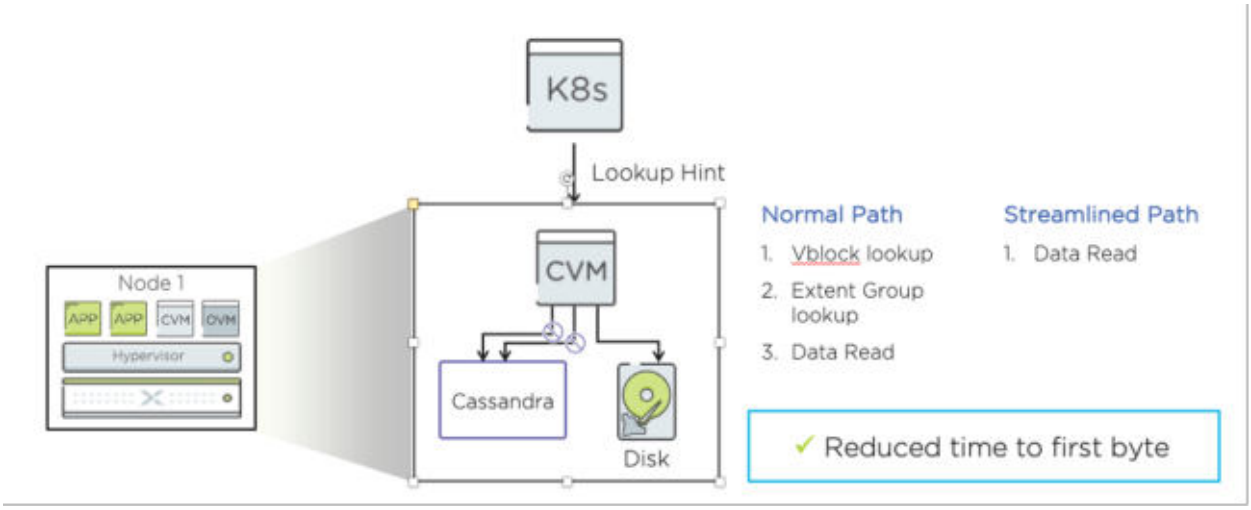


Figure 96: Lookup Hints

For business continuity, Nutanix Objects has replication built in. Streaming replication allows objects to be replicated at the bucket level from one cluster to another. Given that objects are inherently immutable—meaning a particular object cannot be edited—having a replicated copy of the object that may or

may not be versioned gives an organization a powerful business continuity option. Coupling Versioning and WORM (Write Once, Read Many) allows you to keep and store copies of objects for a set period. WORM ensures that even an administrator or Nutanix support engineer cannot delete your objects, allowing you to meet compliance needs or protect against ransomware.

Nutanix Objects may run in a mixed environment along with other workloads, or in a dedicated environment just for Objects. You can license for mixed environments with the Nutanix Objects for AOS add-on to your AOS license. Licensing for dedicated environments requires the Nutanix Objects Dedicated license. In both cases Nutanix Objects is licensed based on Used TiB from the perspective of the S3 protocol, not the amount of storage used on disk after Compression or Erasure Coding.

High-Level Design Considerations

Objects Networking

Table 187: An Object Store Requires Internal and External Networks

OBJECTS-001	PROVIDE INTERNAL AND EXTERNAL NETWORKS
Justification	The internal network is used for communication between the components of an object store. The external network is used for client communication.
Implication	N/A

The Internal network needs to have IPAM enabled for that network in AHV. See the Network Configuration section of the documentation for more info.

Mixed or Dedicated Environments?

Table 188: Mixed or Dedicated Objects Clusters

OBJECTS-002	MIXED OR DEDICATED OBJECTS CLUSTERS?

Justification	Mixed Objects clusters, combining Objects and production workloads, are most often used for small deployments or alongside applications that are actively utilizing object-based storage. Dedicated Objects clusters are for large object-based storage deployments, such as backup/archive, big data, or object stores where a high level of performance is desired.
Implication	Mixed clusters will be utilizing resources alongside production workloads, so proper sizing of existing environments must be considered. Dedicated clusters will require their own hardware, which includes the need to size for rack space, power, cooling, networking, and many other physical factors.

Whether to deploy Nutanix Objects in a mixed or dedicated environment is mostly determined by your workload needs. Like any other workload, the compute and storage density of your applications, performance requirements, failure domain, management domain and other considerations need to be weighed.

Hybrid Storage or All Flash Clusters

Table 189: Hybrid or All Flash Clusters

OBJECTS-003	HYBRID OR ALL FLASH AOS CLUSTERS

Justification	Hybrid clusters that combine HDD and flash storage are used for very large deployments where the cost per TB of raw capacity is the most important concern. Use-cases for Hybrid clusters often include backup, archive, and big data. All Flash clusters are used for smaller deployments or when high levels of performance and the lowest application latency are required.
Implication	Hardware costs are often much higher for All Flash clusters vs Hybrid.

Given that objects are write-cold, the decision between Hybrid vs All Flash infrastructure primarily comes down to workload access patterns and cost. If access patterns are highly random and need very low latency, then consider All Flash. If workloads are highly sequential (e.g., Backup/Archival) hybrid may be best. Utilizing Nutanix's standard sizing methodologies is usually the best way to size infrastructure for Nutanix Objects.

Storage Efficiency

Table 190: Enable Compression

OBJECTS-004	Enable Compression
Justification	Enable compression for all workloads. Erasure Coding is enabled for all Nutanix Objects deployments by default.
Implication	Compression can significantly reduce storage consumption for many data types.

Nutanix Objects utilizes AOS and all its storage efficiency capabilities to ensure that your object store consumes storage as efficiently as possible. After seven days, all objects are erasure-coded because Objects utilizes a "never overwrite" architecture by design.

© 2021 Nutanix, Inc. All rights reserved

Number of Worker Nodes Per Object Server Instance

Table 191: Number of Worker Nodes

OBJECTS-005	NUMBER OF WORKER NODES
Justification	The number of worker nodes should first be based on resiliency requirements. For anything other than Dev/Test, multiple workers are recommended. Light, medium, or heavy performance configurations can be chosen based on your application's required requests per second.
Implication	More memory and CPU capacity are needed as the number of worker nodes increase.

Nutanix Objects is a scale-out platform rather than scale-up. As performance needs increase, you simply add more workers.

Recommended Configurations:

- Light Performance (1,000 requests per second) – (3) Workers, 10vCPU/32GB Memory each, (2) Load Balancers, 2vCPU/4GB Memory each
- Medium Performance (10,000 requests per second) – (4) Workers, 10vCPU/32GB Memory each, (3) Load Balancers, 2vCPU/4GB Memory each
- Heavy Performance (20,000 requests per second) – (5) Workers, 10vCPU/32GB Memory each, (4) Load Balancers, 2vCPU/4GB Memory each
- Custom – Scale out worker nodes as performance needs increase. Maximum number of worker nodes is equal to the number of nodes in an AOS cluster. A maximum of (4) Load Balancers can be deployed
- Single Node – (1) node deployments for Dev/Test can be created by choosing a custom deployment and setting the resources to 10vCPU/32GB Memory

© 2021 Nutanix, Inc. All rights reserved

Multiple Object Store instances

Table 192: Multiple Object Stores

OBJECTS-006	MULTIPLE OBJECT STORES
Justification	Multiple object stores can be deployed on a single cluster to support: - Multi-tenancy: Requires authentication across multiple domains. - Different VLANs: In a highly segmented network with routing concerns between VLANs, a customer may choose to deploy multiple object store instances across each network.
Implication	More resources (CPU, Memory) will be required for the additional Worker and Load Balancer nodes. Prism Central and all AOS nodes will need to have trunked access to all VLANs needed for deployment.

Nutanix Objects supports deploying multiple Object Store servers. This allows you to deploy multiple instances based on your requirements. This could be used in a multi-tenant situation, different networks, etc.

Multi-Cluster Object Stores

Table 193: Multi-Cluster Object Stores

OBJECTS-007	MULTI-CLUSTER OBJECT STORES

Justification	Very large, multi-petabyte object stores can be spread across more than one AOS cluster. This can also allow you to utilize land-locked storage in other clusters to scale out the storage available to Nutanix Objects.
Implication	All clusters utilized for multi-cluster object stores must be within the same datacenter and registered to the same Prism Central instance.

Since Nutanix Objects utilizes Nutanix AOS for storage, it is important that clusters are sized to meet AOS recommendations. For very large multi-petabyte object stores, you may split the object store across multiple AOS clusters.

Objects enables you to utilize storage from other AOS clusters within your datacenter. For large, multi-rack object stores you can utilize Nutanix Objects to deliver a very large namespace spread across multiple clusters and even multiple hypervisors.

Multi-cluster object stores are also useful when you need to utilize “land-locked” storage from other Nutanix clusters, ensuring you utilize your existing capacity efficiently.

Versioning

Table 194: Enabling Versions

OBJECTS-008	ENABLE VERSIONING?
Justification	Versioning is used when an application or organization needs to retain the old versions of changed objects.
Implication	Each version is another copy of the data that lives on disk. Additional capacity is utilized to store the old versions.

Object versioning ensures that previous versions of an object are kept when the object is overwritten with a PUT or POST. Previous versions can be set to be deleted after they reach a certain age.

WORM (Write Once, Read Many)

Table 195: WORM

OBJECTS-009	WORM (WRITE-ONCE, READ-MANY)
Justification	Versioned WORM can be used when an application requires that a DELETE request for an object must occur, but the organization needs to retain the previous version of that object for a specified period. Versioned WORM can be used when an application requires that a DELETE request for an object must occur, but the organization needs to retain the previous version of that object for a specified period. Non-versioned WORM is used when an organization wants to block any DELETE operations on the objects in a particular bucket.
Implication	Versioning utilizes additional storage.Objects effected by a WORM policy cannot be deleted, utilizing cluster storage for the period specified in the policy. Objects effected by a WORM policy cannot be deleted, utilizing cluster storage for the period specified in the policy.

Nutanix Objects supports versioned and non-versioned WORM. WORM ensures that no one can permanently delete any object within a specified window of time (e.g., 3 years). WORM is set at the bucket level.

WORM is very useful for protecting against ransomware, as the object store will deny any object changes or deletions.

Versioned WORM allows DELETE operations to occur, causing the current version of an object to be labeled as a previous version, while retaining no current version. This is useful when the application needs to be able to do

DELETE operations, but the organization requires that the object be retained for a certain period.

Non-versioned WORM will not allow any DELETE operations to occur on any object in a bucket with WORM enabled.

Object Lock

Table 196: Object Lock

OBJECTS-010	OBJECT LOCK
Justification	Used for applications that require the ability to set WORM policies on individual objects via API (e.g., Veeam).
Implication	Objects clusters must be sized for additional storage requirements for objects marked with WORM policy.

Object Lock is object-level WORM that can be set per object by applications via APIs.

Lifecycle Policies

Table 197: Lifecycle Policies

OBJECTS-011	LIFECYCLE POLICY
Justification	A Lifecycle policy can be set when an application or organization requires the ability to delete objects after they reach a certain age. This is often used in conjunction with a WORM policy to ensure that objects are kept for the amount of time necessary to meet regulations, but no longer than that.
Implication	Objects are automatically deleted once they reach the specified age. Ensure this is the desired behavior or retain replica copies on another cluster using streaming replication.

© 2021 Nutanix, Inc. All rights reserved

Lifecycle settings on a bucket allow you define the time when objects are permanently deleted. WORM settings always take precedence over Lifecycle settings, ensuring objects are kept for the required amount of time.

Streaming Replication

Table 198: Streaming Replication

OBJECTS-012	STREAMING REPLICATION
Justification	Backups of objects need to be stored off-site or be made highly available across multiple sites.
Implication	Duplicate hardware required for secondary object store. Prism Central at both sites must be set up as an Availability Zone in order for Streaming Replication to see both sites.

Nutanix Objects has built in, object-based replication. Streaming Replication enables you to replicate objects from one cluster to another with an RPO that can be measured in seconds.

Nutanix Mine for Integrated Backup

Key Design Objectives

Table 199: Nutanix Mine Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	NEW / UPDATE / REASONING
Enable integrated backup	Nutanix Mine integrates backup into the Nutanix ecosystem using Nutanix-controlled infrastructure as a backup target and integrating with proven third-party data protection solutions.	New

© 2021 Nutanix, Inc. All rights reserved

Architecture Overview

Nutanix Mine is an integrated backup solution, combining the performance, scalability, and ease of use of Nutanix HCI with the robust feature set of industry-proven backup vendors. Nutanix Mine automates the provisioning of your secondary storage solution, optimizing performance and ensuring adherence to backup vendor best practices.

Mine can be purchased and configured in a number of predefined sizes and can be scaled-out as needed:

Table 200: Nutanix Mine Configurations

	XSmall	Small	Medium	Scale-out
Rack Size	2U	2U	4U	2U
# of Nodes	4	4	4	2
Raw Capacity	48 TB	96 TB	192 TB	96 TB
Usable Capacity (Extent store in N+1)	15TB	30TB	60TB	+40TB
Effective Capacity (assuming 2:1 comp/dedup	30-50 TB	60-100 TB	120-200TB	60-100TB
Scale-out unit	XS	Single NX8235-G7	Single NX8235-G7	-

At the time of this writing, Nutanix Mine supports two backup vendors: Veeam and HYCU. For future Mine supported backup vendors, reference the Mine section in the Nutanix Portal Documentation

Veeam

Veeam Backup & Replication (VBR) provides a full suite of backup capabilities for all Nutanix-supported hypervisors as well as physical workloads. In addition to VM workloads, Veeam supports the ability to back up Nutanix Files.

- VBR leverages a distributed architecture in which backup components such as Backup Proxies, Backup Repositories, and management consoles can be deployed in multiple sites as needed, to maximize performance and minimize performance bottlenecks.
- Veeam leverages Microsoft Windows as its underlying operating system for the VBR server as well as Backup Proxies for vSphere and Hyper-V hypervisors. Configuring backup jobs for vSphere, Hyper-V, and physical workloads is performed within the VBR management console.
- For Nutanix AHV, a separate Linux-based Veeam AHV Backup Proxy is deployed on the AHV cluster(s) you wish to back up, and backup job configuration and management is performed via a separate web interface.
- Backup jobs are defined on a per-VM-group basis, with backup schedules, archival implementation, and RTO and RPO targets defined by the user for each group.
- Veeam supports the ability to tier to S3-compatible object storage, which, with its WORM capabilities, can help defend against ransomware attacks.

HYCU

HYCU provides native Nutanix backup support with an emphasis on simplicity and ease of use via web-based administration and configuration. HYCU supports:

- Backup for Nutanix AHV and VMware vSphere workloads with native Nutanix snapshot awareness.
- Use of Nutanix Change File Tracking (CFT) APIs for optimized backup of Nutanix Files.
- Ability to back up physical servers as well as integrated application awareness for Oracle Database, SAP HANA, Microsoft SQL Server, and others.
- The ability to backup directly to an S3-compatible object store as well as the ability to schedule an archival job to replicate backups from the Mine appliance to object storage.

© 2021 Nutanix, Inc. All rights reserved

HYCU closely integrates with Nutanix Disaster Recovery Replication technology to allow for a distributed architecture and to accommodate ROBO workloads in which native Nutanix snapshots replicated via Nutanix Protection Domains can be leveraged within the HYCU backup solution.

High-Level Design Considerations

Choosing a Backup Vendor

Table 201: Choose a Backup Vendor

BAC-001	CHOOSE A BACKUP VENDOR TO USE WITH NUTANIX MINE
Justification	Choose a backup vendor from the list of qualified vendors for use with Mine.
Implication	Your current backup vendor may not be qualified.

When deciding which Mine Backup solution to use in your deployment, the choice comes down to which solution best meets your business and technical requirements and your budget. Key areas to consider when choosing a backup vendor are:

- Primary workload support (including legacy and current OSes, applications, and File Storage)
- Backup and retention requirements
- Staff skill set and training (architecture, administration, daily operations)
- Scalability of the solution
- Licensing costs and model (including existing licensing agreements)
- Integration with the existing environment
- Migration of existing backup repositories and data
- Management plane technical features and performance
- Simplicity or complexity of the solution
- Features or products that support only one hypervisor or the other

© 2021 Nutanix, Inc. All rights reserved

- Satisfaction with existing backup solution(s)
- ROI/TCO of the full stack

Freedom of choice is a key tenet of Nutanix. After considering these factors, it is likely that one backup platform stands out as the best option for your deployment. No matter which backup software you choose, the solution is backed by world-class Nutanix support as well as the chosen backup vendor's support.

Deploy Mine in Conjunction with Nutanix Objects

Table 202: Utilize Objects for Archival

BAC-002	UTILIZE NUTANIX OBJECTS AS AN ARCHIVAL TIER FOR BACKUPS
Justification	Both Veeam and HYCU support S3-compatible object stores, so you can deploy Nutanix Objects as an archival tier with either option for ransomware protection and operational and cost efficiency. Optionally HYCU can be deployed to backup directly to an S3-compatible object store
Implication	May require additional licensing for Nutanix Objects.

Nutanix Mine can be deployed in a number of configurations depending on backup storage capacity and performance. Because both Veeam and HYCU support offloading backups to an S3-compatible Object Store, Nutanix Objects can be leveraged as an archival tier for backups. Using Nutanix Objects in conjunction with a Nutanix Mine deployment offers the following benefits:

- Ransomware protection: By using WORM-enabled Buckets and Objects' S3-compatible Object Lock API, backups can be protected against ransomware and deliberate or accidental deletion.
- Greater flexibility: With the use of Nutanix Objects, backup administrators can choose to keep backup data on either or both the Mine Cluster and

© 2021 Nutanix, Inc. All rights reserved

Object Store and leverage the benefits of Objects Replication to distribute backup copies across datacenters or regions.

- Cost-efficiency: For long-term retention requirements, administrators can choose to tier off backups to Objects and free up Mine Secondary storage for more recent or higher priority backups

The following diagram highlights the different architectural patterns in which Nutanix Mine can be deployed across single or multiple datacenters and/or regions.

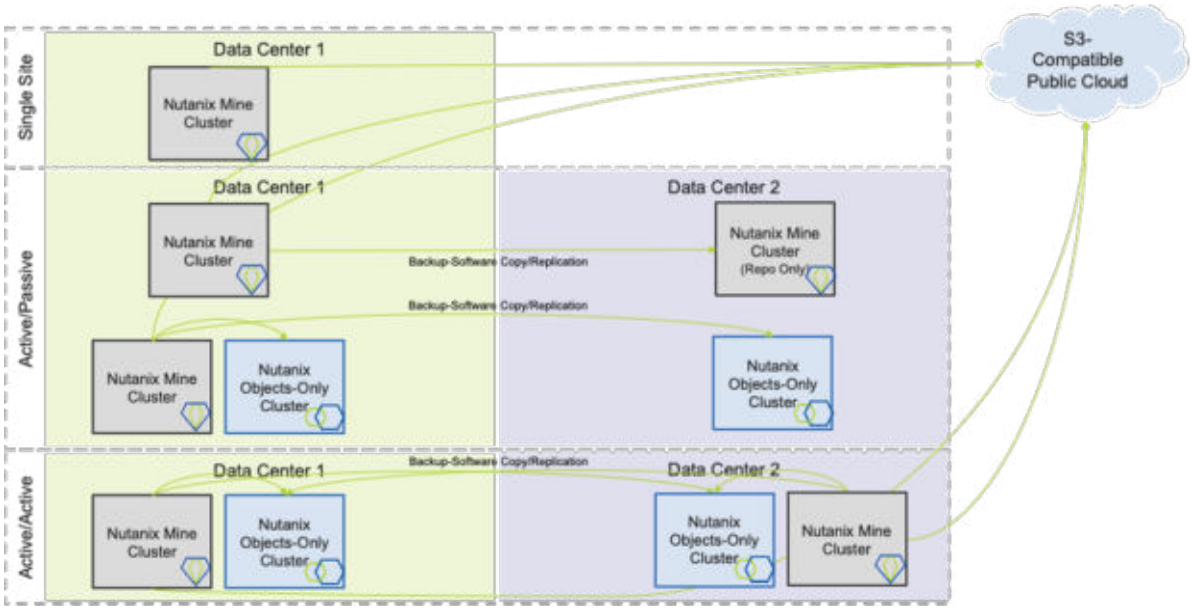


Figure 97: Mine Deployments Across Datacenters or Regions

Calm Application Orchestration

Key Design Objectives

Table 203: Nutanix Calm Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	NEW / UPDATE / REASONING

Private/Public cloud automation and lifecycle management	Virtual Machines, Bare Metal endpoints, Containers	Update

Architecture Overview

Nutanix Cloud Application Lifecycle Management (Calm) is an enterprise-grade, multi-cloud application and infrastructure lifecycle management framework. Nutanix Calm provides application lifecycle management, monitoring, and remediation to manage heterogeneous infrastructure, including VMs, bare-metal servers, and containers. Nutanix Calm supports multiple platforms and endpoints so that you can have a single self-service and automation framework that can manage your applications and infrastructure. The core automation constructs available in Calm are:

- Blueprints. Blueprints provide automation instructions for single or multiple instance(s) (VMs, containers, or native cloud services) governing application creation and lifecycle management of the instance(s). Blueprints can support instance creation, deletion, and any number of standard or custom day 2 management actions.
- Runbooks. Available with Calm version 3.x and later, runbooks provide the ability to create a set of automation tasks that are endpoint agnostic. Any day-to-day management operation that is performed manually, can be encapsulated in a Calm Runbook and executed against any service in your IT estate (VMs, bare-metal instances, containers, or native cloud services.)

Nutanix Calm includes several additional capabilities for efficient management of Hybrid Cloud automation:

- Self-service via the Calm Marketplace, the ServiceNow plugin for Nutanix Calm, or any self-service platform of your choice via Calm API integration.
- Governance via built in RBAC and project resource management
- Resource quota management (for on-premises endpoints only) via Policy Engine
- An intuitive and user-friendly GUI

- A powerful code-based approach with a Python-based Domain Specific Language (DSL) that Nutanix has released and open-sourced.
- The API-centric design of Calm enables extensibility for integration with many industry automation, configuration management, and self-service solutions such as Ansible, Terraform, Puppet, Chef, SCCM, ServiceNow, Cherwell, BMC Remedy, etc. This enables organizations to leverage existing intellectual capital from any of these solutions, allowing for rapid deployment and a short learning/adoption curve.

Hybrid Cloud management enables you to leverage all of the Calm capabilities on non-Nutanix providers (VMWare on Nutanix and non-Nutanix infrastructure) and the major public cloud solutions (AWS, GCP, and Azure). Nutanix Calm is a native service in Prism Central and is automatically enabled in Prism Central for both single Prism Central and scale-out Prism Central deployments.

Design Considerations

Sizing Considerations and Lifecycle Management

When enabling Calm, you must choose between a small or large deployment.

Table 204: Choose a Small or Large Calm Deployment

CALM-001	SMALL OR LARGE CALM DEPLOYMENT
Justification	You can start with a small deployment and easily scale to a large deployment at a later date.
Implication	To make an informed decision, you must first familiarize yourself with the sizing parameters, requirements, and Certified Configurations in the Nutanix Calm Administration and Operations Guide. Your chosen sizing allows you to hot-add necessary resources to Prism Central instance(s) with no disruption to Prism Central services.

It is recommended to upgrade Calm to the latest version as quickly as possible when new releases become available. This allows you to leverage new features and capabilities quickly and ensures you receive the latest available fixes/patches, etc. Calm is deployed through Prism Central and tied to the same user interface. However, the two can be upgraded independently of one another. As with Prism Central, all Calm releases are Long Term Support (LTS); Calm does not provide Short Term Support (STS) releases. See the earlier section Operations Design for information on LTS and STS.)

Table 205: Upgrade to Most Recent Calm Version

CALM-002	UPGRADE TO THE MOST RECENT VERSION OF NUTANIX CALM
Justification	The Calm product team delivers fixes, enhancements, and new features on a regular cadence. You will want to leverage these enhancements as quickly as possible. Upgrading Calm is a non-disruptive LCM function that allows Calm to be upgraded separately from Prism Central or other application services running on Prism Central.
Implication	Calm blueprints from a newer version of Calm are usable with older versions of Calm.

Infrastructure-as-Code solutions require the creation of software code and a developer mindset. Robust processes for code creation and release management are imperative. It is recommended to deploy Calm in a minimum of two separate Prism Central physical instances, one for development and one for production. Larger enterprise deployments may consider up to four separate environments if their existing datacenter architecture includes separate environments and processes for Development, Test, QA, and Prod.

With this implementation methodology, you can follow the recommendation Calm-002 and always work with the latest Calm release in the development environment, and then roll the Calm upgrades through your other environments

as part of a lifecycle management. This enables development and testing of existing and new blueprints with the latest version, while at the same time supporting existing blueprints and applications via the Production Calm instance. This also allows for maintaining a previous Calm version for backwards compatibility for existing blueprints.

Table 206: Deploy Calm for Development and Production

CALM-003	SEPARATE CALM INSTANCES FOR DEVELOPMENT AND PRODUCTION
Justification	Maintaining a development instance of Prism Central and Calm allows for an independent upgrade of Calm and integration testing against all blueprints and runbooks prior to upgrading in production. This provides an additional advantage in cases where new features become available, and you wish to update blueprints or even new automation blueprints or runbooks leveraging the new features prior to upgrading production.
Implication	This requires a separate instance of Prism Central (this could be a single instance even on a single node cluster) on a separate Nutanix cluster. This also adds an additional instance of Calm to manage from the point of view of RBAC access for developers, etc. This is minimal overhead especially since only automation engineers and developers will be leveraging the Dev instance. If you currently leverage a release management process, it may require updates to include Calm.

© 2021 Nutanix, Inc. All rights reserved

Business Continuity and Disaster Recovery

It is recommended that you create a process (automated or manual) to execute Nutanix-provided backup scripts via SSH from Prism Central to protect Calm. These scripts are documented in the Calm support documentation. If a cold Prism Central instance is not immediately available for a restore, then a new Prism central/cluster instance can be provisioned and used as a target for a Calm restore. The only risk with this approach is the impact to RTO due to having to rebuild the Prism Central instance as a new target.

If you have enabled the policy engine for your Calm instance, then there is a separate script to create backups of your policy engine database. The policy engine database backup can be used to restore the policy engine to either an earlier state or to a new policy engine VM.

Table 207: Calm for DR and Backup

CALM-004	DEFINE A PROCESS FOR CALM DR AND BACKUP/ RESTORE
Justification	In the event of data corruption or loss of Prism Central, Nutanix Calm blueprints, runbooks, and configurations need to be backed up. It is recommended that the backup scripts are scheduled to run at a set time each day.
Implication	A separate “cold” Prism Central instance(s) with Calm enabled is required as a restore target for the backup/ restore script. An external scheduler is required to run the scripts via SSH. Customers with Prism Pro can use an X-Play Playbook to call a Calm Runbook to execute the backup script. Keep in mind that Prism Central has security features that limit its ability to remotely execute some SSH commands, and the Calm Runbook will need to SSH to an independent endpoint (VM or container) to run the backup command.

Management and Operations

Calm requires integration with an AD/LDAP source, and Calm can leverage Prism Central's existing AD/LDAP configuration. Local Prism Central users cannot be used for user and role assignments in Calm. The AD/LDAP integration for Calm is tied to the Project construct. It is a best practice to align your AD/LDAP security groups to projects and roles within Calm in order to achieve the desired self-service consumption model for each project. You can create security groups for Calm projects and roles and then add AD users to these groups to simplify access management via standard AD user authentication. (For example, for "ProjectA" all developers are assigned to a group called "ProjectADev" that is associated with the project. Adding new developers to this security group grants them access to the project resources based on their AD/LDAP credentials).

Control of project configuration (creation, group membership, etc.) can also be performed via Calm blueprints and Runbooks via API if you need a more dynamic RBAC project/resource model. This way, self-service features in Calm can be used to automate RBAC policies.

The Prism Central "admin" account can be used to gain/restore access to Calm if something happens to the Active Directory LDAP connection(s).

Table 208: Calm and RBAC Configurations

CALM-005	ALIGN PROJECT STRUCTURE AND RBAC CONFIGURATION
Justification	Design Calm Projects and Security Groups to align with AD/LDAP users and roles that are specific to entitlements for administering, developing, managing, and consuming automation constructs (blueprints, runbooks. etc.)
Implication	AD/LDAP configurations in Prism Central are required in order to leverage Calm RBAC features.

Calm has a native feature to help manage and share code across environments. End-users can select blueprints to provision into any environment in any geography at runtime. This allows blueprints to be created once and delivered to any Calm physical instance without having to manage multiple copies of

the same blueprint. Through RBAC, administrators can assign users/groups to projects that control access to who can provision which blueprints in which environments. It is a good design principle to align your physical Calm instances to both your environments (e.g., Dev, Test, QA, Prod) and your deployment geographies (e.g., Americas, Europe, Asia). Here is an example of what this might look like for a global enterprise deployment of Calm.

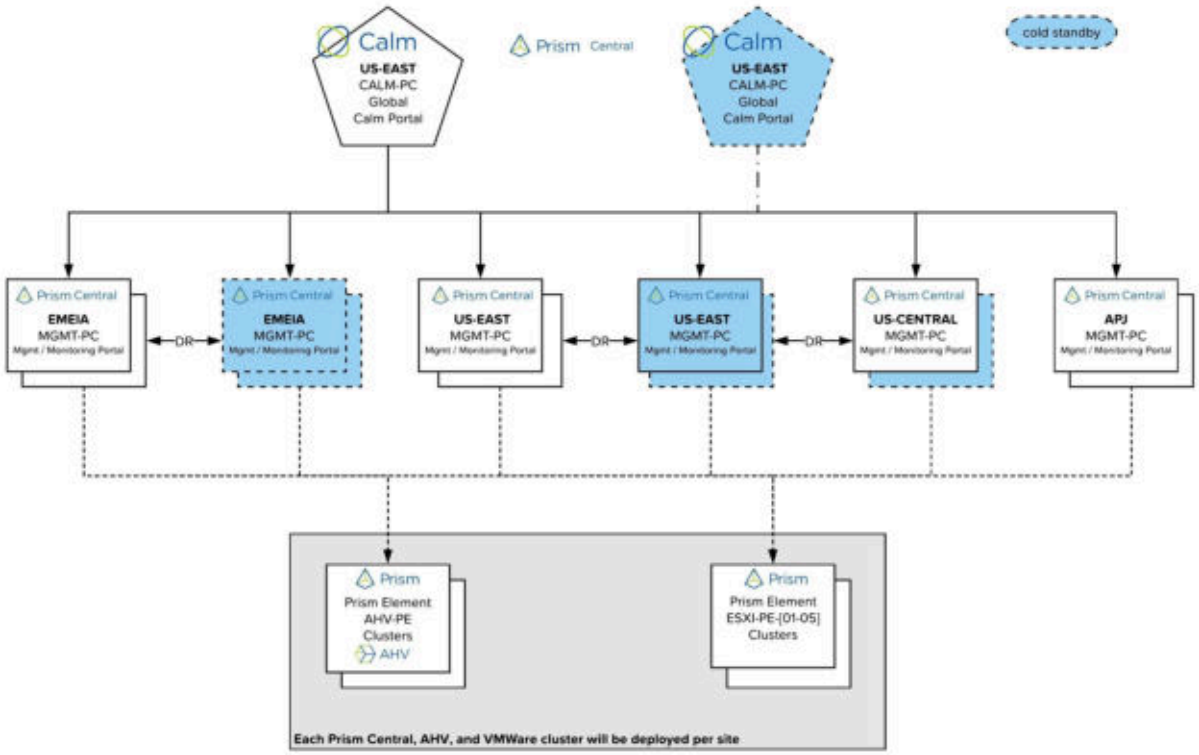


Figure 98: Global Enterprise Calm Deployment

Table 209: Calm Releases and Geographic Regions

CALM-006	DESIGN CALM ENVIRONMENT TO ALIGN WITH RELEASE METHODOLOGY AND GEOGRAPHIC REGIONS

Justification	This feature maximizes efficiency by allowing you to have one automation blueprint that can be chosen as a self-service request from end-users in the same cost center, but in different sites or geographies.
Implication	A Calm decision should be made while planning your infrastructure to establish physical clusters that align with your needs.

Nutanix Calm has the ability to set resource quotas for projects via its Policy Engine. Maximum quotas for compute, memory, and storage can be set for projects so that the total amount provisioned by all users in the project does not exceed the threshold. When you enable the Policy Engine for a Calm instance, a new VM is created and deployed. For quota policy enforcement, VMs must be provisioned through Calm. In order to enforce the desired policy for VMs that are not provisioned through Calm, migrate the VMs to Calm once you have enabled the Policy Engine.

Table 210: Calm Policy Engine

CALM-007	ENABLE CALM POLICY ENGINE
Justification	Prevents teams from over-provisioning or hogging resources and allows you to better manage consumption among business units and projects.
Implication	You need an available IP address that belongs to the same network as that of your Prism Central VM for the Policy Engine VM.

Integration with Nutanix Beam enables showback for Calm without any additional licensing costs. Nutanix Beam is a SaaS product that helps customers optimize their multi-cloud spend. Cost calculations in Nutanix Beam are configured for you, and Beam will automatically calculate your compute, memory, and storage costs. Since Beam runs in the Public Cloud, the automatic configuration of your cost data requires an active internet connection to Prism Central. In the event you do not have an active internet connection, you can still leverage Beam showback features by manually entering your cost data. Showback is currently available for Nutanix and VMware environments.

© 2021 Nutanix, Inc. All rights reserved

Table 211: Showback with Nutanix Beam

CALM-008	ENABLE SHOWBACK VIA NUTANIX BEAM
Justification	Showback helps you control IT spend by tracking utilization of IT resources to each business unit. Reporting can identify how much each business unit is consuming and helps in budgeting and cost optimization. You may also be able to implement Chargeback processes whereby business units pay for their consumption on a monthly or yearly basis.
Justification	Showback helps you control IT spend by tracking utilization of IT resources to each business unit. Reporting can identify how much each business unit is consuming and helps in budgeting and cost optimization. You may also be able to implement Chargeback processes whereby business units pay for their consumption on a monthly or yearly basis.

Self-service enables end-users to create new workloads and manage existing workloads provisioned by Calm without accessing the administrative consoles or needing help from IT Admins with privileged access. Self-service can be achieved via a variety of mechanisms: the Calm Marketplace, the ServiceNow plugin for Nutanix Calm, or the self-service platform of your choice via Calm API integration. The Calm Marketplace is native to Nutanix Prism and included with Calm.

The ServiceNow plug-in for Calm may be ideal for companies who already use the ServiceNow ITSM platform; you can expose Calm services to all end-users that have access to ServiceNow. In addition, ServiceNow approvals, incidents, and CMDB updates are integrated with the Calm Plug-in. The ServiceNow Plug-in for Calm is available from the ServiceNow store and adds no additional Nutanix licensing costs. For customers who prefer to use an in -house self-

© 2021 Nutanix, Inc. All rights reserved

service interface or some other third-party interface, Calm blueprints and runbooks can be invoked via REST API, enabling easy integration.

Table 212: Self-Service with Calm

CALM-009	CHOOSE A METHOD FOR SELF-SERVICE
Justification	Self-service allows end-users to perform provisioning and Day 2 operations without the need for privileged access or assistance from IT Admins.
Implication	Potential third-party licenses needed for self-service integration are not provided by the Calm Marketplace.

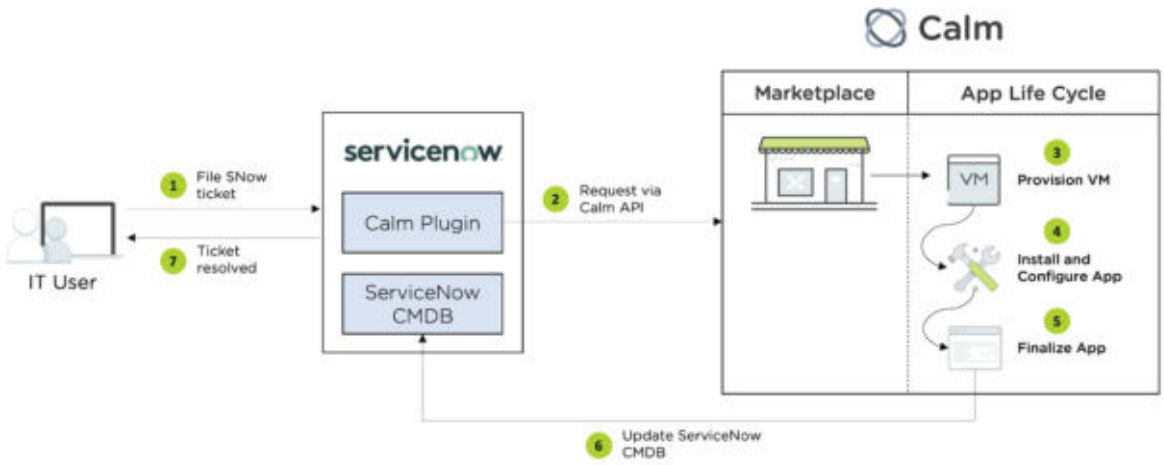


Figure 99: Self-Service with Calm

Karbon for Kubernetes and Container Support on Nutanix

Key Design Objectives

Table 213: Nutanix Karbon Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	NEW / UPDATE / REASONING

Enable support for containers and Kubernetes	Nutanix Karbon enables teams to incorporate containers and Kubernetes as part of the Nutanix ecosystem with simplified provisioning and management.	New

Architecture Overview

Nutanix Karbon is an enterprise-grade Certified Kubernetes distribution that simplifies the provisioning, operations and lifecycle management of Kubernetes, all from within Nutanix Prism. Karbon is an opt-in service running on Prism Central; upon enablement, Karbon Core and Karbon UI docker containers are instantiated on Prism Central VM(s), allowing you to be up and running in 5 to 10 minutes.

Once enabled, Nutanix Karbon can deploy and manage Kubernetes clusters to any AHV cluster that's registered to Prism Central. Any number of Kubernetes clusters can be deployed with any configuration, assuming adequate resources on the physical clusters. Day 2 management operations, such as host OS upgrades, Kubernetes version upgrades, storage class management, and worker node additions or removals are feasible via the UI or APIs.

The UI provides a mechanism to download a hardened host OS image provided by Nutanix (which is similar to the CVM) in a single click. Nutanix quickly reacts to any published security vulnerability and will push out a new host OS image with the appropriate vulnerabilities patched. Karbon admins can quickly download these new host OS images, and initiate host OS upgrades to bring a given Kubernetes cluster to the updated OS level in a couple of clicks. Alternatively, if the vulnerability is at the Kubernetes level, admins can also easily perform a Kubernetes version upgrade.

© 2021 Nutanix, Inc. All rights reserved

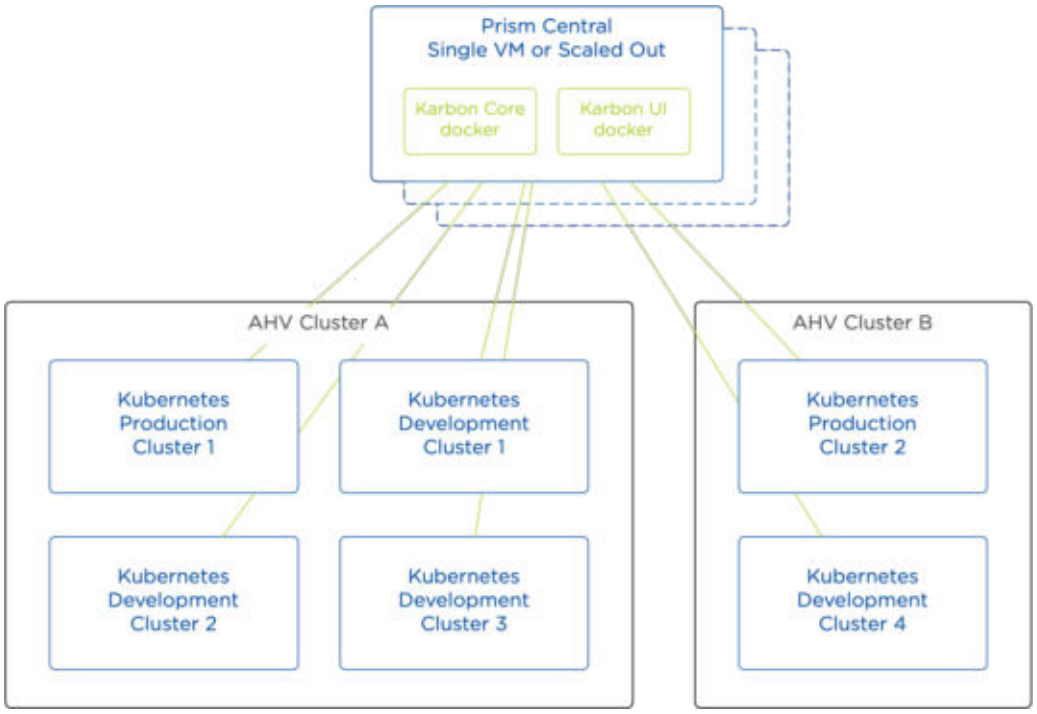


Figure 100: Karbon Provides a Single Control Plane for Multiple Kubernetes Clusters and Locations

Karbon can deploy both development and production Kubernetes clusters, depending upon use case. Development clusters are restricted to a single master node (which runs the Kubernetes control plane), a single ETCD node (Kubernetes' key-value store), and any number of worker nodes (which run end-user applications). The number of worker nodes can be specified during cluster creation and can be increased or decreased at any point in the cluster's life. Due to the single master and ETCD nodes, these clusters cannot tolerate failures in the control plane, and applications will experience downtime during upgrades. However, they're light on resources, making them perfect for development.

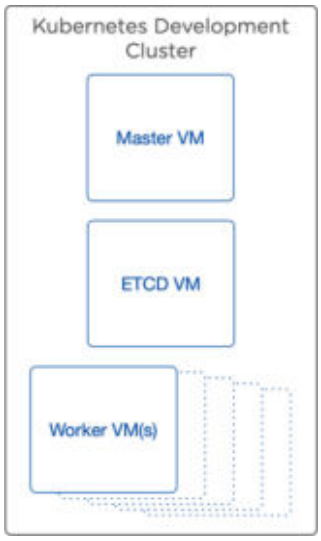


Figure 101: A Karbon Development Cluster is not Highly Available

Production Kubernetes clusters deployed by Nutanix Karbon have two master nodes in active-passive mode by default, utilizing VRRP for high availability. If the active master node goes down (either due to an upgrade or loss), the passive node will assume the virtual IP and the cluster will continue to function.

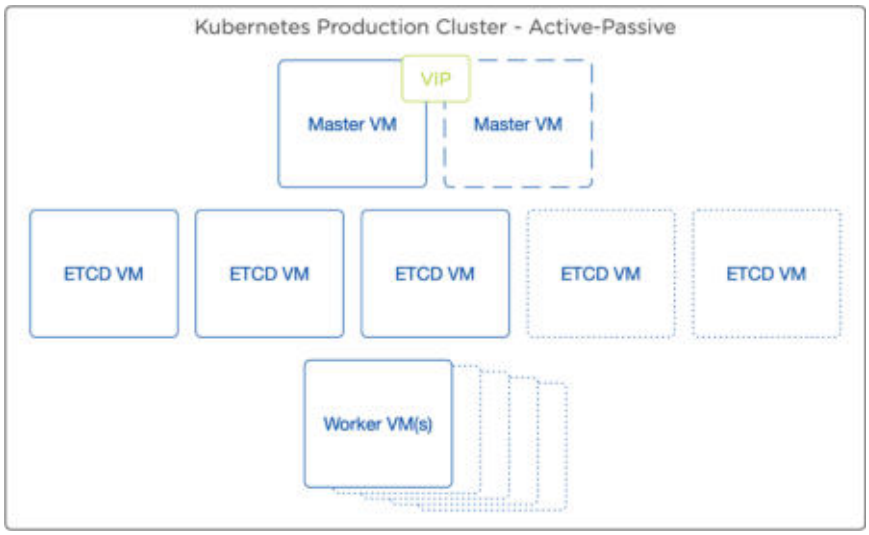


Figure 102: A Karbon Cluster for Production is Highly Available

Alternatively, if you have an external load balancer, anywhere from two to five master nodes can be deployed in an active-active configuration. This

increases performance, and as long as a single master node is online, the cluster continues to function.

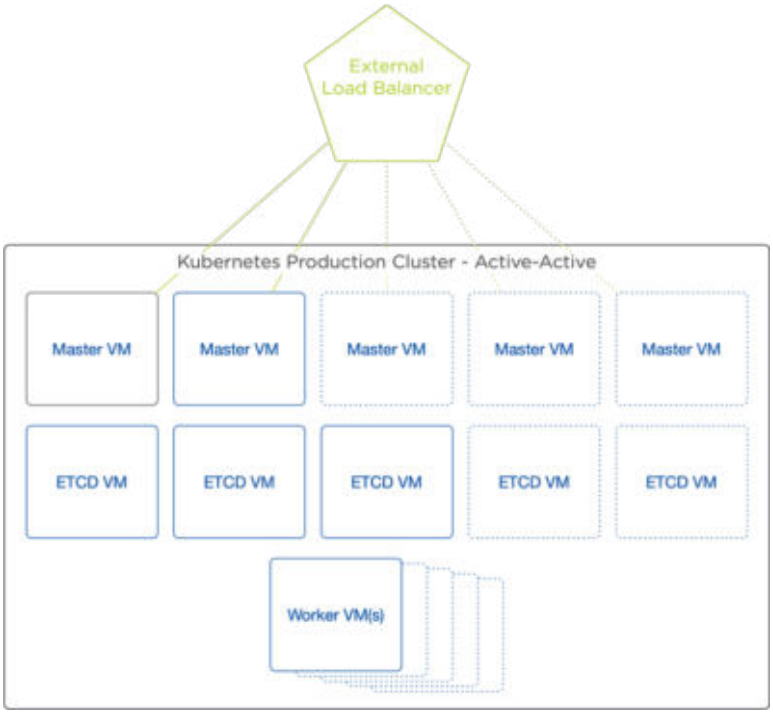


Figure 103: An External Load Balancer Enables Karbon to Have More than Two Master Nodes

Either three or five ETCD nodes can be deployed for production clusters; three enables the cluster to sustain a single ETCD node failure, and five enables two ETCD nodes to fail without causing a disruption. Stated differently, a quorum of two is required when three ETCD nodes are used, and a quorum of three is required when five nodes are used. As with development clusters, any number of worker nodes can be deployed at cluster creation, and later modified to scale with workload changes.

During deployment, admins are able to choose the physical network and VLAN which the nodes that make up the Kubernetes cluster belong to. All nodes must live in the same network. For development clusters, these networks can be backed by a DHCP server, or can utilize AHV managed networks (IPAM). For production clusters, you are required to select an AHV managed network. For Kubernetes cluster networking, either the Flannel or Calico CNIs can be utilized.

Nutanix Karbon Kubernetes clusters have the Nutanix CSI driver installed during cluster creation for persistent storage. A default storage class based on Nutanix Volumes is also defined and created during cluster creation, which provides Read Write Once storage (RWO). After cluster creation, any number of additional storage classes based on either Nutanix Volumes or Nutanix Files, which provides Read Write Many (RWX), can be created.

Karbon admins are also able to download a secure, 24-hour valid kubeconfig file from the Karbon UI, which allows for kubectl usage, which is the standard Kubernetes command line interface. This effectively provides root access to the cluster. Utilizing this, Karbon admins can create Kubernetes authentication objects, such as a Role, ClusterRole, RoleBinding, and ClusterRoleBinding, for end users. End users are added (either directly or as part of group membership) via Active Directory as Nutanix Read-Only users within the role mapping configuration section of Prism Central. This allows end users to log into the Karbon UI and download kubeconfig files.

Once these authentication objects are appropriately created, Karbon end users can easily create and manage applications utilizing kubectl, assuming they have been granted privilege to do so. This includes, but is not limited to, creating Kubernetes deployments, services, persistent volume claims, persistent volumes, and stateful sets. Ultimately the types of objects end users can create or manage is dependent upon the roles configured for them by the Karbon admin.

Since Kubernetes is a cloud native technology, it is recommended to implement Business Continuity at the application level. This means running multiple instances of your applications in different regions so that, in the event of a disaster, the secondary application continues to operate. For environments or applications where this is not possible, open source tools like Velero can provide Business Continuity at the Kubernetes cluster level.

Nutanix Karbon is included in every licensed version of Nutanix software. The only requirements are at least one AHV cluster and Prism Central.

High-Level Design Considerations

When configuring Karbon, first determine the availability needs of your applications and Kubernetes clusters. Development and test workloads are

likely okay to run on development clusters with single points of failure. Any production workloads should run on production clusters designed to ensure that a single VM loss will not bring down an application or an entire cluster. Business critical workloads should be distributed across multiple production clusters in multiple availability zones to ensure availability in the event of a disaster.

Next, determine your preferred CNI, either Calico or Flannel. Generally, Flannel offers simpler configuration, while Calico has more features. It is possible to have some clusters with Flannel and some with Calico to address specific use cases, however this may complicate management.

It is recommended to use the latest version of Kubernetes supported by Karbon unless there's a specific reason for using an older version. When possible, upgrade existing Kubernetes clusters to newer versions of Kubernetes.

Utilize a consistent naming convention for Storage Classes, based on volume type (Volumes or Files), reclaim policy, file system, and performance.

Table 214: Minimize Admin Kubeconfig Use

KARBON-001	MINIMIZE USE OF THE ADMIN KUBECONFIG
Justification	The admin kubeconfig provides root access to the cluster, so regular use of it is a security vulnerability. It should only be used for initial cluster configuration. Use AD-based Prism Read-Only credentials (with the corresponding Role and RoleBinding) instead, or Kubernetes Service Accounts for automation use cases.
Implication	Configuration of Prism role mappings, and Kubernetes objects (such as Roles, ClusterRoles, RoleBindings, ClusterRoleBindings, and ServiceAccounts) must be performed before running end-user applications on a cluster.

© 2021 Nutanix, Inc. All rights reserved

Table 215: Use Multiple Availability Zones for Production

KARBON-002	DISTRIBUTE PRODUCTION APPLICATIONS ACROSS AVAILABILITY ZONES
Justification	Production applications should run in an active-active or active-passive mode on multiple Nutanix clusters in different availability zones. In the event of a disaster, the secondary site can take over transparently.
Implication	A minimum of 2 AHV clusters and AZs are required.

Table 216: Use Namespaces to Segment Users and Apps

KARBON-003	UTILIZE KUBERNETES NAMESPACES TO SEGMENT USERS AND APPLICATIONS RATHER THAN CLUSTERS
Justification	Where possible, use of namespaces to segment work is preferable to reduce Kubernetes cluster sprawl and resource utilization.
Implication	This requires an approach to cluster configuration similar to that prescribed above in Karbon-001.

Table 217: Leave Room for a Worker Node to Fail

KARBON-004	LEAVE ENOUGH SPACE IN A CLUSTER TO TOLERATE A WORKER NODE FAILURE
Justification	Whether there is an unplanned outage, or a planned upgrade, sizing the cluster to tolerate a failure of a worker node allows applications to continue to run in the event of a failure.
Implication	At minimum, the cluster must be sized for n-+1, and workers must be added if the workload grows.

© 2021 Nutanix, Inc. All rights reserved

Table 218: Use Consistent CIDRs

KARBON-005	UTILIZE CONSISTENT POD AND SERVICE CIDRS
Justification	Whether utilizing Calico or Flannel as the CNI provider, during cluster deployment the admin must provide a large (/16 by default) subnet for use by Kubernetes Pods and Services. While these are internal subnets which do not go over the wire, they must not overlap with physical host networking. It is recommended to utilize the same subnets for all clusters for consistency and to minimize the chance of IP overlap.
Implication	Two large subnets are needed for Karbon Kubernetes clusters which do not overlap with physical host networking.

Table 219: Keep the Host OS Updated

KARBON-006	UPGRADE THE HOST OS AS SOON AS POSSIBLE
Justification	Due to the nature of Common Vulnerabilities and Exposures (CVEs), it is critical to upgrade Host operating systems quickly whenever a new version is released. This greatly reduces the chances of a cluster being compromised.
Implication	For development clusters, an outage is required. For production clusters, there will be a single point of failure during upgrades.

© 2021 Nutanix, Inc. All rights reserved

Era for Database as a Service

Key Design Objectives

Table 220: Nutanix Era Design Objectives

DESIGN OBJECTIVES	DESCRIPTION	UPDATE

Enable Database as a Service (DaaS	• Deliver Database as a Service (DBaaS) consumable via a user interface and REST API. • Perform a standardized and repeatable deployment of Postgres database instances. • Deploy and manage Postgres database instances in both: › HA mode › Single mode • Deploy Postgres HA instances across Nutanix clusters. • Protect (back up) database data to a location remote from the database instance. • Perform lifecycle and day2 operations for Postgres databases. • Provide copy data management functionality such as the ability to clone database instances. • Manage 100+ DB instances and DB VMs • Network communication is required between the Era management plane and the database instances managed by Era.	New

Architecture Overview

Era automates and simplifies database administration and brings one-click simplicity and invisible operations to database provisioning and lifecycle management, facilitating database as a service (DBaaS). Era

enables database administrators to perform operations such as database registration, provisioning, cloning, patching, restore, and much more. It allows administrators to define standards for database provisioning with end-state driven functionality that includes HA database deployments.

Era enables multi-cluster database management. Database administration for different databases across multiple Nutanix clusters can be performed with a single Era instance. With the extension of support for Nutanix Clusters, you can utilize all the capabilities of Era both in the cloud and on-prem.

The Era management plane operates on one or more VMs running in a Nutanix cluster with at least three Nutanix nodes (physical servers). The default Era management plane deployment in a single Nutanix cluster consists of one VM running all services:

- Front-end services (API, agent, web service).
- Back-end service (database or repository).

The following figure presents a logical implementation of Era DBaaS, including the Era management VM, database server VMs (DB), Nutanix cluster, and physical hardware

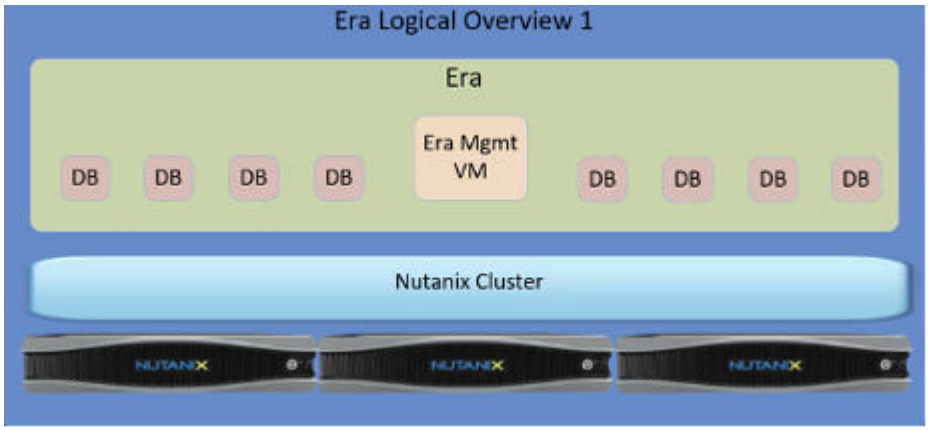


Figure 104: Nutanix Era DBaaS

In a multicluster Era environment, where Era manages database server VMs in multiple Nutanix clusters, the Era management plane requires one Era management agent VM running only the agent service in each additional Nutanix cluster it manages.

The following figure expands on the previous one to include the two Nutanix clusters and two Era management agent VMs.

Figure 105: Nutanix Era VMs and DBs

It is possible to run the Era management plane in a highly available mode to avoid service disruption from a node failure. This is done by provisioning an HA PostgreSQL instance where Era deploys a multi-node PostgreSQL instance. One of the nodes in the cluster acts as the leader, and other nodes serve as replicas. The leader node is used for both read and write operations while the replicas are only used for read-only requests. When a primary node becomes unavailable, one of the replicas is promoted to the role of master, ensuring availability of the database for write operations.

When Era is deployed in a high availability (HA) configuration, it includes six VMs:

As described earlier, Era supports managing database instances across multiple Nutanix clusters. The figure below illustrates the Era management architecture in a multi-cluster scenario where C1, C2, and C3 are different Nutanix clusters.

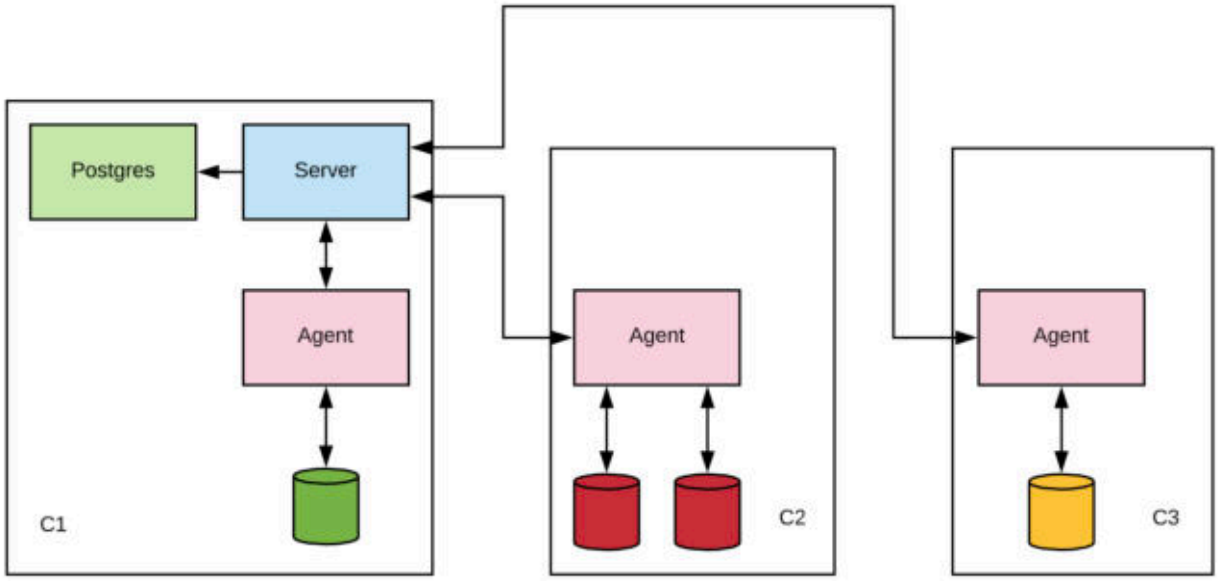


Figure 106: Era Management Layer in a Multi-Cluster Deployment

Based on the rich API provided by Era, you can automate any action available via the Era UI so that Era can easily be consumed from an overlying automation, self-service, or orchestration solution such as:

- Nutanix Calm
- ServiceNow
- VMware vRealize Automation

The Era management plane and the database instances managed by Era communicate during initial provisioning and registration, as well as during ongoing operation. Communication is mostly initiated from the database instances to the Era management plane, accept during provisioning and registration.

No specific hardware is required to run Era; it runs as a VM on any Nutanix AOS-supported hardware running Nutanix AHV or VMware ESXi.

Era employs role-based access control to govern access to an Era system. RBAC allows system administrators to restrict user access and limit the

operations they can perform. For example, a database operator can be granted access to only perform database-related operations.

With RBAC, you can add users either in Era or Active Directory and then later add those users to a group. Groups allow you to assign the same roles to multiple users at the same time. After adding a user, you can assign one or more roles to the user based on the operations they can perform.

There are multiple license options available for Era:

- License the entire Nutanix cluster
- License per physical CPU cores
- License per DB VM vCPU

High-Level Design Considerations

The design decisions presented in this section cover the Era management plane but not the database instances managed by Era.

Table 221: Era to Enable DBaaS

DBAAS-001	ENABLE DATABASE AS A SERVICE
Justification	Era can easily be integrated with overlying automation, self-service, and orchestration solutions.Era deploys DB instances with a predefined operating system, database engine binary and DB instance configuration in a repeatable fashion. Era can deploy and manage (clone, take a snapshot, perform log catchup, refresh) Postgres in single and highly availability mode in one or multiple Nutanix clusters. Snapshots of the managed database instances can be stored in the same Nutanix cluster where they are running or in remote Nutanix clusters.
Implication	N/A

Table 222: Era Hardware Requirements

DBAAS-002	MINIMUM HARDWARE REQS FOR NUTANIX CLUSTERS RUNNING ERA
Justification	The minimum configuration requirements shown below provide a solution with predictable performance. Since no specific performance requirements are established for this section, these requirements can serve as a baseline for running DB workloads.
Implication	Since workload characteristics are undefined the configuration below could be either too performant (increasing CapEx unnecessarily) or not performant enough and therefore need to be upgraded. Careful sizing should be performed for your intended DB workloads.

A Nutanix cluster supporting Era and database instances should meet the following minimum hardware requirements:

- Minimum 2.7GHz, 16-core CPUs
- An optimal memory configuration with minimum 768 GB memory per node. See Nutanix Physical Memory Configuration document for Nutanix NX recommendations.
- Minimum All flash storage
- Minimum 2x10 Gbps networks.

Table 223: Dedicated DB Clusters

DBAAS-003	USE DEDICATED NUTANIX CLUSTERS TO SUPPORT DATABASE WORKLOADS

Justification	The requirement to host more than 100 VMs justifies a minimum of two separate Nutanix clusters for database workloads to ensure that database workloads do not have to compete with non-database workloads for resources. The vCPU:pCPU ratio is typically lower for database workloads compared to other server virtualization workloads.
Implication	Hardware may be underutilized if not enough workloads are provisioned in the Nutanix clusters.

Table 224: HA Management Plane

DBAAS-004	DEPLOY ERA MANAGEMENT PLANE FOR HIGH AVAILABILITY
Justification	Provides the highest possible uptime for the Era management plane, ensuring that no changes are needed to address future changes in SLA uptime requirements. The solution protects against single-node failures.
Implication	When an Era VM providing front-end services or the hypervisor host (where the Era VM runs) becomes unavailable, there will be a short period of downtime for any user or service accessing Era.

Table 225: Management Network

DBAAS-005	PLACE ERA ON A SEPARATE MANAGEMENT NETWORK
Justification	The Era service is treated like any other management service and placed on a separate management network accordingly.

Implication	Firewall openings between the network where an Era server is hosted and the network(s) where the DB workloads are hosted are required if the networks are not the same.

Table 226: Era Platform License

DBAAS-006	USE THE ERA PLATFORM LICENSE
Justification	The Platform license lets you run as many DB workloads in a cluster as permitted by the underlying hardware and your acceptable vCPU:pCPU ratio. In addition to the Era license, the Platform license also includes the Nutanix AOS Ultimate license, so both Era and AOS licenses are managed as one entity.
Implication	Some AOS Ultimate features may go unused.

Table 227: Active Directory Integration

DBAAS-007	ACTIVE DIRECTORY IS USED FOR USER, GROUP, AND SERVICE ACCOUNT ACCESS
Justification	Using AD provides a single source of truth for user, group, and service management for traceability and auditing purposes. The local Era admin account can be used to gain access to the environment for super admins if the Active Directory connection is lost.
Implication	Era requires a connection to Active Directory to provide end user functionality.

© 2021 Nutanix, Inc. All rights reserved

Table 228: Active Directory Access

DBAAS-008	ACTIVE DIRECTORY GROUPS ARE USED TO PROVIDE ACCESS TO ERA
Justification	One place—Active Directory—is used to configure Era permissions (add and remove users from Active Directory groups) during normal operations.
Implication	Era requires a connection to Active Directory to provide end user functionality.

Table 229: Era Test Environment

DBAAS-009	AN ERA TEST ENVIRONMENT MUST EXIST
Justification	Upgrades and configuration changes must be applied and tested in a test environment before being applied in the production environment to avoid problems with new versions or upgrade processes.
Implication	An Era test system must be available.

Table 230: Service Windows

DBAAS-010	ERA UPGRADES ARE PERFORMED DURING SERVICE WINDOWS
Justification	There will be a very short service interruption during an upgrade. If a critical patch must be applied for stability or security reasons an upgrade can be performed during business hours after the service window is approved by the change control function
Implication	Could create some delay before a patch can be applied since a service window has to be requested.

Table 231: Era Roles

DBAAS-0011	USE ERA SUPER ADMIN AND DATABASE ADMIN ROLES
Justification	Provides a clear division between those managing the Era solution and those consuming DBaaS
Implication	N/A

7. Conclusion

This document is intended to demonstrate valuable methods and practices that organizations, both large and small, can use to implement Nutanix Solutions to solve their IT and business problems. There is no one-size-fits-all solution for Nutanix Hyperconverged Infrastructure, and the contents of this document are informational only and intended to provide customers with suggested best practices from which they can design and evolve their private, hybrid, and multi-cloud solutions.

For more information on any details of this document, please visit our website at www.nutanix.com or reach out to our sales team. You may also contact one of our many global support phone numbers listed on our website.

© 2021 Nutanix, Inc. All rights reserved

Appendix

Table of Design Decisions

The following table summarizes all the design decisions described in this document.

Table 232: Design Decisions

REFERENCE	DECISION NAME	DECISION
Region-001	Number of regions to be used	
AZ-001	Number of availability zones to be used	
DC-001	Number of datacenters to be used	
VRT-001	Choose Nutanix AHV or VMware ESXi as the hypervisor for your deployment.	
PFM-001	Management Cluster Architecture: Deploy a separate Management Cluster or share a cluster with other workloads. When choosing a separate management cluster, consider a redundant configuration.	
PFM-002	Mixed or dedicated workload per cluster.	
PFM-003	Select Nutanix software licensing level	
PFM-004	Select physical node vendor	
PFM-005	Select node model(s) per use case	
PFM-006	Number, type, and size of clusters	
PFM-007	Decide which workload domains will span a single or multiple racks	
NET-001	Use a large buffer datacenter switch at 10Gbps or faster.	

© 2021 Nutanix, Inc. All rights reserved

NET-002	Use a leaf-spine network topology for new environments.	
NET-003	Populate each rack with two 10GbE or faster ToR switches.	
NET-004	Avoid switch stacking to ensure network availability during individual device failure.	
NET-005	Ensure that there are no more than three switches between any two Nutanix nodes in the same cluster.	
NET-006	Reduce network oversubscription to achieve as close to a 1:1 ratio as possible.	
NET-007	Configure the CVM and hypervisor VLAN as native, or untagged on server facing switch ports.	
NET-008	Use tagged VLANs on the switch ports for all guest workloads.	
NET-009	Use a Layer 2 network design.	
NET-010	Connect at least one 10 GbE or faster NIC to each top-of-rack switch.	
NET-011	Use a single br0 bridge with at least two of the fastest uplinks of the same speed.	
NET-012	Use NICs of the same vendor within a bond	
NET-013	Use VLANs to separate logical networks.	
NET-014	Use active-backup uplink load balancing.	
NET-015	Use standard 1,500-byte MTU and do not use jumbo frames.	
NET-016	Use virtual distributed switch (vDS).	
NET-017	Connect at least one 10 GbE NIC to each top-of-rack switch.	
NET-018	Use a single vSwitch0 with at least two of the fastest uplinks of the same speed.	

NET-019	Use Route Based on Physical NIC Load uplink load balancing.
NET-020	Use standard 1,500-byte MTU and do not use jumbo frames.
CMP-001	If running a non-NUMA aware application on a VM, configure the VM's memory and vCPU to be within a NUMA node on AHV host.
STR-01	When creating vDisks in ESXi, always use thin- provisioned vDisks.
STR-02	When sizing a hybrid cluster, make sure to have enough usable SSD capacity to meet active data set of application.
STR-03	Do not mix node types from different vendors in the same cluster.
STR-04	Do not mix nodes that contain NVMe SSDs in same cluster with hybrid SSD/ HDD nodes.
STR-05	Minimum 2:1 HDD to SSD ratio required for Hybrid clusters.
STR-06	Size for N+1 node redundancy for storage and compute when sizing. For mission critical workloads that need higher SLAs, use N+2 node redundancy.
STR-07	Use FT=2 and RF=3 for workloads and clusters that need higher SLAs or for cluster sizes >32.
STR-08	Enable Inline Compression.
STR-09	Enable/Disable Deduplication.
VRT-001	Deploy scale-out Prism Central for enhanced cluster management.
VRT-002	Deploy prism central in each region or az using runbook dr automation.
VRT-003	Use VM-HA Guarantee.

© 2021 Nutanix, Inc. All rights reserved

VRT-004	Deploy HA vCenter instance with embedded PSC to manage all ESXi based Nutanix clusters.	
VRT-005	Deploy a vCenter in each region or AZ using runbook DR automation.	
VRT-006	Enable EVC mode and set to the highest compatibility level the processors in the cluster will support.	
VRT-007	Enable HA.	
VRT-008	Enable DRS with default automation level.	
VRT-009	Disable VM PDL and APD component protection.	
VRT-010	Configure das.ignoreInsuffi-cientHbDatastore if one Nutanix container is presented to the ESXi hosts.	
VRT-011	Disable Automation level for all CVMs.	
VRT-012	Set host failures clusters tolerate to 1 for RF2 and to 2 for RF3	
VRT-013	Set host isolation response to “Power off and restart VMs”.	
VRT-014	Set host isolation response to “Leave Powered on” for CVMs.	
VRT-015	Disable HA/DRS for each Nutanix CVM.	
VRT-016	Disable SIOC.	
DEP-001	A minimum of three NTP servers for all infrastructure components should be provided.	
DEP-002	A minimum of two DNS Servers should be configured accessible to all infrastructure layers.	
SEC-001	Use active directory authentication. This applies for user and service accounts.	
SEC-002	Use SSL/TLS connection to Active Directory.	

© 2021 Nutanix, Inc. All rights reserved

SEC-003	Use signed Certificate Authority (CA) certificates for the components where certificates can be replaced. This can be either internal or external signed certificates.
SEC-004	Do not use Nutanix Cluster lockdown.
SEC-005	Do not use vSphere Cluster lockdown.
SEC-006	Enable CVM and hypervisor AIDE.
SEC-007	Configure SCMA to run hourly.
SEC-008	Stop unused ESXi services and close unused firewall ports.
SEC-009	Send log files to a highly available syslog infrastructure.
SEC-010	Include all Nutanix modules in the logging.
SEC-011	Use Error log level for the Nutanix components.
SEC-012	Use default esxi logging level, log rotation, and log file sizes.
SEC-013	If extra security and reliability are required, then use tcp for log transport. Otherwise, use the default syslog protocol, udp.
SEC-014	Use port 514 for logging.
SEC-015	Use VLAN for traffic separation of management and user workloads.
SEC-016	Place CVM and hypervisor on the same VLAN and subnet.
SEC-017	Place out of band management on separate VLAN or physical network.
SEC-018	Use the least privileged access approach when providing access. and Align RBAC structure and usage of default plus customer roles according to the company requirement.

© 2021 Nutanix, Inc. All rights reserved

SEC-019	Align RBAC structure and usage of default plus custom roles according to the company requirements defined via SEC-019.
SEC-020	Do not use storage encryption.
SEC-021	Do not use a key management server.
OPS-001	Deploy Prism Central Pro for enhanced cluster management.
OPS-002	Review monthly capacity planning.
OPS-003	Perform updates after hours for performance or migration sensitive applications.
OPS-004	Utilize the current LTS branch.
OPS-005	Update to the next maintenance version 4 weeks after release. Update to the current patch version 2 weeks after release.
OPS-006	Maintain a pre-production environment for testing any changes needed (firmware, software, hardware) prior to executing the change actual production.
OPS-007	Configure alerts and alert policies in prism central, not prism element.
OPS-008	Utilize SMTP for alert transmission.
BCN-001	Utilize application- consistent snapshots when needed by the application
BCN-002	Place nearsync vm's in their own protection domain.
BCN-003	Configure snapshot schedules to be more frequent than the desired rpo
BCN-004	Configure snapshot schedules to retain the lowest number of snapshots while still meeting the retention policy.

© 2021 Nutanix, Inc. All rights reserved

BCN-005	Group applications together in unique protection domains. Keep number of vm's per protection domain as small as reasonably possible	
BCN-006	Application requires RPO=0 and RTO near zero	
BCN-007	Application requires zZero downtime DR avoidance solution	
BCN-008	Application requires RPO=0	
BCN-009	Provide RPO between 1 min and 15 min for the application	
BCN-010	Application requires RPO=>1h	
DCD-001	Decide on cluster licensing model for each ROBO (CBL or per-VM)	
DCD-002	Decide on cluster type used (if not ROBO per- VM)	
DCD-003	Decide whether to provide compute for ROBO failover	
DCD-004	Decide on ROBO failover compute over commit	
MULTI-NET-001	Calculate required storage replication bandwidth based on RPO.	
MULTI-NET-002	Calculate application- specific bandwidth.	
MULTI-NET-003	Place metro Availability and synchronous replication sites within the same region (Within 100km or less than 5msec RTT)	
MULTI-NET-004	Place metro witness within 200msec	
MULTI-NET-005	Do not use network segmentation for DR unless required. Routing segmentation is preferred over CVM segmentation	

MULTI-NET-006	Ensure redundancy of each network device. Track the complete network path between protected sites.	
MULTI-NET-007	Create and follow a redundancy test plan. Capture network components in plan.	
MULTI-NET-008	Ensure that the link between protected sites and a witness does not use the link between sites.	
MULTI-NET-009	For each workload, decide whether to keep the same IP addresses or change them during failover.	
OPS-001	Use X-PLAY for low-code/no-code operational automation.	
OPS-002	Monitor non-AOS VMs	
OPS-003	Discover the applications running on your infrastructure	
OPS-004	Monitor SQL Server from Prism	
AWS-001	Create separate subnets for user VMS	
AWS-002	Use AWS Direct Connect or create a second AWS cluster in a different AZ	
AWS-003	Protect your AWS cluster by replication back to on-prem.	
AWS-004	Plan your AWS VPC not to overlap with on-prem subnets	
AWS-005	Allow outbound internet access through an existing connection to the internet to allow the cluster portal to manage your AWS cluster	
AWS-006	Set AWS VPC limits to accommodate growth and auto-remediate features	
FLOW-001	Use Microsegmentation with App-Type Policy to segment applications	
FLOW-002	Use Microsegmentation with AD-Type Policy For VDI Deployments	

© 2021 Nutanix, Inc. All rights reserved

FILES-001	Dedicate Nutanix clusters to file services workloads
FILES-002	Use Nutanix clusters with hybrid storage for Files
FILES-003	Use a minimum of 3 FSVMs
FILES-004	Use an N+1 configuration of FSVMs to physical nodes
FILES-005	Allocate more than the minimum FSVM CPU and memory recommended for your planned workload
FILES-006	Use the minimum number of file server instances for your needs
FILES-007	Place the file server storage network on the same VLAN as the eth0 interface of the controller VMs.
FILES-008	Place the FSVM client network so it uses the shortest network route for the majority of file services clients
FILES-009	Use Files distributed shares
FILES-010	Use file server share-level inline compression
FILES-011	Use AOS container-level erasure coding (EC-X)
FILES-012	Do not enable AOS container-level deduplication
FILES-013	Use telescoping snapshot schedules
FILES-014	Use the minimum number of DR snapshots needed for your RPO requirements
OBJECTS-001	Provide internal and external networks
OBJECTS-002	Choosing between mixed and dedicated Objects clusters
OBJECTS-003	Hybrid or all-flash Nutanix clusters for Objects?
OBJECTS-004	Enable compression

© 2021 Nutanix, Inc. All rights reserved

OBJECTS-005	Number of worker nodes
OBJECTS-006	Multiple object stores
OBJECTS-007	Multi-cluster object stores
OBJECTS-008	Enable versioning?
OBJECTS-009	WORM (write-once, read-many)
OBJECTS-010	Object lock
OBJECTS-011	Lifecycle policy
OBJECTS-012	Streaming replication
BAC-001	Choose a backup vendor to use with Nutanix Mine
BAC-002	Use Nutanix Objects as an archival tier for backups
CALM-001	Small or Large Calm deployment
CALM-002	Upgrade to the most recent version of Calm
CALM-003	Create separate Calm instances for development and production
CALM-004	Define processes for Calm DR and backup/restore
CALM-005	Align Calm project structure and RBAC configuration
CALM-006	Design Calm environment to align with release methodology and geographic regions
CALM-007	Enable the Calm Policy Engine
CALM-008	Enable Showback via Nutanix Beam
CALM-009	Choose a method for self-service
KARBON-001	Minimize use of the admin kubeconfig
KARBON-002	Distribute production applications across AZs

© 2021 Nutanix, Inc. All rights reserved

KARBON-003	Use Kubernetes namespaces to segment users and applications rather than clusters	
KARBON-004	Leave enough space in a cluster to tolerate a worker node failure	
KARBON-005	Utilize consistent pod and service CIDRs	
KARBON-006	Upgrade the host OS as soon as possible	
DBAAS-001	Enable Database as a Service (DBaaS)	
DBAAS-002	Minimum hardware reqs for Nutanix clusters running Era	
DBAAS-003	Use dedicated Nutanix Clusters for Database	
DBAAS-004	Deploy Era management plane for HA	
DBAAS-005	Place Era on a separate management network	
DBAAS-006	Use the Era Platform license	
DBAAS-007	Use Active Directory for user, group, and service account access	
DBAAS-008	Use Active Directory groups to provide access to Era	
DBAAS-009	Provide an Era test environment	
DBAAS-010	Perform Era upgrades during service windows	
DBAAS-011	Use Era super admin and database admin roles	

© 2021 Nutanix, Inc. All rights reserved

List of Figures

Figure 1: Nutanix Hybrid and Multicloud Strategy.....7
Figure 2: Architecture Overview.....13
Figure 3: Hyperconverged Infrastructure.....14
Figure 4: A Nutanix Node.....15
Figure 5: Chassis Configurations.....16
Figure 6: Cluster Design.....18
Figure 7: Business Continuity Layer.....22
Figure 8: Designing for Security and Compliance.....23
Figure 9: Nutanix Security.....25
Figure 10: Regions.....28
Figure 11: Availability Zones.....28
Figure 12: Datacenters.....29
Figure 13: Single Location.....29
Figure 14: Multiple Locations.....30
Figure 15: Active - Passive.....30
Figure 16: Active - Active - Passive.....31
Figure 17: Datacenter ROBO.....32
Figure 18: Chain structure with Datacenter-ROBO-ROBO.....33
Figure 19: Separate Management Clusters.....39
Figure 20: Block and Pod Architecture.....45
Figure 21: Example VDI Building Block.....47
Figure 22: Planning the Technical Design.....56
Figure 23: Management Plane Failure Domain.....66

Figure 24: The Nutanix Cluster as a Failure Domain....67
Figure 25: Datacenter Rack and Server Room Failure Domains....68
Figure 26: Datacenter Building Failure Domain....70
Figure 27: Designing Datacenter Locations....70
Figure 28: Single Rack Single Workload Domain....75
Figure 29: Single Rack, Multiple Workload Domains....75
Figure 30: Multiple Racks, Single Workload Domain....77
Figure 31: Pure Layer 2 Design....83
Figure 32: Leaf Spine Topology....83
Figure 33: Core Aggregation Access Design....84
Figure 34: Layer 2 Network Design....89
Figure 35: Layer 3 Network Design....90
Figure 36: Layer 3 Network Design with Overlay....91
Figure 37: Logical View of Leaf Spine Scale....92
Figure 38: Leaf Switches Connected Using Link Aggregation....92
Figure 39: Rack Level View of Leaf Spine....93
Figure 40: Clusters Connecting to Different Aggregation Layers....95
Figure 41: AHV Networking....96
Figure 42: Relationship between the Components of the Nutanix Node....104
Figure 43: Example Host with 2 NUMA Nodes....105
Figure 44: Example of a Single NUMA Node....106
Figure 45: vCPU and Memory Assignment Spanning NUMA Boundaries....107
Figure 46: Best Performance vNUMA Configuration for Wide VMs....108
Figure 47: The Nutanix Control Plane....122
Figure 48: Small vs Large Prism Central VM Sizes....129
Figure 49: Scale Out Prism Central Deployment....130

Figure 50: vCenter Appliances..... 132
Figure 51: Life Cycle Manager..... 151
Figure 52: Foundation Pre-Configuration Service..... 153
Figure 53: Capacity Runway..... 155
Figure 54: Expansion Planning..... 156
Figure 55: Alert Email Configuration..... 160
Figure 56: General Guidance for Business Continuity..... 161
Figure 57: BCDR Workflow..... 162
Figure 58: Active – Active..... 164
Figure 59: Active - Passive..... 165
Figure 60: Disaster Recovery Decision Tree..... 176
Figure 61: Two AZ within the Same Region..... 180
Figure 62: Two AZ across Multiple Regions..... 181
Figure 63: Multiple AZ within a Single Region..... 182
Figure 64: Multiple AZ across Multiple Regions..... 183
Figure 65: Multiple AZ within a Single Region..... 184
Figure 66: Multiple AZ across Multiple Regions..... 185
Figure 67: Single Region..... 187
Figure 68: Multiple Regions..... 188
Figure 69: Latency Requirements..... 196
Figure 70: Network Routing..... 198
Figure 71: Network Redundancy..... 199
Figure 72: Single Piece of Conduit with Multiple WAN Links..... 201
Figure 73: Metro Witness Link Requirements..... 203
Figure 74: IT Ops Triangle..... 208
Figure 75: With Prism X-Play, a single trigger can precipitate multiple actions..... 209

Figure 76: Example Workflows Based on Various Triggers....212
Figure 77: Monitoring non-AOS VMs with Prism Ultimate....213
Figure 78: Monitoring SQL Server with Prism Ultimate....214
Figure 79: App Discovery....215
Figure 80: Nutanix Clusters on AWS....216
Figure 81: Nutanix Clusters on AWS Architecture....217
Figure 82: Example IP Address Assignment for an NCA Deployment....219
Figure 83: Security groups used by NCA....222
Figure 84: Effects of a Quarantine Policy....228
Figure 85: Effects of an Isolation Policy....229
Figure 86: Illustration of an App-Type Policy....229
Figure 87: Using an AD-Type Policy to Secure a VDI Environment....230
Figure 88: Using an AD-Type Policy to Secure a VDI Environment....231
Figure 89: Example of a Nutanix Files Instance....234
Figure 90: Three Separate Nutanix Files Instances Deployed on a Single Cluster....235
Figure 91: Relationship Between File Server VMs, CVMs, and Cluster Nodes....236
Figure 92: Distributed Shares are Distributed Across Multiple Volume Groups and FSVMs....237
Figure 93: Standard Shares are Assigned to a Single FSVM and a Single Volume Group....237
Figure 94: Nutanix Objects Deployments....248
Figure 95: Internal and External Networks....249
Figure 96: Lookup Hints....249
Figure 97: Mine Deployments Across Datacenters or Regions....263
Figure 98: Global Enterprise Calm Deployment....270
Figure 99: Self-Service with Calm....273
Figure 100: Karbon Provides a Single Control Plane for Multiple Kubernetes Clusters and Locations....275

Figure 101: A Karbon Development Cluster is not Highly Available....276

Figure 102: A Karbon Cluster for Production is Highly Available.... 276

Figure 103: An External Load Balancer Enables Karbon to Have More than Two Master Nodes....277

Figure 104: Nutanix Era DBaaS.... 284

Figure 105: Nutanix Era VMs and DBs....285

Figure 106: Era Management Layer in a Multi-Cluster Deployment....286

List of Tables

Table 1: Cloud Configurations....6
Table 2: Architectural Principles....6
Table 3: Design Objectives of this Document for Private Cloud Implementations....8
Table 4: Design Decision Example....11
Table 5: Number of Regions to be used....34
Table 6: Number of Availability Zones to be used....34
Table 7: Number of Datacenters to be used....34
Table 8: RPO and RTO requirements....34
Table 9: Choose a Hypervisor....37
Table 10: Management Cluster Architecture....40
Table 11: Mixed or Dedicated Workloads Per Cluster....42
Table 12: Nutanix Software Licensing Level....48
Table 13: Physical Node Vendor....51
Table 14: Node Models....55
Table 15: Software Versions....57
Table 16: Cluster Size Design Considerations....58
Table 17: Maximum Limits for Management Plane Software Components in AHV Deployments....63
Table 18: Min and Max Number of Nodes in a Single AHV Cluster....63
Table 19: Maximum Limits for Management Plane Software Components in VMware Deployments....63
Table 20: Min and Max Number of Nodes in a Single VMware Cluster....64
Table 21: Number of Clusters....64
Table 22: Management Plane Risk Mitigations....65

Table 23: Designing for Redundancy at the Cluster Level....67

Table 24: Datacenter Rack and Server Room Risk Mitigation....68

Table 25: Datacenter Building Risk Mitigation....69

Table 26: Two region implementation for DR....71

Table 27: Single region implementation for DR....71

Table 28: Long term retention backups are placed in Region 2....72

Table 29: Three availability zones in two regions....72

Table 30: Two availability zones in Region 1....73

Table 31: One availability zone in Region 1 to provide disaster recovery....73

Table 32: One availability zone in Region 2 hosts backups....74

Table 33: Pros and cons....78

Table 34: Workload Domains Spanning Single or Multiple Racks....78

Table 35: Examples of Recommended Switch Models....80

Table 36: Examples of Robo Or Smb Switch Models....81

Table 37: Switches that are not acceptable....81

Table 38: Using a Large Buffer Datacenter Switch at 10GBPS or Faster....82

Table 39: Using a Leaf-Spine Network Topology for New Environments....86

Table 40: Populate Each Rack with Two 10GBE or Faster TOR Switches....86

Table 41: Avoid Switch Stacking to Ensure Network Availability During Individual Device Failure....86

Table 42: Use No More Than Three Switches Between Two Nutanix Nodes in the Same Cluster....87

Table 43: Reduce Network Oversubscription to Achieve a Near-1:1 Ratio....87

Table 44: Configure CVM and Hypervisor VLAN on Server-facing Switch Ports....87

Table 45: Use Tagged VLANs on Switch Ports for All Guest Workloads....88

Table 46: Use a Layer 2 Network Design....91

Table 47: Connecting NICs to TOR Switches..... 99
Table 48: Using Bridges with Uplinks..... 99
Table 49: Using NICs of the same vendor within a bond..... 99
Table 50: Using VLANs to separate logical networks..... 100
Table 51: Load Balancing..... 100
Table 52: Using Standard 1500 Byte MTU..... 100
Table 53: Use Virtual Distributed Switch (VDS)..... 101
Table 54: Connecting NICs to the TOR Switch..... 102
Table 55: Using Virtual Switches and Uplinks..... 102
Table 56: Load Balancing..... 103
Table 57: Using Standard 1500 Byte MTU..... 103
Table 58: Running Non-NUMA-aware Applications on a VM..... 105
Table 59: Minimum CPU and Memory Values as of AOS 5.15..... 109
Table 60: Creating vDisks in ESXi..... 110
Table 61: SSD Capacity Considerations When Sizing a Cluster..... 113
Table 62: Mixing Nodes from Different Vendors in the Same Cluster..... 114
Table 63: Mixing NVMe and Hybrid SSD/HDD Nodes in the Same Cluster..... 114
Table 64: HDD to SSD Ratio..... 115
Table 65: Sizing for Redundancy..... 115
Table 66: Data Awareness and Fault Tolerance Levels Supported..... 117
Table 67: Fault Tolerance for Workloads/Clusters with Higher SLAs..... 118
Table 68: Expected Overhead for EC-X vs Standard RF2/RF3..... 118
Table 69: Enabling Inline Compression for All Containers..... 120
Table 70: Enabling/Disabling Deduplication..... 120
Table 71: Deploying Prism Central for Enhanced Cluster Management..... 122
Table 72: Deploy Prism Central in Each Region or Az Using Runbook Dr Automation..... 122

Table 73: Using VM-HA Guarantee....123
Table 74: Managing ESXi-based Nutanix Clusters....124
Table 75: Deploy a vCenter in Region or Az Using Runbook Dr Automation....125
Table 76: Enabling EVC Mode....125
Table 77: Enabling HA....126
Table 78: Enabling DRS....126
Table 79: Disabling VM PDL and APD Component Protection....126
Table 80: Configuring DAS.IGNOREINSUFFICIENTHBDATASTORE....126
Table 81: Disabling Automation Level for All CVMs....127
Table 82: Setting Amount of Reserved Resources in the Cluster....127
Table 83: Setting Host Isolation Response for VMs....127
Table 84: Setting Host Isolation Response for the CVM....127
Table 85: Disabling HA/DRS for Each Nutanix CVM....128
Table 86: Disabling SIOC....128
Table 87: Prism Central Recommendations....131
Table 88: vCenter Server Recommendations....132
Table 89: NTP Servers for Infrastructure Components....133
Table 90: DNS Servers for Infrastructure Layers....134
Table 91: Active Directory Authentication....137
Table 92: Using SSL/TLS Connection to Active Directory....137
Table 93: Using a Certificate Authority....138
Table 94: Nutanix Cluster Lockdown....138
Table 95: vSphere Cluster Lockdown....139
Table 96: CVM and Hypervisor AIDE....139
Table 97: Configuring SCMA....139
Table 98: Unused ESXi Services and Unused Firewall Ports....140

Table 99: Logging....141
Table 100: Including All Nutanix Modules in Logging....141
Table 101: Error Log Level for Nutanix Components....141
Table 102: ESXi Logging Level, Rotation, and File Size....142
Table 103: Protocols for Log Transport....142
Table 104: Using Port 514 for Logging....142
Table 105: VLAN for Traffic Separation....143
Table 106: Place CVM and Hypervisor on Same VLAN and Subnet....143
Table 107: Place Out-of-band Management on a Separate VLAN or Physical Network....144
Table 108: Default Prism Central Roles....144
Table 109: Prism Element Default Roles....146
Table 110: Use Least-Privilege Approach when Determining Access....146
Table 111: Align RBAC to Company Structure....146
Table 112: Do Not Use Storage Encryption....148
Table 113: Do Not Use a Key Management Server....148
Table 114: Overview of Different Foundation Software Options....153
Table 115: Deploying Prism Pro for Enhanced Cluster Management....154
Table 116: Reviewing Monthly Capacity Planning....156
Table 117: Performing Updates....157
Table 118: Utilizing the Current LTS Branch....158
Table 119: Upgrading to Maintenance and Current Patch Versions....158
Table 120: Maintaining a Pre-Production Environment....159
Table 121: Configuring Alerts and Alert Policies....159
Table 122: Utilizing SMTP for Alert Transmission....160
Table 123: Utilizing Application-consistent Snapshots....166
Table 124: Place Nearsync VMs in their Own Protection Domain....169

Table 125: Configure Snapshot Schedules to be More Frequent....170
Table 126: Configure Snapshot Schedules to Retain Lowest Number of Snapshots....170
Table 127: Schedule to Save Recovery Points....170
Table 128: Protection Domain 2-Site Support Matrix....171
Table 129: Protection Domain Multi-site Support Matrix....172
Table 130: Protection Policy 2 Site Support....172
Table 131: Protection Policy Multi-site Support Matrix....172
Table 132: Group Applications Together in Unique Protection Domains....173
Table 133: Application Requires RPO....177
Table 134: Application Requires Zero Downtime....178
Table 135: Application Requires RPO....178
Table 136: Provide RPO Between 1 and 15 Min....179
Table 137: Application Requires RPO....179
Table 138: Decide on Cluster Licensing Model....189
Table 139: Decide on Cluster Type Used....190
Table 140: Decide Whether to Provide Compute for ROBO Failover....190
Table 141: Decide on ROBO Failover Compute....190
Table 142: Required Networking Characteristics....193
Table 143: Bandwidth Needed for an RPO....194
Table 144: Calculate Required Storage Replication Bandwidth....195
Table 145: Calculate Application Specific Bandwidth....195
Table 146: Place Metro and Synchronous Replication Sites within the Same Region....196
Table 147: Place Metro Witness within 200msec....197
Table 148: Do Not Use Network Segmentation for DR....198
Table 149: Ensure Redundancy of Each Network Device....201
Table 150: Create and Follow a Redundancy Test Plan....202

Table 151: Ensure that the Link Between Protected Sites and a Witness does not Use the Link Between Sites....203

Table 152: For Each Workload Decide Whether to Maintain or Change....205

Table 153: Prism Ops Design Objectives....207

Table 154: Leverage X-Play Playbooks....211

Table 155: Monitor Non-AOS VMs....212

Table 156: Discover Applications with Prism Ultimate....213

Table 157: Monitor Microsoft SQL Server....213

Table 158: Nutanix Clusters Design Objectives....215

Table 159: Supported EC2 Bare-Metal Instances....220

Table 160: Separate Subnets....223

Table 161: Create Availability Zones....224

Table 162: Use Replication....224

Table 163: Don't Overlap Subnets....225

Table 164: Allow Outbound Internet Access....225

Table 165: Plan VPC Limits for Growth and Availability....226

Table 166: Flow Design Objectives....226

Table 167: Flow Policy Types....228

Table 168: Use Microsegmentation to Isolate Applications....232

Table 169: Use AD-Type Policy to Isolate VDI....232

Table 170: Nutanix Files Design Objectives....233

Table 171: Mixed or Dedicated Environments....238

Table 172: Choose Hybrid Clusters....239

Table 173: Minimum FSVMs....240

Table 174: N+1 Configuration....240

Table 175: CPU and Memory Allocation....241

Table 176: Minimum Number of File Servers....242
Table 177: File Server Storage Network....242
Table 178: Shortest Route Between Client and Server....242
Table 179: Utilize Distributed Shares....243
Table 180: Compression....244
Table 181: Erasure Coding (EC-X)....244
Table 182: Deduplication....244
Table 183: Scheduling Snapshots....245
Table 184: Disaster Recovery....245
Table 185: Disaster Recovery....246
Table 186: Design Objectives....246
Table 187: An Object Store Requires Internal and External Networks....250
Table 188: Mixed or Dedicated Objects Clusters....250
Table 189: Hybrid or All Flash Clusters....251
Table 190: Enable Compression....252
Table 191: Number of Worker Nodes....253
Table 192: Multiple Object Stores....254
Table 193: Multi-Cluster Object Stores....254
Table 194: Enabling Versions....255
Table 195: WORM....256
Table 196: Object Lock....257
Table 197: Lifecycle Policies....257
Table 198: Streaming Replication....258
Table 199: Nutanix Mine Design Objectives....258
Table 200: Nutanix Mine Configurations....259
Table 201: Choose a Backup Vendor....261

Table 202: Utilize Objects for Archival....262
Table 203: Nutanix Calm Design Objectives....263
Table 204: Choose a Small or Large Calm Deployment....265
Table 205: Upgrade to Most Recent Calm Version....266
Table 206: Deploy Calm for Development and Production....267
Table 207: Calm for DR and Backup....268
Table 208: Calm and RBAC Configurations....269
Table 209: Calm Releases and Geographic Regions....270
Table 210: Calm Policy Engine....271
Table 211: Showback with Nutanix Beam....272
Table 212: Self-Service with Calm....273
Table 213: Nutanix Karbon Design Objectives....273
Table 214: Minimize Admin Kubeconfig Use....279
Table 215: Use Multiple Availability Zones for Production....280
Table 216: Use Namespaces to Segment Users and Apps....280
Table 217: Leave Room for a Worker Node to Fail....280
Table 218: Use Consistent CIDRs....281
Table 219: Keep the Host OS Updated....281
Table 220: Nutanix Era Design Objectives....282
Table 221: Era to Enable DBaaS....287
Table 222: Era Hardware Requirements....288
Table 223: Dedicated DB Clusters....288
Table 224: HA Management Plane....289
Table 225: Management Network....289
Table 226: Era Platform License....290
Table 227: Active Directory Integration....290

Table 228: Active Directory Access....291
Table 229: Era Test Environment....291
Table 230: Service Windows....291
Table 231: Era Roles....292
Table 232: Design Decisions....294

Flow 6.1

Flow Microsegmentation Guide

February 24, 2022

Contents

1. Security Policies....4

2. Types of Policies....5

3. Security Policy Model....7

4. Requirements....9

5. Enabling Microsegmentation....10
   - Disabling Microsegmentation....10

6. Built-in Categories for Security Policies....12

7. Services....14
   - Creating a Service....14

8. Addresses....16
   - Creating an Address....16

9. Application Security Policy Configuration....18
   - Creating an Application Security Policy....18
   - Modifying an Application Security Policy....24
   - Applying an Application Security Policy....24
   - Monitoring an Application Security Policy (Visualizing Network Flows)....24
   - Deleting an Application Security Policy....26

10. Isolation Environment Policy Configuration....28
   - Creating an Isolation Environment Policy....29
   - Modifying an Isolation Environment Policy....31
   - Applying an Isolation Environment Policy....32
   - Monitoring an Isolation Environment Policy (Visualizing Network Flows)....32
   - Deleting an Isolation Environment Policy....33

11. Quarantine Policy Configuration....34
   - Configuring the Quarantine Policy....34
   - Quarantining a VM....36
   - Removing a VM from the Quarantine....36

12. VDI Policy Configuration..... 37
Creating a VDI Policy..... 37
Default VDI Policy..... 40
Configuring Active Directory Domain Services..... 41
Modifying the VDI Policy..... 43
Applying the VDI Policy..... 43
Monitoring the VDI Policy..... 44
Deleting the VDI Policy..... 44

13. Applying Filtering and Grouping to a Security Policy..... 45

14. Exporting and Importing Security Policies..... 49

1

SECURITY POLICIES

Traditional data centers use firewalls to implement security checks at the perimeter—the points at which traffic enters and leaves the data center network. Such perimeter firewalls are effective at protecting the network from external threats. However, they offer no protection against threats that originate from within the data center and spread laterally, from one compromised machine to another.

The problem is compounded by virtualized workloads changing their network configurations and hosts as they start, stop, and migrate frequently. For example, IP addresses and MAC addresses can change as applications are shut down on one host and started on another. Manual enforcement of security policies through traditional firewalls, which rely on network configurations to inspect traffic, cannot keep up with these frequent changes and are error-prone.

Network-centric security policies also require the involvement of network security teams that have intimate knowledge of network configuration in terms of VLANs, subnets, and other network entities.

Nutanix Flow includes a policy-driven security framework that inspects traffic within the data center. The framework works as follows:

- Security policies inspect traffic that originates and terminates within a data center and help eliminate the need for additional firewalls within the data center.
- The framework uses a workload-centric approach instead of a network-centric approach. Therefore, it can scrutinize traffic to and from VMs no matter how their network configurations change and where they reside in the data center. The workload-centric, network-agnostic approach also enables the virtualization team to implement these security policies without having to rely on network security teams.
- Security policies are applied to categories (a logical grouping of VMs) and not to the VMs themselves. Therefore, it does not matter how many VMs are started up in a given category. Traffic associated with the VMs in a category is secured without administrative intervention, at any scale.
- Prism Central offers a visualization-based approach to configuring policies and monitoring the traffic to which a given policy applies.
- Using Prism Central, you can configure syslog monitoring by forwarding Flow logs to an external syslog server. See Configuring Syslog Monitoring in the *Prism Central Guide* for details.

> Note: Nutanix Flow supports only AHV hypervisor; security policies can not be applied to VMs running on other hypervisors.

2

TYPES OF POLICIES

The types of policies in Prism Central and their use cases are described here.

Table 1: Types of Policies

Policy Type	Use Case
Application Security Policy	Use an application security policy when you want to secure an application by specifying allowed traffic sources and destinations. This method of securing an application is typically called *application ring fencing*. For example, use an application security policy when you want to allow only those VMs in the categories `department: engineering` and `department: customersupport` (the allowed sources) to communicate with an issue tracking tool in the category `AppType: IssueTracker` (the secured application), and you want the issue tracking tool to be able to send traffic only to an integrated customer relationship management application in the category `AppType: CRM`. The secured application itself can be divided into tiers by the use of categories (the built-in AppTier category). For example, you can divide the issue tracking tool into web, application, and database tiers and configure tier-to-tier rules. For more information, see Application Security Policy Configuration on page 18.

Policy Type	Use Case
Isolation Environment Policy	Use an isolation environment policy when you want to block all traffic, regardless of direction, between two groups of VMs identified by their category. VMs within a group can communicate with each other. For example, use an isolation environment policy when you want to block all traffic between VMs in the category `Environment: sandbox` and VMs in the category `Environment: production`, and you want to allow all the VMs within each of those categories to communicate with each other. For more information, see Isolation Environment Policy Configuration on page 28.
Quarantine Policy	Use a quarantine policy when you want to isolate a compromised or infected VM and optionally want to subject it to forensics. For more information, see Quarantine Policy Configuration on page 34.
VDI Policy	Use a VDI policy when you want to secure your VDI environment. For more information, see VDI Policy Configuration on page 37

3

SECURITY POLICY MODEL

Application-centricity

The security policy model uses an application-centric policy language instead of the more complex, traditional network-centric policy language. Configuring an application security policy involves specifying which VMs belong to the application you want to protect and then identifying the entities or networks, in the inbound and outbound directions, with which you want to allow communication.

All the entities in an application security policy are identified by the categories to which they belong and not by their IP address, VLAN, or other network attributes. After a VM is associated with a category and the category is specified in a security policy, traffic associated with the VM is monitored even if it migrates to another network or changes its IP address.

The default options for allowing traffic on the inbound and outbound directions are also inherently application centric. For application security policies, the default option for inbound traffic is **Allowed List**, which means that **Allowed List** is usually the recommended option for inbound traffic. The default option can be changed to **Allow All** traffic. The default option in the outbound direction allows the application to send traffic to all destinations, but you can configure a destination **Allowed List** if desired.

For forensic quarantine policies, the default option in both directions is **Allowed List**, but you can **Allow All** traffic in both directions. For strict quarantine policies, no traffic is allowed in either direction.

All the VMs within a category can communicate with each other. For example, in a tiered application, regardless of how you configure tier-to-tier rules, the VMs within a given tier can communicate with each other.

Whitelist-Based Policy Expression

An application security policy is expressed in terms of the categories and subnets with which you want the application to communicate and therefore, by extension, the traffic you want to allow. A more granular policy expression can be achieved by specifying which protocols and ports can be used for communication.

Any category or subnet that is not in the allowed list is blocked. You cannot specify the categories and subnets you want to block because the number of such entities are typically much larger and grow at a much higher rate than the categories and subnets with which an application should be allowed to communicate. Expressing a policy in terms of allowed traffic results in a smaller, tighter policy configuration that can be modified, monitored, and controlled more easily.

Enforcement Modes

All policies, whether associated with securing an application, isolating environments, or quarantining VMs, can be run in the following modes:

Apply Mode

Blocks all traffic that is not allowed by the policy.

Monitor Mode

Allows all traffic, including traffic that is not allowed by the policy. This mode enables you to visualize both allowed and disallowed traffic and fine-tune the policy before applying it.

You can switch a policy between these two modes as many times as you want.

Automated Enforcement

A policy uses categories to identify the VMs to which it must apply. This model allows the automatic enforcement of a policy to VMs regardless of their number and network attributes. Connectivity between Prism Central and a registered AHV cluster is required only when creating and modifying policies, or when changing the mode of operation (applied or monitoring) of a policy. Policies are applied to the VMs in a cluster even if the cluster temporarily loses network connectivity with the Prism Central instance with which it is registered. New policies and changes are applied to the cluster when connectivity is restored.

Priorities Between Policies

Prism Central does not provide a way for you to specify priorities between policies of a single type. For example, you cannot prioritize one security policy over another. There is no limit to the number of inbound and outbound rules that you can add to a security policy, allowing you to define all of an application's security requirements in a single policy. This makes priorities between policies unnecessary.

However, priorities exist between the different policy types. Quarantine policies have the highest priority followed by isolation environment policies, and application security policies, in that order. The VDI Policy takes the last precedence, for example, if an application security is protecting a VM, it cannot simultaneously be protected with the VDI policy.

Isolation environment rules take precedence over application security rules, so make sure that isolation environment policies and application security policies are not in conflict. An isolation environment rule and an application security rule are said to be in conflict if they apply to the same traffic (a scenario that is encountered when VMs in one of the categories in the isolation environment send traffic to an application in the other category, and some or all of that traffic is either allowed or disallowed by the application security policy). The effect that an isolation environment policy has on a conflicting application security policy depends on the mode in which the isolation environment policy is deployed, and is as follows:

- If the isolation environment policy is in the applied mode, it blocks all traffic to the application, including the traffic that is allowed by the application security policy.
- If the isolation environment policy is in the monitoring mode, it allows all traffic to the application, including any traffic that is disallowed by the application security policy.

4

REQUIREMENTS

The Security Policies feature has the following requirements:

- The feature is supported only on AHV clusters running AOS 5.6 or later and AHV version 20170830.115 or later.
- The Prism Central instance must be hosted on one of the AHV clusters registered with it. The AHV cluster that hosts the Prism Central instance must be running AOS 5.6 or later.
- The host must have at least 1 GB of additional memory for each Prism Central VM hosted on it.
- If you are running a Prism Central scale-out instance, all the VMs in the Prism Central cluster must be powered on.
- The AHV hosts must be allowed to communicate with the Prism Central VMs over TCP port 9446. Keeping the port open enables the hosts to send the Prism Central VMs connection tracking data. Prism Central uses that data to show network flows.

Caution:

- When Flow is enabled, a Kafka container is automatically created on the cluster where Prism Central is hosted. The container is used to store data that is required for flow visualization to work and must not be deleted.
- Cross cluster live migration of guest VMs that are part of Flow security policy is not supported.
- Security Policies are not supported for VMs that are on the advanced networking stack. An alert is raised for VMs that are part of both VPC and Flow policy, and Flow policies are not enforced for VMs on VPCs.
- Overlapping or conflicting policy configuration is not supported and might cause unintended interruption of network services.

5

ENABLING MICROSEGMENTATION

Microsegmentation is disabled by default. Before you can configure and use application security policies, isolation environment policies, and quarantine policies, you must enable the feature. The feature requires a Flow license. If you have not installed a Flow license, you can try the feature for a period of 60 days. After this period expires, you will be required to install the license to continue using the feature.

Before you begin

Ensure that you meet Microsegmentation requirements.

About this task

To enable microsegmentation, do the following:

Procedure

1. Log on to the Prism Central web console.
2. Click the collapse menu ("hamburger") button on the left of the main menu and then select **Prism Central Settings** to display the Settings page.
3. Click **Microsegmentation** from the Settings menu (on the left).
   The **Enable Microsegmentation** dialog box is displayed.
4. To determine whether the registered clusters are capable of supporting microsegmentation, do the following:
   a. Click **View Cluster Capability**, and then review the results of the capability checks that Prism Central performed on the registered clusters.
   b. Click **Back**.
5. Select the **Enable Microsegmentation** check box.
6. Click **OK**.

Disabling Microsegmentation

Prism Central web console provides you the ability to disable the microsegmentation feature.

About this task

To disable microsegmentation, do the following:

Procedure

1. Log on to the Prism Central web console.

2. Click the gear icon in the main menu and then select **Microsegmentation** in the **Settings** page.

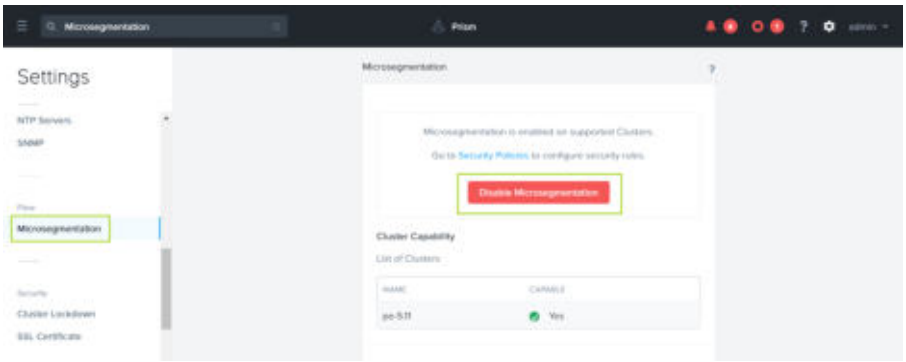


Figure 1: Settings Page - Disabling Microsegmentation

3. Click **Disable Microsegmentation**.
   A confirmation message appears.

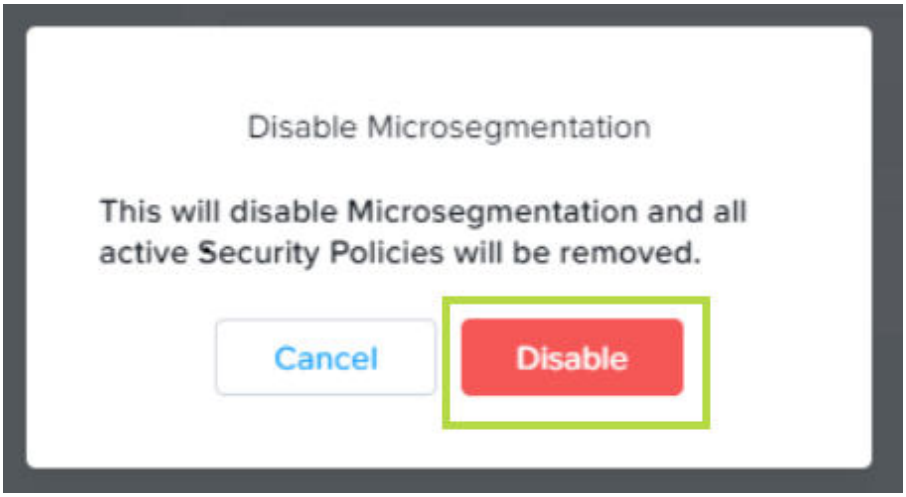


Figure 2: Microsegmentation - Confirmation message

4. Click **Disable** to confirm disabling the microsegmentation feature.

6

BUILT-IN CATEGORIES FOR SECURITY POLICIES

Prism Central includes built-in categories that you can use in application security policies and isolation policies. It also includes a built-in category for quarantining VMs.

Table 2: Built-In Categories

Category	Description
AppTier	Add values for the tiers in your application (such as web, application_logic, and database) to this category and use the values to divide the application into tiers when configuring a security policy.
AppType	Associate the VMs in your application with the appropriate built-in application type such as Exchange and Apache_Spark. You can also update the category to add values for applications not listed in this category.
Environment	Add values for environments that you want to isolate from each other and then associate VMs with the values.
Quarantine	Add a VM to this category when you want to quarantine the VM. You cannot modify this category. The category has the following values: **Strict** Use this value when you want to block all inbound and outbound traffic. **Forensic** Use this value when you want to block all inbound and outbound traffic except the traffic to and from categories that contain forensic tools.

Category	Description
ADGroup	This category is managed by ID Based Security (ID Firewall). Each ADGroup value represents an imported group from Active Directory. To add or remove values to use in Flow policies use the **ID Based Security** configuration page (**Prism Central Settings** > **Flow** > **ID Based Security**). The category values may be used in VDI policies, see VDI Policy Configuration on page 37 for details.
ADGroup:Default	This category is applied to the VDI VMs of the AD group when the VM inclusion criteria is set and allows you to apply a default set of rules for the VDI VMs (without the requirement of user logons).

7

SERVICES

Service is a group of protocol-port combination. You can use any of the default services or create a custom service. The ability to use the service entities in the policy creation workflow reduces any manual configuration error and enables reusability of available entities.

- To create or update a custom service, see Creating a Service on page 14.
- To view the list of available services (built-in and custom services), go to **Policies** > **Security** > **Services**.

Creating a Service

About this task

To create a custom service, do the following.

Procedure

1. Log on to the Prism Central web console.
2. Click the collapse menu ("hamburger") button on the left of the main menu and go to **Policies** > **Security** > **Services**.

3. Click **Create Service**.

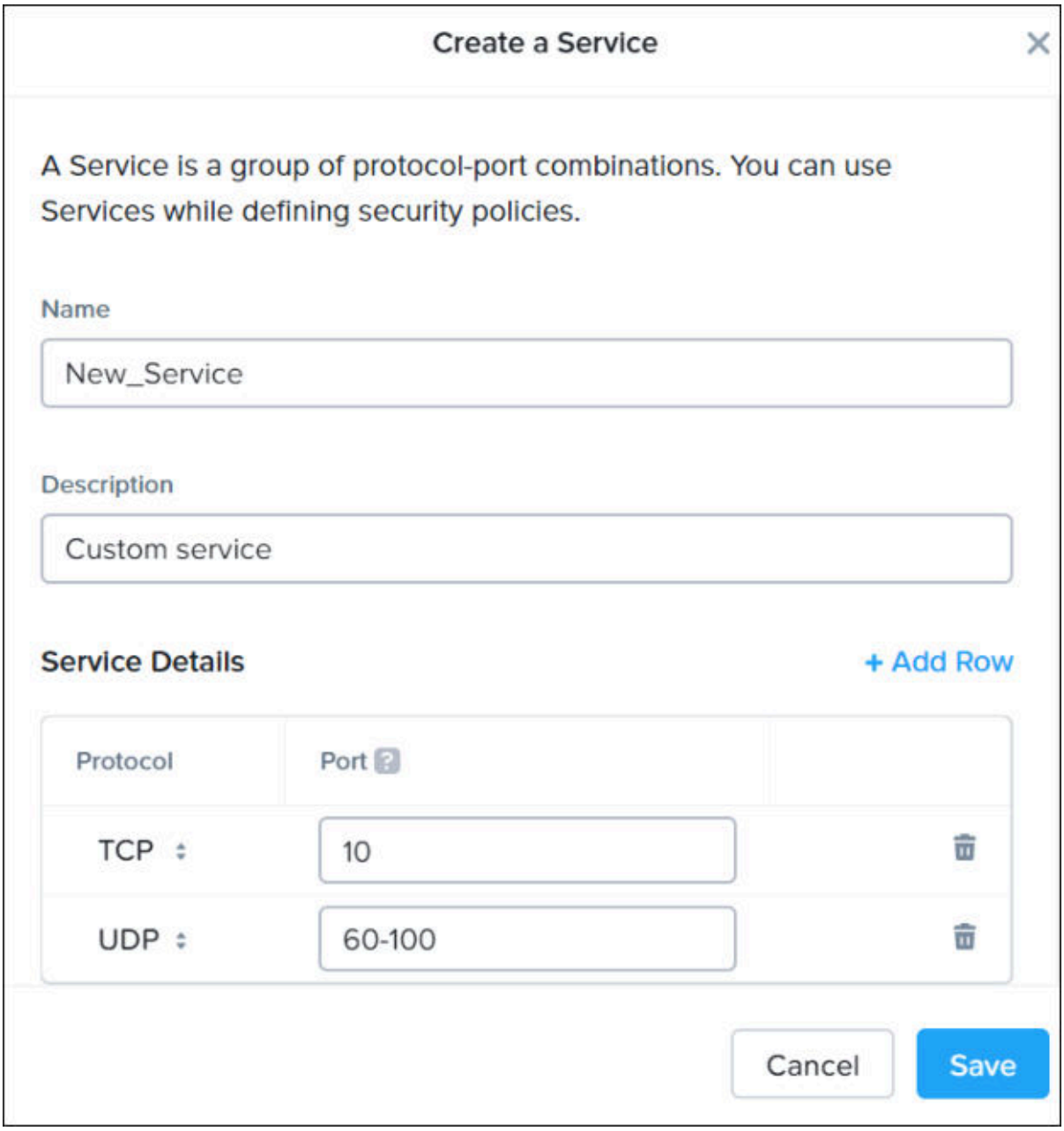


Figure 3: Create Service Tab

4. Enter a name and description for the service.
5. Select the **Protocol** from the drop-down menu and enter the port number or port range in the **Port** field.

   You can add multiple protocol-port combinations in a single service. To add more protocol-port combination, click **Add Row** and specify the required values.
6. Click **Save** to save the service.

8

ADDRESSES

Address is a way to group one or many IP addresses or ranges. You can create an address entity and use that address entity while creating policies. The ability to use the addresses in the policy creation work flow reduces any manual configuration error and enables reusability of available entities.

- To create or update an Address, see Creating an Address on page 16.
- To view the list of available services (built-in and custom services), go to **Policies** > **Security** > **Address**.

Creating an Address

About this task

To create an Address, do the following.

Procedure

1. Log on to the Prism Central web console.
2. Click the collapse menu ("hamburger") button on the left of the main menu and go to **Policies** > **Security** > **Addresses**.

3. Click **Create Address**.

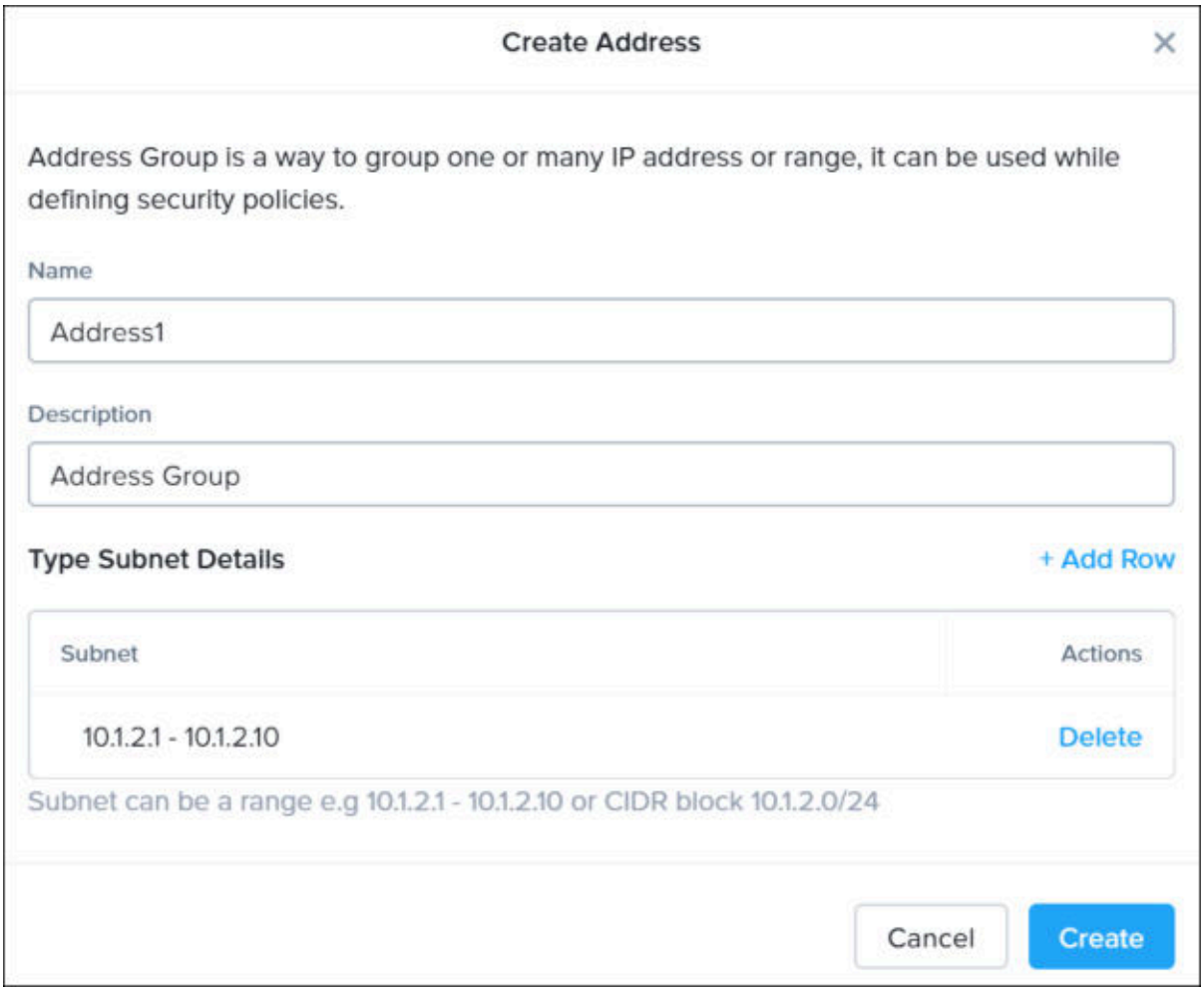


Figure 4: Create Address Tab

4. Enter a name and description for the address.
5. Enter the IP address or a IP range in the **Subnet** field.
   You can add multiple subnets in a single address entity. To add more subnets, click **Add Row** and specify the required values.
6. Click **Save** to save the service.

9

APPLICATION SECURITY POLICY CONFIGURATION

Creating an Application Security Policy

Before you begin

- Create the categories you need and associate the VMs that you want to protect with those categories. You might be required to create categories for the following purposes. Some categories or category values are required while others are optional:

  - Every security policy must be associated with a value in the AppType category, so make sure that you update the AppType category with appropriate values if the built-in values do not work for you. For information about this category and its values, see "Category Details View" in the *Prism Central Guide*.
  - If you need to apply the policy to an application in a specific environment (for example, development, test, or production) or an application at a specific location, create the category you need and apply it to the application. Prism Central includes a built-in Environment category that you can use or update with values of your own. You can also create your own categories.
  - If you want to specify categories for traffic sources and destinations instead of allowing all inbound and outbound traffic, create those categories and apply them to the traffic sources and destinations.
  - If you want to divide the application into tiers in a security policy, add tiers to the AppTier category. The AppTier category has a built-in `default` value, but you can update the category to add values of your choice.

  For information about categories and their values, see "Category Management" in the *Prism Central Guide*.

- Security policy configuration might require more time than the default session timeout allows you. You might want to increase the session timeout so that you do not lose a configuration that is left unattended while you perform associated tasks such as referring to this documentation. For more information, see "Modifying UI Settings" in the *Prism Central Guide*.

About this task

To secure an application, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), click **Create Security Policy**, and then click **Secure an Application**.
   The **Create App Security Policy** page is displayed.

2. On the **Define Policy** tab, do the following in the indicated fields, and then click **Next**:

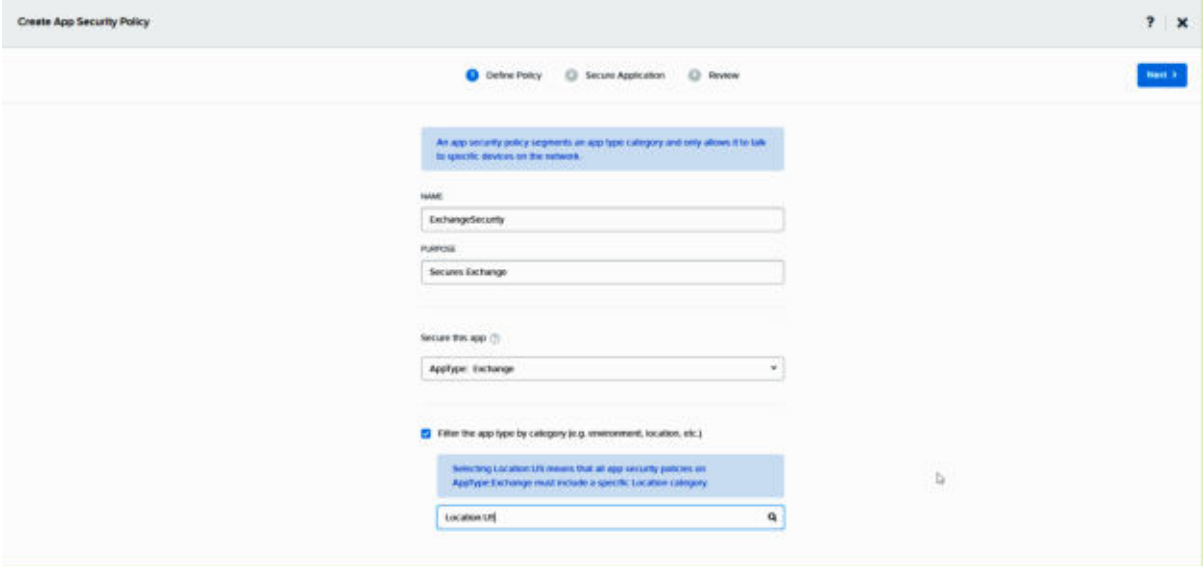


Figure 5: Define Policy Tab

a. **Name**: Enter a name for the security policy.

b. **Purpose**: Describe the purpose of the security policy.

c. **Secure This App**: Select the type of application that you want to secure.

   The **Secure This App** list displays available values in the AppType category. It uses the format `AppType: value`, where value represents a type of application. Every application that you want Prism Central to secure must be associated with a value from the built-in AppType category. The AppType category includes values for frequently encountered applications, such as Exchange and Hadoop. The AppType category also includes a built-in `default` value that you can use if your application cannot be associated with one of the other built-in values. You can also update the `AppType` category to add a value of your choice. For information about categories and their values, see *Category Management*.

d. If you want to filter the VMs by an additional category, select **Filter the app type by category**, and then enter the name of the category in the text box that is displayed.

   This option enables you to apply the policy to an additional category. For example, if you are configuring a policy for an application in the category `AppType: Exchange`, this option enables you to further restrict the policy to specific locations (such as `Location: US` and `Location: EU`) or environments (such as `Environment: Production`, `Environment: Development`, and `Environment: Test`).

e. Optionally, in the **Advanced Configuration** section, select the **Allow** radio button to allow **IPv6 traffic**. The policy rules apply to IPv4 traffic only and all IPv6 traffic is blocked by default.

   > Note: If you choose to block IPv6 traffic, the IPv6 traffic remains blocked even in the monitoring mode.

f. Optionally, click the toggle button against **Policy Hit Logs** to log traffic flow hits on the policy rules.

   You can configure syslog monitoring for the policy hit logs for Flow, see Configuring Syslog Monitoring in the *Prism Central Guide* for details.

   > Note: Policy hit logs are not generated if both source and destination are in inbound or outbound category.

3. In the **Securing an App** dialog box, review the schematic that illustrates the flow of traffic through a secured app, and then click **OK, Got it!**
   The **Secure Application** tab is displayed. The schematic on this tab can be divided into three areas of configuration: the **Inbound** side, (for adding traffic source allowlist), the application at the center (for configuring inbound, outbound, and tier-to-tier rules), and the **Outbound** side (for adding traffic destination allowlist).

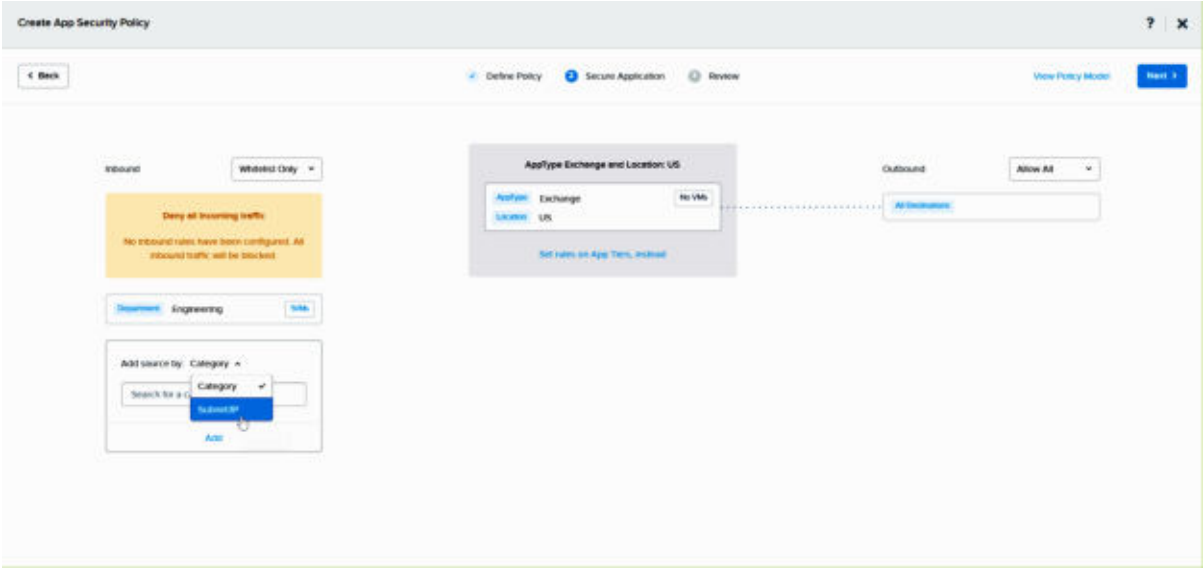


   Figure 6: Secure Application Tab

4. On the **Secure Application** tab, do the following, and then click **Next**:

   a. On the application at the center of the tab, do the following in the indicated fields:

      - If you want to divide the application into tiers (such as a web tier, an application tier, and a database tier) and configure tier-to-tier rules, first configure the application as described in this step, and then configure inbound and outbound rules. This approach ensures that the individual tiers are available when you want to configure inbound and outbound rules at the tier level. Skip this step if you want to treat the application as a single entity in the security policy.

        To divide your application into tiers and create tier-to-tier rules, do the following:

        1. On the application, click **Set Rules on App Tiers, Instead**.

           > Note: After you click **Set Rules on App Tiers, Instead**, the link text, **Set rules on the whole app, instead**, is displayed in its place. Click **Set rules on the whole app, instead** if you want to discard the tiered configuration and return to configuring rules on the application as a whole.

2. Click **Add Tier**, and then select a tier.

   Repeat this step to add as many tiers as you require. The following figure shows an application with a web tier, an application tier, and a database tier:

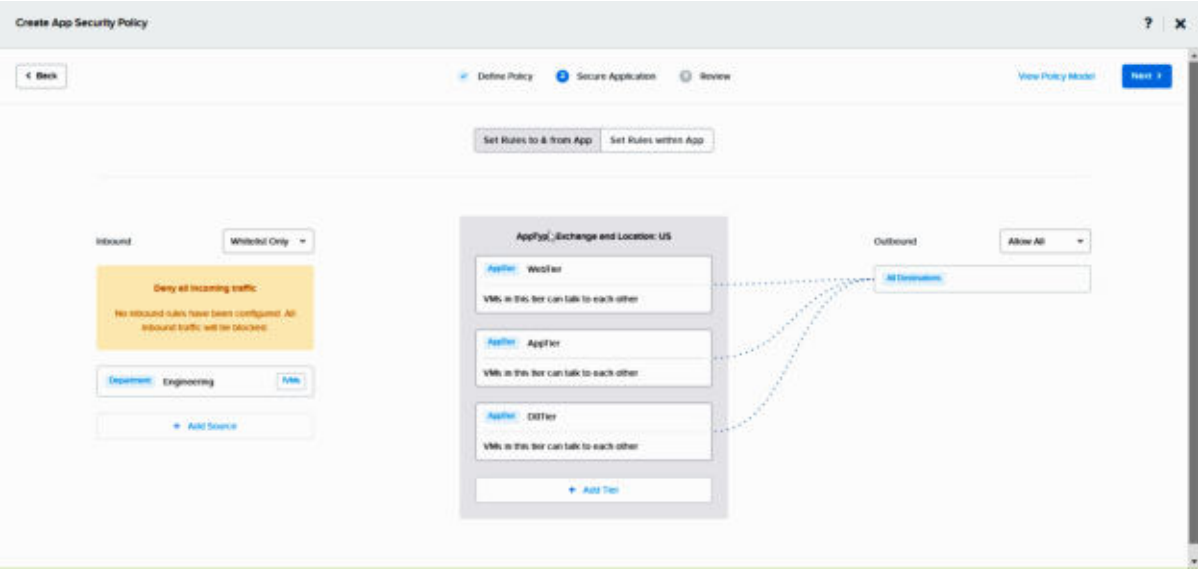


   Figure 7: Tiered Application

3. To delete a tier, pause over the tier you want to delete and click the delete button that is displayed.
4. Click **Set Rules Within App**.

   > Note: When configuring tier-to-tier rules, two modes are made available to you through the buttons **Set Rules to & from App** and **Set Rules Within App**. The **Set Rules to & from App** option enables you to add application tiers and to specify allowed inbound and outbound traffic. The **Set Rules Within App** option enables you to specify tier-to-tier rules within the application. These buttons enable you to switch between the two modes.

5. Click each tier in the application and click **Yes** or **No** to specify whether or not you want to allow the VMs in the tier to communicate with each other.

6. Configure a tier-to-tier rule as follows:
   1. Click the source tier (for example, if the tiers are WebTier and AppTier and you want to configure a tier-to-tier rule from WebTier to AppTier, click the source tier, WebTier).
   2. Click the plus sign that is displayed on the destination tier (in this example, click the destination tier, AppTier). The **Create Tier to Tier Rule** dialog box
   3. Enter a description for the rule.

      Note: The policy rule description is captured in the policy hitlog data.

      - Policy hitlog must be enabled
      - Rule description is added to the hitlog only for allowed traffic

   4. In **Service Details**, click **Allow all traffic** to allow all types of traffic or click **Select a service** to choose any default or custom service.
   5. Click **Save**.

   Configure tier-to-tier rules for as many source and destination tiers as you want.

b. To add traffic sources, on the **Inbound** side of the **Secure Application** tab, do the following:

- From the drop-down list, select one of the following options:
  - **Allow All**: Allows traffic from all sources.
  - **Whitelist Only**: Allows traffic only if the traffic originates from entities on the security policy's source allowlist. This option is the default option. If this option is selected, you must also configure the source allowlist by clicking **Add Source**.
- Click **Add Source**, and then do the following:
  1. Select one of the following options from the drop-down list:
     - **Category**: Allows traffic only if that traffic originates from entities that are in the selected category.
     - **Subnet/IP**: Allows traffic only if that traffic originates from entities that are in the selected subnet.
     - **Addresses**: Allows traffic only if the traffic originates from the entities that are in the selected address.
  2. Enter the value (category name or subnet) in the text box, and then click **Add**.

     When entering the name of a category, a list of matching names is displayed, and you can select the name you want to specify. The subnet mask must be specified in the CIDR format.
  3. To add another category or subnet, click **Add Source**. Add as many categories or subnets as you want to allow.

  Each entry in this list represents a stream of inbound traffic.

c. To add traffic destinations, on the **Outbound** side, do the following:

- From the drop-down list, select one of the following options:
  - **Allow All**: Allows traffic to all destinations. This option is the default option.

- **Whitelist Only**: Allows traffic only if the traffic is destined for entities on the security policy's destination allowlist. If this option is selected, you must also configure the destination allowlist by clicking **Add Destination**.

- Click **Add Destination**, and then do the following:

  1. Select one of the following options from the drop-down list:

     - **Category**: Allows traffic only if that traffic is destined for entities in the selected category.
     - **Subnet/IP**: Allows traffic only if that traffic is destined for entities in the selected subnet.
     - **Addresses**: Allows traffic only if the traffic originates from the entities that are in the selected address.
  2. Enter the value (category name or subnet) in the text box, and then click **Add**.

     When entering the name of a category, a list of matching names is displayed, and you can select the name you want to specify. The subnet mask must be specified in the CIDR format.
  3. To add another category or subnet, click **Add Destination**. Add as many categories or subnets as you want to allow.

  Each entry in this list represents a stream of outbound traffic.

- To specify the protocols that you want to allow from each stream of inbound and outbound traffic, do the following:

  1. If you added application tiers and configured tier-to-tier rules, first click **Set Rules to & from App**.
  2. Click the traffic source or traffic destination (a category or subnet if you have configured a allowlist or **All Sources** if you have chosen to allow all sources) for which you want to create a rule.
  3. Click the plus icon that appears on the application (if you are treating the application as a single entity) or application tier (if you have divided the application into tiers). The **Create Inbound Rule** or **Create Outbound Rule** dialog box appears.
  4. Enter a description for the rule.
  5. In **Service Details**, click **Allow all traffic** to allow all types of traffic or click **Select a service** to choose any default or custom service.
  6. Click **Save**.

After you configure a rule, a dotted line appears between the two endpoints of the rule. Point to the dotted line to show the list of ports that the rule allows.

5. On the **Review** tab, review the security policy configuration, and then do one of the following:

   - If you want to apply the configuration, click **Apply Now**.

     Applying a security policy enforces the security policy on the application, and traffic from entities that are not defined as sources in the policy is blocked.

   - If you want to save the configuration and monitor how the security policy works, click **Save and Monitor**.

     When a policy is in the monitoring state, the application continues to receive all traffic, but disallowed traffic is highlighted on the monitoring page. Traffic is not blocked until the policy is applied.

     > Note: A policy that you have chosen to save and monitor can be applied from the policy update page.

Modifying an Application Security Policy

About this task

To modify a security policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the policy that you want to modify, click **Actions**, and then click **Update**.

2. Make the changes you want and then apply or save and monitor the policy.

   The update options are the same as those for creating a policy. For information about the options, see Creating an Application Security Policy on page 18.

Applying an Application Security Policy

Applying a security policy enforces the security policy on the application, and any traffic from sources that are not allowed is blocked.

About this task

To apply a security policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the policy that you want to apply, click **Actions**, and then click **Apply**.

2. Confirm by typing **Apply** in the dialog box, and then click **OK**.

Monitoring an Application Security Policy (Visualizing Network Flows)

About this task

When a policy is in the monitoring state, the application continues to receive all traffic, but disallowed traffic is highlighted on the monitoring page. Traffic is not blocked until the policy is applied.

To monitor a security policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the policy that you want to monitor, click **Actions**, and then click **Monitor**.

2. Confirm by typing **Monitor** in the dialog box, and then click **OK**.
   Allowed network flows and disallowed network flows are shown on the monitoring page, as shown in the following figure. Allowed flows are depicted with a blue dotted line and disallowed network flows are depicted with a red dotted line:

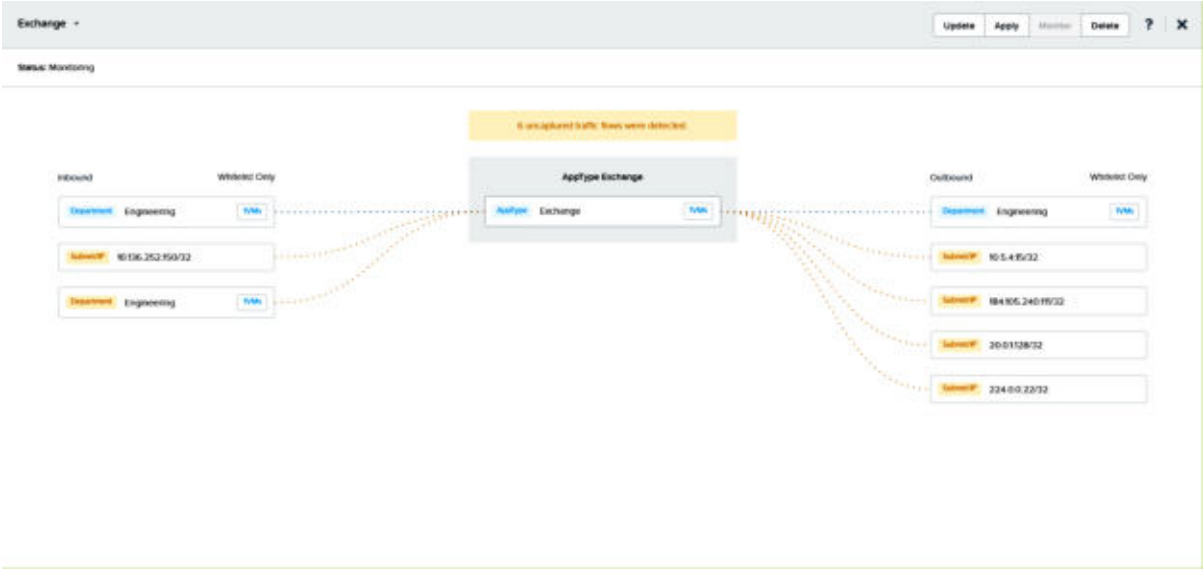


Figure 8: Monitoring Page for an Application Security Policy

3. To show a preview of the network flow in a tooltip, pause over the dotted line that depicts the network flow in the diagram.
   A tooltip similar to the following is displayed. The tooltip shows a graph for each connection:

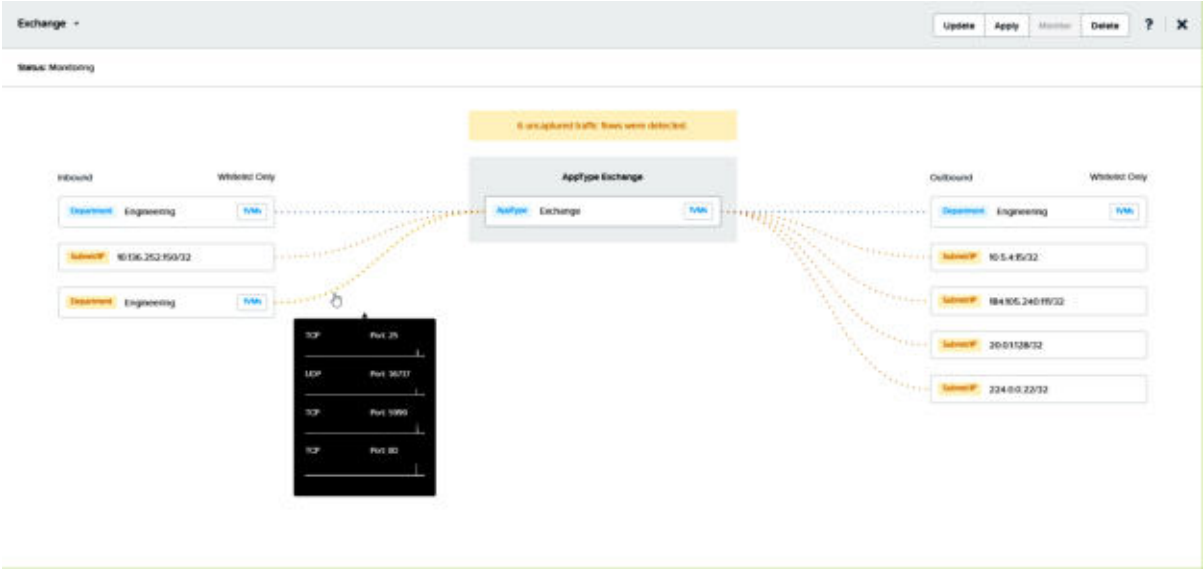


Figure 9: Tooltip Showing a Preview of the Network Flow

4. To see a graph of a network flow, click the dotted line that depicts the network flow in the visualization.
   A more detailed graph of the network flows is displayed, as shown in the following figure:

Figure 10: Network Flows Graph

5. To block unwanted flows, click **Update**, and then update the policy. For information about updating an application security policy, see Modifying an Application Security Policy on page 24.

6. To apply the policy, click **Apply**.
   Applying a policy enforces the policy and traffic from sources that are not allowed is blocked.

Deleting an Application Security Policy

About this task

To delete an application security policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the policy that you want to delete.

   You can select multiple policies and delete them all at once.

2. Click **Delete** in the **Actions** menu.

10

ISOLATION ENVIRONMENT POLICY CONFIGURATION

An isolation environment identifies two groups of VMs by category, and it blocks communications between the groups.

You can also specify an additional category to restrict the scope of the isolation environment to that category.

For example, consider that you have an `application` category with values `app1` and `app2` and that you have associated some VMs with `application: app1` and some VMs with `application: app2`. Also, consider that these same VMs are distributed between two sites, and have accordingly been associated with values site1 and site2 in a category named location (`location: site1` and `location: site2`).

In this example, you might want to block communications between the VMs in the two locations. Additionally, you might want to restrict the scope of the policy to VMs in category `application: app1`. In other words, `app1` VMs in `site1` cannot communicate with `app1` VMs in `site2`. The following diagram illustrates the desired outcome. The red connectors illustrate blocked traffic. The green connectors illustrate allowed traffic.

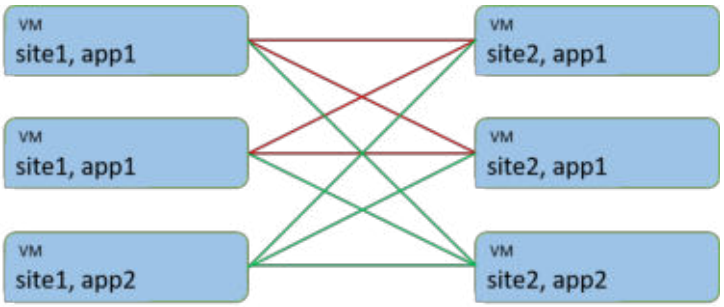


Figure 11: Applications Across Sites

You can configure an isolation policy for this by creating the following categories and isolation policy in Prism Central:

Table 3: Sample Configurations For Categories and the Isolation Policy

Entity	Values
Categories	• **Name**: `application` • **Values**: `app1` and `app2`

Entity	Values
	• **Name**: `location` • **Values**: `site1` and `site2`
Isolation Policy	• **Name**: `eng_isolation_policy_across_sites` • **Description**: Isolate engineering VMs across sites • **Isolate This Category**: `location: site1` • **From This Category**: `location: site2` • **Apply the isolation only within a subset of the data center**: `application: app1`

Layer 2 Isolation

Flow supports Layer 2 isolation to enable filtering of the layer 2 packets across all isolated entities. When an isolation policy is applied between two category-based VM groups, all ingress and egress traffic (broadcast, unknown-unicast, and multicast traffic) is dropped at the destination VM group.

Note:

- If VMs are part of both isolation policy and quarantine policy, the quarantine policy takes priority of processing over the isolation policy. For example, if VMs with category app1 are isolated from VMs with category app2 using an isolation policy, the traffic between these VM groups are not dropped if the VM groups are also part of a quarantine forensic policy that allows communication between these VMs. In this case, since the quarantine forensics policy matches the VMs, and this policy allows the traffic, the isolation policy is not enforced.
- IPv6 traffic between isolated VMs is blocked by default with the introduction of layer 2 isolation.

Creating an Isolation Environment Policy

An isolation environment policy identifies two groups of VMs and blocks communications between the groups. The two groups are identified by category. You can specify an additional category to restrict the scope of the policy to that category.

About this task

To create an isolation environment, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), click **Create Security Policy**, and then click **Isolate Environments**.
   The **Create Isolation Policy** page is displayed.

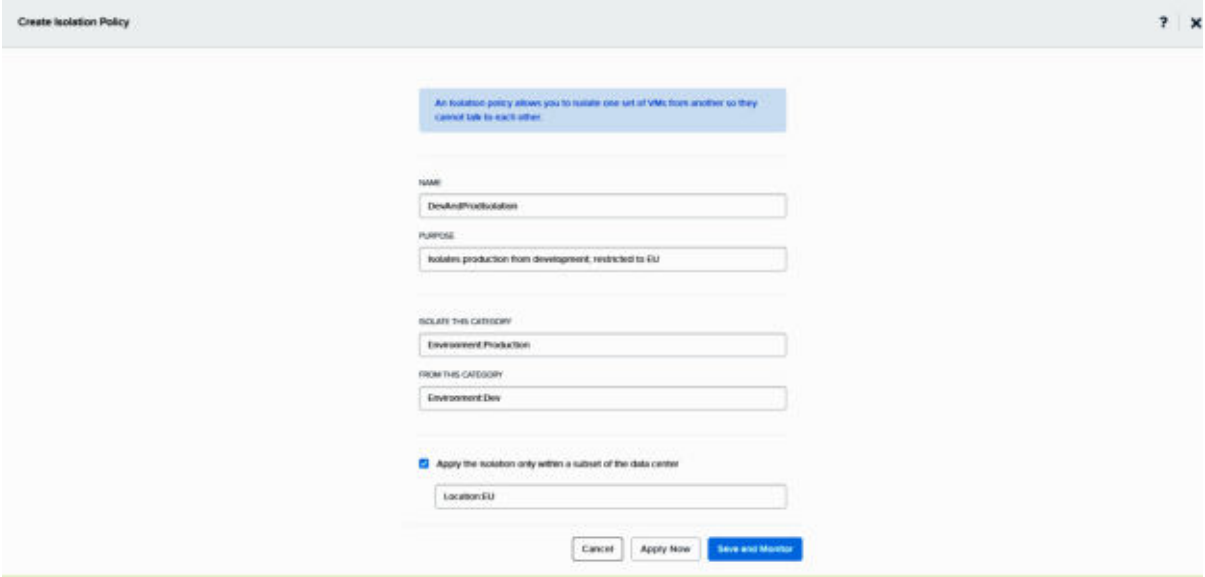


Figure 12: Create Isolation Policy

2. Do the following in the indicated fields:

   - **Name**: Enter a name for the isolation policy.
   - **Purpose**: Describe the purpose of the isolation policy.
   - **Isolate this category**: Type the name of one of the two categories that you want to isolate from each other.

     Matching names appear in a list as you type. You can click the name of the category you want.
   - **From this category**: Type the name of other category.
   - Apply the isolation only within a subset of the data center. If you want to restrict the scope of the policy to a specific category of VMs, select this check box, type the name of the category in the text box, and select the category from the list of matches.

     If you isolate VMs in category `Environment: Production` from VMs in category `Environment: Staging`, and you restrict the scope of the policy to VMs in the category `Environment: Dev`, Prism Central applies the isolation policy to the following groups:

     - VMs that are in both `Environment: Production` and `Environment: Dev`
     - VMs that are in both `Environment: Staging` and `Environment: Dev`.
   - IPv6 Traffic. Optionally, in the **Advanced Configuration** section, select the **Allow** radio button to allow **IPv6 traffic**. The policy rules apply to IPv4 traffic only and all IPv6 traffic is blocked by default.
   - Policy Hit Logs. Optionally, click the toggle button against **Policy Hit Logs** to log traffic flow hits on the policy rules. You can configure syslog monitoring for the policy hit logs for Flow, see "Configuring Syslog Monitoring" in the *Prism Central Guide* for details.

     > Note: Policy hit logs are not generated if both source and destination are in inbound or outbound category.

3. Do one of the following:

   » Click **Apply Now** to apply the isolation environment.

   » Click **Save and Monitor** to save the configuration and place the isolation environment in the monitoring mode.

   You can switch between the monitoring and applied states by selecting the isolation environment on the **Security Policies** page and clicking the appropriate option in the **Actions** menu.

Modifying an Isolation Environment Policy

About this task

To modify an isolation environment, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the isolation policy that you want to modify, click **Actions**, and then click **Update**.

2. Make the changes you want and then apply or save and monitor the policy.

   The update options are the same as those for creating a policy. For information about the options, see Creating an Isolation Environment Policy on page 29.

Applying an Isolation Environment Policy

Applying an isolation environment policy enforces the policy on the specified categories, and any traffic between the categories is blocked.

About this task

> Note: Changing the state of an isolation environment policy affects the functioning of any conflicting application security policies. For more information, see Priorities Between Policies on page 8.

To apply an isolation environment policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the policy that you want to apply, click **Actions**, and then click **Apply**.
2. Confirm by typing **Apply** in the dialog box, and then click **OK**.

Monitoring an Isolation Environment Policy (Visualizing Network Flows)

About this task

The VMs in the two categories in an isolation environment policy are allowed to communicate with each other when the policy is in the monitoring state. Traffic is blocked only during the time the policy is applied.

> Note: Changing the state of an isolation environment policy affects the functioning of any conflicting application security policies. For more information, see Priorities Between Policies on page 8.

To monitor a security policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the policy that you want to monitor, click **Actions**, and then click **Monitor**.
2. Confirm by typing **Monitor** in the dialog box, and then click **OK**.
   The monitoring page shows the flows between the two categories.

3. To view information about a particular network flow, pause over the flow.
   A tooltip similar to the following is displayed:

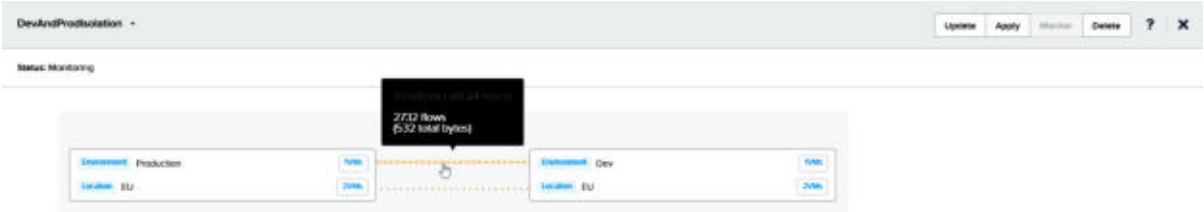


Figure 13: Monitoring Page for an Isolation Environment Policy

Deleting an Isolation Environment Policy

About this task

To delete an isolation environment policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the policy that you want to delete.

   You can select multiple policies to delete them all at once.

2. Click **Delete** in the **Actions** menu.

11

QUARANTINE POLICY CONFIGURATION

Prism Central includes a system defined quarantine policy that enables you to perform the following tasks:

- Completely isolate an infected VM that must not have any traffic associated with it.
- Isolate an infected VM but specify a set of forensic tools that can communicate with the VM.

For these use cases, Prism Central includes built-in categories that are included in the system defined quarantine policy.

> Note: You cannot create a quarentine policy. However, you can modify existing (system defined) quarantine policy.

Prism Central also enables you to monitor the quarantine policy before applying it.

The quarantine policy cannot be deleted.

Configuring the Quarantine Policy

In the built-in quarantine policy, you specify categories that can communicate with VMs that have been added to the **Quarantine: Forensics** category.

About this task

To configure the quarantine policy, do the following;

Procedure

1. In the **Security Policies** dashboard, select **Quarantine**, and then click **Update** in the **Actions** menu.
2. Optionally, in the **Advanced Configuration** under the **Define Policy** tab, do the following.
   a. Select the **Allow** radio button to allow **IPv6 traffic**. The policy rules apply to IPv4 traffic only and all IPv6 traffic is blocked by default. You can configure the allow option for both **Forensic** and **Strict** modes.
   b. Optionally, click the toggle button against **Policy Hit Logs** to log traffic flow hits on the policy rules.

      You can configure syslog monitoring for the policy hit logs for Flow, see Configuring Syslog Monitoring in the *Prism Central Guide* for details. You can enable the policy hit log option for both **Forensic** and **Strict** modes.

      > Note: Policy hit logs are not generated if both source and destination are in inbound or outbound category.

3. On the **Add Forensic Tools** tab, do the following, and then click **Next**:

   a. To specify the categories that contain forensic tools, on the **Inbound** and **Outbound** sides of the policy diagram, do the following:

      - From the drop-down list, select one of the following options:

        - **Allow All**: Allows traffic associated with all sources or destinations.
        - **Whitelist Only**: Allows traffic only if the traffic is associated with the categories and subnets on the allowlist. This option is the default option. If this option is selected, you must also configure the allowlist by clicking **Add Source** or **Add Destination**.

      - Click **Add Source** or **Add Destination**, and then do the following:

        1. Select one of the following options from the drop-down list:

           - **Category**: Allows traffic to or from the specified category.
           - **Subnet/IP**: Allows traffic to or from the specified subnet.
           - **Addresses**: Allows traffic only if the traffic originates from the entities that are in the selected address.
        2. Enter the value (category name or subnet) in the text box, and then click **Add**.

           When entering the name of a category, a list of matching names is displayed, and you can select the name you want to specify. The subnet mask must be specified in the CIDR format.
        3. To add another category or subnet, click **Add Source** or **Add Destination**. Add as many categories or subnets as you want to allow.

   b. To specify the protocols and ports over which the forensic tools can communicate with the VMs in the forensic category, do the following:

      - 1. On the **Inbound** and **Outbound** sides of the policy diagram, click a category or subnet (if you have configured a allowlist) or **All Sources** (if you have chosen to allow all sources) for which you want to create a rule.
        2. Click the plus icon that appears on the **Quarantine: Forensic** category. The **Create Inbound Rule** or **Create Outbound Rule** dialog box
        3. Enter a description for the rule.

           Note: The policy rule description is captured in the policy hitlog data.

           - Policy hitlog must be enabled
           - Rule description is added to the hitlog only for allowed traffic
        4. In **Service Details**, click **Allow all traffic** to allow all types of traffic or click **Select a service** to choose any default or custom service.
        5. Click **Save**.

   After you configure a rule, a dotted line appears between the two endpoints of the rule. Point to the dotted line to show the list of ports that the rule allows.

4. On the **Review** tab, do one of the following:

   » Click **Apply Now** to apply the quarantine policy.

   » Click **Save and Monitor** to save the configuration and place the quarantine policy in the monitoring mode.

   You can switch between the monitoring and applied states by selecting **Quarantine** on the **Security Policies** page and clicking the appropriate option in the **Actions** menu.

Quarantining a VM

You quarantine a VM by adding the VM to a quarantine category.

About this task

To add an infected VM to a quarantine category, do the following:

Procedure

1. In the VMs dashboard **List** tab (see "VMs Summary View" in the Prism Central Guide), select the infected VM, click **Actions**, and then click **Quarantine VMs**.

2. Under Quarantine Method, click one of the following options:

   » **Strict.** Isolates the VM from all traffic. No exceptions can be made for forensics.

   » **Forensic.** Isolates the VM from all traffic except traffic from categories specified in the built-in quarantine policy. The allowed categories contain forensic tools that enable you to perform forensics on the VM.

   For VMs added to the strict quarantine, a red icon is displayed in the name column.

3. Click **Quarantine**.

Removing a VM from the Quarantine

About this task

To remove a VM from the quarantine, do the following:

Procedure

1. In the VMs dashboard **List** tab (see "VMs Summary View" in the Prism Central Guide), select the VM that you want to remove from the quarantine, click **Actions**, and then click **Unquarantine VMs**.

   You can select multiple VMs and remove them from the quarantine in a single step.

2. In the **Unquarantine VMs** dialog box, click **Unquarantine**.

12

VDI POLICY CONFIGURATION

The **VDI Policy** is based on identity-based categorization of the VDI VMs using Active Directory group membership. Configuring VDI policy includes adding an Active Directory domain that is used for the ID firewall (**ID Based Security**) and configuring a service account for the domain.

ID Based Security

ID firewall is an extension to Flow that allows you to write security policies based on users and groups in an Active Directory domain in which your VDI VMs are attached. When using ID firewall, you can import groups from Active Directory into Prism Central as categories (in the category key ADGroup), and then write policies around these categories, just as you would for any other category. A new type of policy has been added for this purpose - the **VDI Policy**. ID firewall takes care of automatically placing VDI VMs in the appropriate categories on detecting user logons into the VM hosted on Nutanix infrastructure associated with Prism Central, thus allowing user and group based enforcement of Flow policies.

- See Configuring Active Directory Domain Services on page 41 to import user groups for identity-based security policies.
- See Creating a VDI Policy on page 37 to create a VDI policy.
- See Default VDI Policy on page 40 configuration to define a default VDI policy.

Note:

- It is recommended to disable credential caching on VDI VMs for Flow ID Firewall. The Flow ID Firewall checks the domain controller events for logon attempts. If the VM connection to the domain controller is not available, a user is able to logon (if credential caching enabled) but no event is generated on the domain controller inhibiting the ID Firewall to detect the logon.
- To disable credential caching, see *Interactive logon: Number of previous logons to cache (in case domain controller is not available)* on *Microsoft* documentation website.
- A basic assumption of VDI Policies is that a single end-user is logged on to each desktop VM at a point in time. As a result, if multiple users log into a single desktop VM at once, the security posture of the VM may change in unpredictable ways. Please ensure that for predictable behavior, only one user is logged into the desktop VMs at a time.

Creating a VDI Policy

ID firewall integrates Nutanix Flow with Microsoft Active Directory (AD), such that the groups in the AD can be imported into Prism Central as categories. These imported categories can then be used in the VDI policy as target groups, inbound traffic, and outbound traffic. Prism Central automatically places VMs inside the imported AD group categories when user logons

are detected on VMs that are part of the Active Directory domain and also present on Nutanix managed clusters, thus applying security policies based on user group membership.

Before you begin

> Note:
>
> - Flow ID firewall is supported only for AHV host compatible with AOS version 5.17 and above and Prism Central version 5.17 and above.
> - Flow ID firewall does not detect user logoffs. The policy applied to a VM is kept applied until next user logon on the same VM.
> - VMs with an **AppType** category assignment do not get categorized by **ID Based Security**.
> - You can use the Default VDI Policy on page 40 to apply a default set of rules for the VDI VMs (without the requirement of user logons).
> - Since a VM user can be a member of multiple ADGroups that are mapped into Prism Central from Active Directory, when a user logs on, a VM may be placed in multiple ADGroups at once. This is the correct behavior, and the policy applied to the VM will be a union of the respective combination of inbounds and outbounds across all ADGroups the VM is placed into.

- If not already available, configure an Active Directory domain that is used for ID firewall, see Configuring Active Directory Domain Services on page 41.
- Configure a service account with required configuration for the Active Directory domain, see Configure Service Account for ID Firewall on page 42.

About this task

To secure a VDI environment, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide) and click **Create Security Policy**. Select **Secure VDI Groups (VDI Policy)** and click **Create**.

   You can create only one VDI policy for securing applications through ID Firewall.

   The **Define Policy** page is displayed.

2. On the **Define Policy** tab, do the following in the indicated fields, and then click **Next**:

   a. The **Policy Name** and **Purpose** fields are auto-populated.

   b. Select either **Include all VMs** or **Include VMs by name** as the **VDI VM Filter**.

   You can use the VDI VM Filter for the following scenarios.

   - **Include VMs by name** - Select **Include VMs by name** and enter the matching criteria in the **VM Name Contains** field. Select the **Assign matching VMs to an optional default category (ADGroup:Default)** check-box to to apply a default posture to the VMs, see Default VDI

Policy on page 40 for details. Optionally, select the **Keep the default category upon user logon** check-box to preserve the default category even after user logon.

> Note:
>
> - Assign ADGroup categories only when the VM matches the filter criteria, otherwise ADGroups apply to all VMs where a logon is detected.
> - VMs with an AppType category assigned is never categorized with an ADGroup.
> - While updating the VDI policy, if inclusion criteria is changed to exclude and then re-include previously included VMs (that were previously logged on and categorized), upon re-inclusion the previous categories will not be applied; consecutively, a new logon must occur for the VM for categories to apply.

- **Include all VMs** - Select **Include all VMs** to include all the VMs in the AD group in the policy. Note that non-VDI VMs will also be included in the policy if **Include all VMs** option is selected.

a. Optionally, in the **Advanced Configuration** section, select the **Allow** option to allow **IPv6 traffic**. The policy rules apply to IPv4 traffic only and all IPv6 traffic is blocked by default.

> Note: If you choose to block IPv6 traffic, the IPv6 traffic remains blocked even in the monitoring mode.

b. Optionally, turn on the **Policy Hit Logs** option to log traffic flow hits on the policy rules.

   You can configure syslog monitoring for the policy hit logs for Flow, see "Configuring Syslog Monitoring" in the *Prism Central Guide* for details.

> Note: Policy hit logs are not generated if both source and destination are in inbound or outbound category.

3. In the **Secure AD Groups** tab, do the following in the indicated fields and click **Next**.

   a. For **Inbound Traffic**, click **+ Add Source** and enter the category or subnets that the VDI group can receive the traffic from, as the source.

   b. For each **VDI ADGroup**, click **+Add AD Group** to select the AD groups (categorized VDI VMs) that you want to secure. You can click **Import all AD Groups** to add all imported ADGroup categories to the VDI policy.

   c. For **Outbound Traffic**, click **+ Add Destination** and enter the category or subnets that the VDI group can send the traffic to, as the destination.\

   > Note: If you have not used the default VDI option in *Step 2b*, ensure that you add all of your Active Directory domain controllers as part of this step, using either categories or subnets, for each ADGroup.

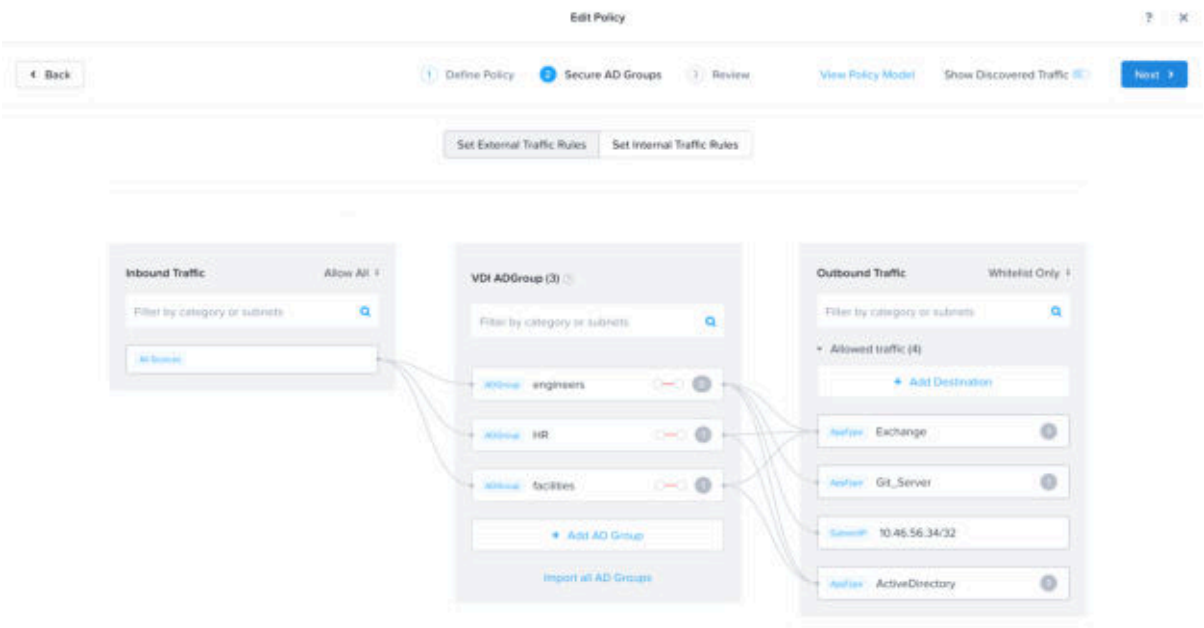


Figure 14: Secure AD Groups Tab

4. Do one of the following:

   » Click **Apply Now** to apply the VDI Policy.

   » Click **Save and Monitor** to save the configuration.

   You can switch between the monitoring and applied states on the **Security Policies** page and clicking the appropriate option in the **Actions** menu.

Default VDI Policy

The Default VDI policy feature allows you to apply a default set of rules as defined by the desktop administrator for VDI VMs and users. There are two primary use cases for Default VDI Policy (**ADGroup:Default**).

- To ensure that a VDI VM is secure even before a user logs on to the VDI VM.
- To enable access to common network resources without the need to add the resources to every tier of a VDI policy.

You can define a default VDI policy at the time of creating a new VDI policy, or by updating any existing VDI policy. See *Step 2b* of the VDI Policy Configuration topic for details.

Configuring Active Directory Domain Services

Active Directory Domain Services configuration is used to import user groups for identity based security policies.

Before you begin

- Microsegmentation must be enabled to be able to use the ID Firewall feature, see Enabling Microsegmentation on page 10.
- You must allow WMI access from Prism Central to all the Active Directory Domain Controllers in your network firewall and Active Directory firewall.
- Active Directory Requirements:
  - Minimum supported domain functional level in Active Directory is Windows Server 2008 R2.
  - ID Firewall checks the membership of Security Groups only, Distribution Groups are not supported.
  - NTP must be configured on Active Directory and Prism Central.
  - DNS must be configured on Prism Central if you want to use host name for domain controllers.

About this task

To configure an Active Directory domain, do the following.

Procedure

1. Log on to the Prism Central web console.
2. Click the collapse menu ("hamburger") button on the left of the main menu and then select **Prism Central Settings** to display the Settings page.
3. Click **ID Based Security** from the Settings menu (on the left).
   The **ID Based Security** page is displayed. This page allows you to **Add New Domain** or use an **Existing AD**.
4. If you select **Use Existing AD** in step 3, do the following in the indicated fields:
   a. Click the **Manually Add Domain Controller** button, then click **+ Domain Controller**.
   b. Enter the **IP Address** or **Host Name** of the domain controllers that you want to monitor for user logons events. You must add all the domain controllers associated with your Active Directory manually.

      Click **+** and add each domain controller individually, then click the blue check mark icon to save.

      > Note: DNS must be configured on Prism Central for the host name option to work.

5. If you select **Add New Domain** in step 3, a set of fields is displayed. Do the following in the indicated fields:

   a. **Name**: Enter a directory name.

      This is a name you choose to identify this entry; it need not be the name of an actual directory.

   b. **Domain**: Enter the domain name.

      Enter the domain name in DNS format, for example, `nutanix.com`.

   c. **Directory URL**: Enter the LDAP address of the directory, including the port number.

   d. **Service Account Username**: Enter the service account user name in the `user_name@domain.com` format that you want Prism Central to use to detect logons and query user and group information from Active Directory.

      > Caution: Do not use the Domain Admin account as the service account considering the security best practices. Create a new domain user and grant it required permissions as described in Configure Service Account for ID Firewall on page 42.

      A service account is a special user account that an application or service uses to interact with the Active Directory. Enter your Active Directory service account credentials in this (username) and the following (password) field.

      > Note: Ensure that you update the service account credentials here whenever the service account password changes or when a different service account is used.

   e. **Service Account Password**: Enter the service account password.

   f. When all the fields are correct, click the **Save** button (lower right).

      ID Firewall uses the service account for ID based security with additional requirements, see Configure Service Account for ID Firewall on page 42.

   Once saved, the **Referenced AD Groups** section is displayed. You can add a new user group by clicking **+ Add User Group** and edit the auto-generated **Category Value**. After the active directory configuration is complete, you can create the VDI Policy, see Creating a VDI Policy on page 37

6. Select **Add Inclusion Criteria** under **Manage the VM Inclusion Criteria** to specify which VMs are assigned to AD Group categories upon user logon based on VM name.

   > Note: It is recommend that users add inclusion criteria if at all possible to prevent any unintended categorizations.

   > Note: The VMs with AppType category assigned cannot be categorized by ID Based Security.

Configure Service Account for ID Firewall

Active Directory service account in Prism Central is used for connectivity with the Active Directory domain services. ID Firewall also uses the same service account for ID based security.

To configure a service account for ID firewall, do the following.

1. Create a new user in the Active Directory.
2. Add the user to the **Distributed COM Users** group and the **Event Log Readers** domain groups.
3. Start the **dcomcnfg.exe** utility and go to **Component Services** > **Computers** > **My Computer** > **DCOM Config**.

4. Right-click on **Windows Management and Instrumentation** and select **Properties** from the menu.
5. Switch to **Security** tab, select **Customize** option in the **Access Permissions** section and then click **Edit**.
6. Add the user and grant **Local Access** and **Remote Access** permissions to the user. Click **OK** to confirm changes.
7. Run the `WMIMGMT.msc` command to start **Windows Management Instrumentation** snap-in.
8. Right-click on **WMI control (local)** and select **Properties** from the menu.
9. Switch to **Security** tab and expand **Root** tree.
10. Select **CIMV2** in the expanded tree and click **Security**.
11. Go to **Advanced** > **Add** > **Principal** and enter the user name.
12. Change scope by selecting This namespace and subnamespaces in the Applies to drop-down menu.
13. Click the check-box to grant the **Enable Account** and **Remote Enable** permissions. Click **OK** to confirm changes.
14. Restart the `winmgmt` service.

    ```
    C:\> net stop winmgmt
    C:\> net start winmgmt
    ```

    Alternatively, reboot the domain controller.
15. Repeat step 3 to step 14 on every domain controller.

Modifying the VDI Policy

About this task

To modify the VDI policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the *Prism Central Guide*), select the policy that you want to modify, click **Actions**, and then click **Update**.

2. Make the changes you want and then apply or save and monitor the policy.

   The update options are the same as those for creating a policy. For information about the options, see Creating a VDI Policy on page 37.

Applying the VDI Policy

Applying the VDI policy enforces the policy on the specified categories (VDI AD groups), and any traffic between the categories is blocked.

About this task

To apply the VDI policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the policy that you want to apply, click **Actions**, and then click **Apply**.

2. Confirm by typing **Apply** in the dialog box, and then click **OK**.

Monitoring the VDI Policy

About this task

The VMs in VDI AD Groups in the VDI policy are allowed to communicate with each other when the policy is in the monitoring state. Traffic is blocked only during the time the policy is applied.

To monitor a security policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the *Prism Central Guide*), select the policy that you want to monitor, click **Actions**, and then click **Monitor**.
2. Confirm by typing **Monitor** in the dialog box, and then click **OK**.

Deleting the VDI Policy

About this task

To delete the VDI policy, do the following:

Procedure

1. In the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide), select the VDI policy.
2. Click **Delete** in the **Actions** menu.

13

APPLYING FILTERING AND GROUPING TO A SECURITY POLICY

You can apply different types of filters to view results based on properties like source , destination, category, ports, and more. You can also group related rule attributes together for easier visualization of connection flows. Grouping and Filtering work together to provide an intuitive view for the security policy.

About this task

To apply filtering and grouping to a security policy, do the following.

Procedure

1. Log on to the Prism Central web console.
2. Click the collapse menu ("hamburger") button on the left of the main menu and go to **Policies** > **Security**. The **Policies** page is displayed.
3. Click any policy to view the inbound, application, and outbound configuration.

4. To view specific rule properties, do one of the following.

   » In the **Search** box, search for the required string using the default **All** filter.

   » Click the filter drop-down menu to search the policy based on any of the following filter types.

   Category
   search category name and value

   Address
   search address and subnet IP address

   Subnet IP
   search subnet IP address

   Service
   search service name

   Rule Description
   search rule description

   Ports (TCP/UDP)
   search TCP/UDP ports and services

   ICMP
   search ICMP ports and services

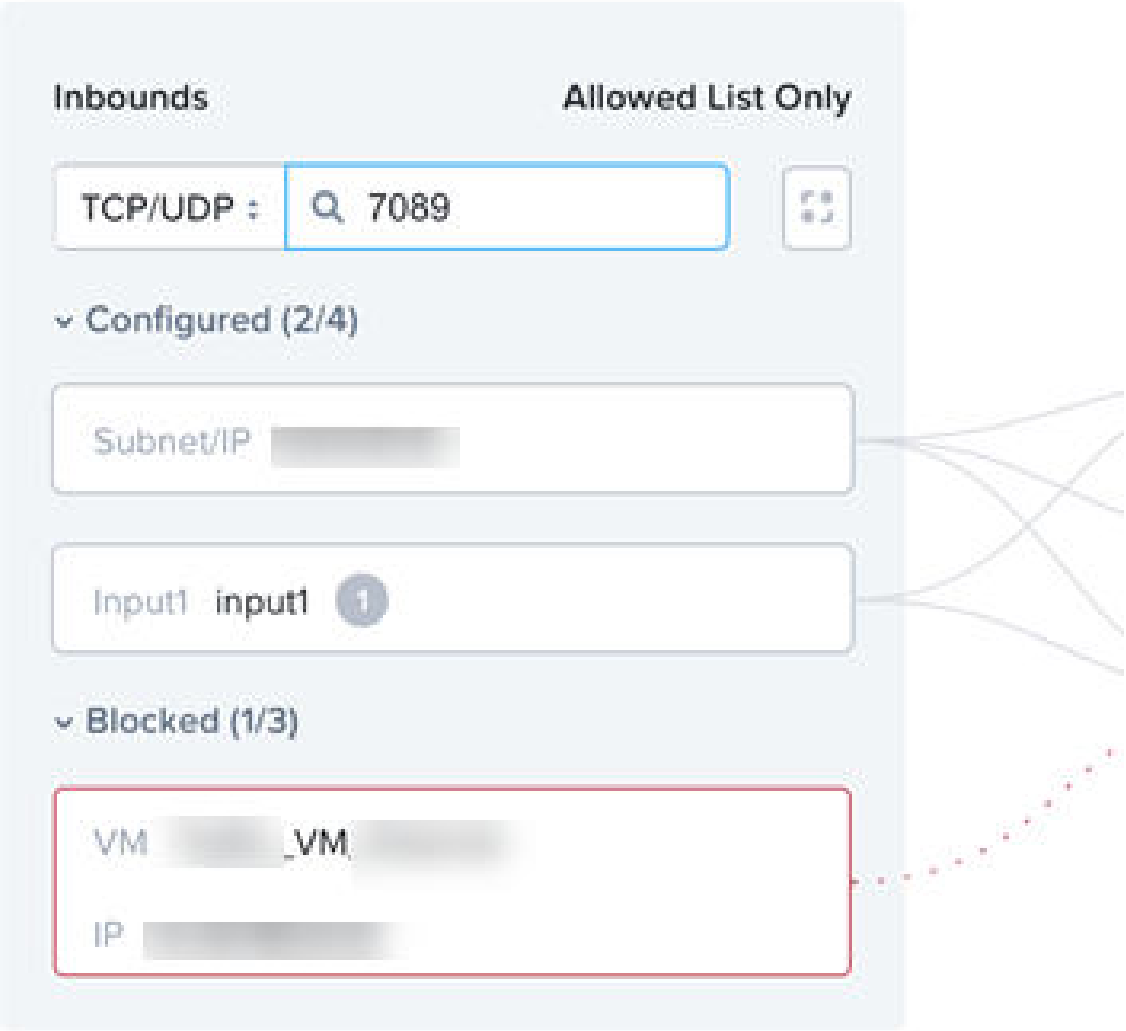


Figure 15: Filtering Policies

5. To group related rule entities together, click the group icon.
   The group option organizes related rule attributes like subnet IP, categories, and service in distinct boxes. Also, the connection flows for all the entities in a group are displayed as a

single connection flow. To view all the entities belonging to a group, click the down-arrow icon to expand the group.

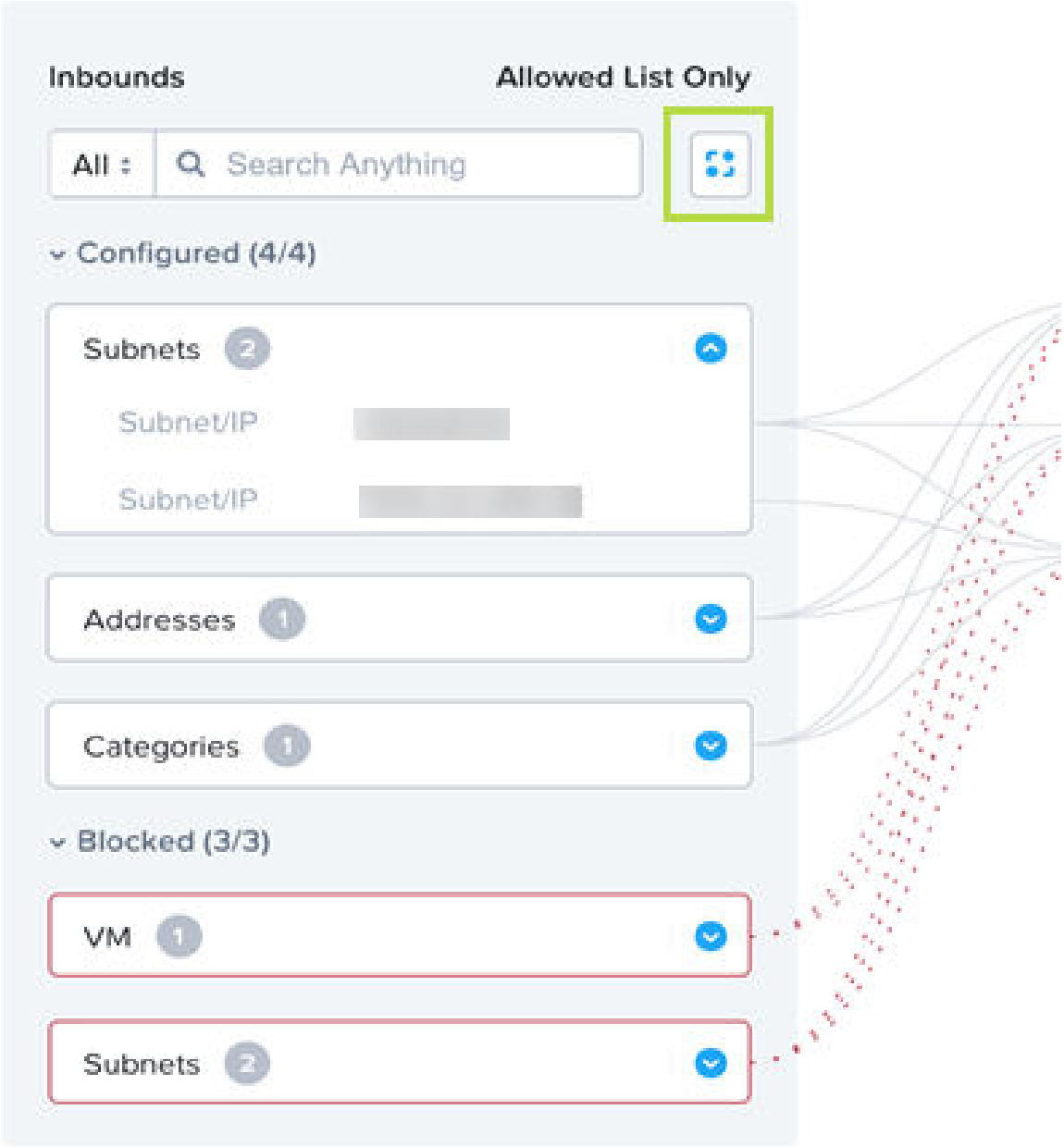


Figure 16: Filtering Policies

14

EXPORTING AND IMPORTING SECURITY POLICIES

Prism Central allows you to export and import security policies for the following security administration aspects.

- Have a snapshot of a working security configuration so that system can be restored to the desired state when needed.
- Ability to apply security policies as templates. This scenario is useful in ROBO environments (disaster recovery deployments) where the datacenters are being managed by multiple Prism Central instances.

Exporting Security Policies

To export or import security policy, do the following in the **Security Policies** dashboard (see "Security Policies Summary View" in the Prism Central Guide)).

> Note: For VDI policy, the inclusion criteria and default VDI category settings are not included in the export process. You must set these manually after an import if required.

- Click the **Export & Import** drop down menu.
- To export the security policies, select **Export Security Policy**. The security policies binary file is downloaded.
- To import any previously exported security policies binary file, select **Import Security Policy**, then click **Browse** to select the binary file. Click Import. The security policies are imported.

> Note: Existing policies are overridden with new policies. Policies that are not part of this import are deleted.

AOS 6.1

Flow Networking Guide

March 23, 2022

Contents

1. Purpose 4

2. Related Documentation 5

3. Flow Networking Overview 6
   - Flow Networking Architecture 7
   - Essential Concepts 8

4. Prerequisites for Enabling Flow Networking 14

5. Flow Networking Configurations 18
   - Enabling Flow Networking 18
   - Disabling Flow Networking 21
     - Unregistering a PE from the PC 22
   - Upgrading Flow Networking 23
   - Deploying Flow Networking at a Dark Site 25
   - Troubleshooting Tips 26

6. Network Gateway Upgrades 28
   - Identifying the Gateway Version 28
   - Detecting Upgrades for Gateways 29
   - Upgrading the PC-managed Onprem Nutanix VPN Gateways 29

7. Network and Security View 31
   - Subnets 31
     - Subnet Summary View 34
   - Virtual Private Clouds 36
     - Customizing the VPC List View 37
     - VPC Details View 39
   - Floating IPs 41
   - Connectivity 42
     - Gateways Summary View 42
     - VPN Connections Summary View 44
     - Subnet Extensions Summary View 48
   - Security Policies Summary View 53
     - Security Policy Details View 54

8. Virtual Private Cloud 56
   - VM IP Address Management 56
     - Creating Secondary IP Addresses 57
     - Assigning Secondary IP Addresses to Interfaces 61
     - Assigning Secondary IP Addresses to Floating IPs 62

VPC Workflow ... 62
VPC Management ... 62
Creating Virtual Private Cloud ... 62
Requesting Floating IPs ... 66
Creating a Subnet ... 68
Attaching a Subnet to a Virtual Machine ... 75
Creating a Policy ... 78
Creating Static Routes ... 83
Updating Virtual Private Cloud ... 84
Updating a Subnet ... 85
Updating a Policy ... 89
Updating Static Routes ... 89
Deleting a Virtual Private Cloud ... 90
Deleting Subnets, Policies or Routes ... 90

9. Connections Management ... 92
Virtual Network Connections ... 92
VPN Workflow ... 93
Prerequisites for VPN Configurations ... 93
Creating a Network Gateway ... 96
Updating a Network Gateway ... 107
Deleting a Network Gateway ... 108
Creating a VPN Connection ... 108
Updating VPN Connection ... 109
Deleting a VPN Connection ... 110
Layer 2 Virtual Network Extension ... 110
Layer 2 Virtual Subnet Extension Over VPN ... 113
Layer 2 Virtual Subnet Extension Over VTEP ... 118
Updating an Extended Subnet ... 127
Removing an Extended Subnet ... 127

Copyright ... 129
License ... 129
Conventions ... 129
Version ... 129

1

PURPOSE

This *Flow Networking Guide* describes how to enable and deploy Nutanix Flow Networking on Prism Central.

Upgrading from EA Versions

If you have enabled the early access (EA) version of Flow Networking, disable it before upgrading the Prism Central and enabling the general availability (GA) version of Flow Networking.

2

RELATED DOCUMENTATION

Links to Nutanix Support Portal software and documentation.

The Nutanix Support Portal provides software download pages, documentation, compatibility, and other information/

Documentation	Description
Nutanix Security Guide	Prism Element and Prism Central security, cluster hardening, and authentication.
AOS guides and release notes	Covers AOS Administration, Hyper-V Administration for Acropolis, Command Reference, Powershell Cmdlets Reference, AOS Family Release Notes, and AOS release-specific Release Notes
Acropolis Upgrade Guide	How to upgrade core and other Nutanix software.
AHV guides and release notes	Administration and release information about AHV.
Prism Central and Web Console guides and release notes	Administration and release information about Prism Central and Prism Element.

3

FLOW NETWORKING OVERVIEW

Enabled and administered from Prism Central, Flow Networking powers network virtualization to offer a seamless network experience with enhanced security. It is disabled by default.

It is a software-defined network virtualization solution providing overlay capabilities for the on-prem AHV clusters. It integrates tools to deploy networking features like Virtual Private Cloud (VPC) and Virtual Private Network (VPN) to support flexible app-driven networking that focuses on VMs and applications instead of virtual LANs and network addresses.

After you enable it on Prism Central, Flow Networking delivers the following.

- A simplified, Prism Central-based workflow that deploys the application-driven network virtualization feature.
- A secure multi-tenancy solution allowing per-tenant isolation using VPC-based network segmentation and namespace isolation.
- A secure VPN-based connectivity solution for multiple sites, with automated VPN bundle upgrades.
- NAT-based secure egress to external networks, with IP address retention and policy-based routing.
- Self-serve networking services using REST APIs.
- Enhanced networking features for more effective disaster recovery.

> Note: You can enable network segmentation on a Layer 2 extended virtual subnet that does not have a gateway. For more information about Layer 2 subnet extensions, see Layer 2 Virtual Network Extension on page 110. For information about network segmentation of an extended layer 2 subnet, see *Segmenting a Stretched L2 Network for Disaster Recovery* in the *Securing Traffic through Network Segmentation* section of the *Security Guide*.

Deployment Workflow

You can enable Flow Networking using a simple Prism Central driven workflow, which installs the network controller. The network controller is a collection of containerized services that run directly on the Prism Central VM(s). The network controller orchestrates all the virtual networking operations.

- Ensure that microservices infrastructure is enabled in **Prism Central Settings** > **Prism Central Management**. See *Prism Central Guide* for information about enabling microservices infrastructure.
- Enable Flow Networking in **Prism Central Settings** > **Advanced Networking**. It is disabled by default. See Enabling Flow Networking on page 18
- You can opt out of Flow networking by disabling the **Advanced Networking** option subject to prerequisites to disable advanced networking. See Disabling Flow Networking on page 21.

- You can deploy Flow Networking in a dark site (a site that does not have Internet access) environment. See the Deploying Flow Networking at a Dark Site on page 25 topic for more information.
- You can upgrade the Flow networking controller. Nutanix releases an upgrade for the Flow networking controller with AOS and Prism Central releases. See Upgrading Flow Networking on page 23.

  See the *AOS Family Release Notes* and *Release Notes | Prism Central*.
- Flow networking allows you to create and manage virtual private clouds (VPCs) and overlay subnets to leverage the underlying physical networks that connect clusters and datacenters. See Virtual Private Cloud on page 56.
- You can upgrade the network gateway version. Network gateway is used to create VPN or VTEP gateways to connect subnets using VPN connections, or Layer 2 subnet extensions over VPN or VTEP.

Flow Networking Architecture

The Flow Networking architecture uses a three-plane approach to simplify network virtualization.

Prism Central provides the management plane, the network controller itself acts as the control plane while the AHV nodes provide the data plane. This architecture provides a strong foundation for Flow Networking. This architecture is depicted in the following chart.

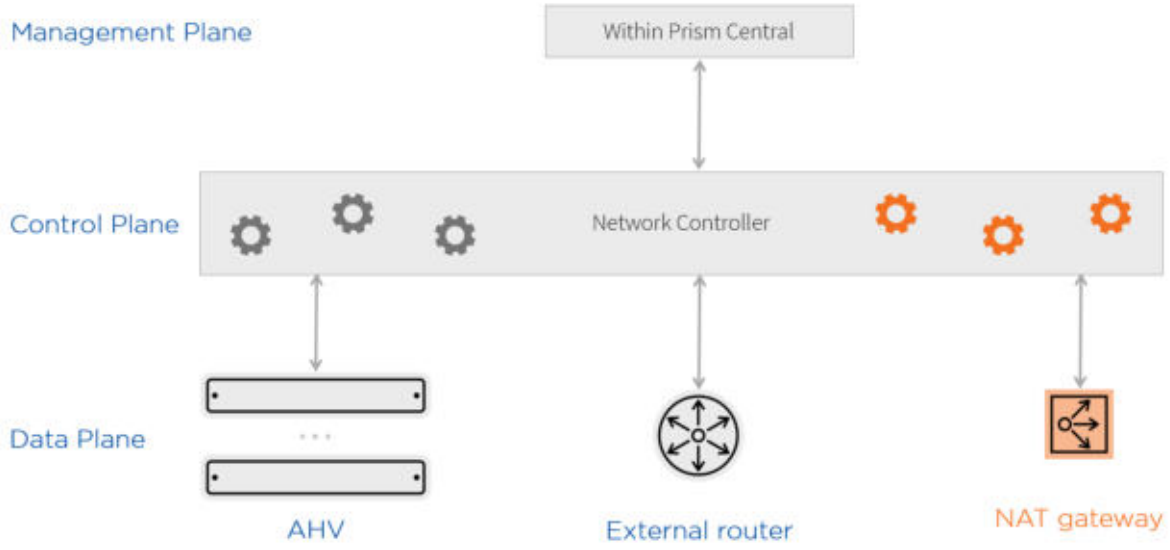


Figure 1: Flow Networking Architecture

Deployment Scale

Flow Networking supports the following scale:

Entities	Scale
Virtual Private Clouds	500
Subnets	5,000

Entities	Scale
Ports	50,000
Floating IPs	2,000 per networking controller-enabled Prism Central.
Routing Policies	1,000 per Virtual Private Cloud. 10,000 per networking controller-enabled Prism Central.

Essential Concepts

VPC

A Virtual Private Cloud (VPC) is an independent and isolated IP address space that functions as a logically isolated virtual network. A VPC could be made up of one or more subnets that are connected through a logical or virtual router. The IP addresses within a VPC must be unique. However, IP addresses may overlap across VPCs. As VPCs are provisioned on top of another IP-based infrastructure (connecting AHV nodes), they are often referred to as the overlay networks. Tenants may spin up VMs and connect them to one or more subnets within a VPC. Virtual Private Cloud (VPC) is a virtualized network of resources that are specifically isolated from the rest of the resource pool. VPC allows you to manage the isolated and secure virtual network with enhanced automation and scaling. The isolation is done using network namespace techniques like IP-based subnets or VLAN based networking.

VPC Subnets

You can use IP address-based subnets to network virtual machines within a VPC. A VPC may use multiple subnets. VPC subnets use private IP address ranges. IP addresses within a single VPC must be unique, in other words, IP addresses inside the same VPC cannot be repeated. However, IP addresses can overlap across multiple VPCs. The following figure shows two VPCs named Blue and Green. Each VPC has two subnets, 192.168.1.0/24 and 192.168.2.0/24, that are connected by a logical router. Each subnet has a VM with an IP address assigned. The subnets and VM IP addresses overlap between the two VPCs.

Figure 2: VPC Subnet

The communication between VMs in the same subnets or different subnets in the same VPC (also called East-West communication) is enabled using GEneric NEtwork Virtualization Encapsulation (GENEVE). If a Prism Central manages multiple clusters, then the VMs that

belong to the same VPC could be deployed across different clusters. The virtual switch on the AHV nodes provide distributed virtual switching and distributed virtual routing for all VPCs.

The communication from a VM in a VPC to an endpoint outside the VPC (called external communication or North-South communication) is enabled by an external network connection. Such a connection may be secured using VPN. The following figure shows the logical connectivity of the VPCs to the external network, and subsequently to the Internet.

> Note: You must configure the default route (0.0.0.0/0) to the external subnet as the next hop for connectivity outside the cluster (north-south connectivity).

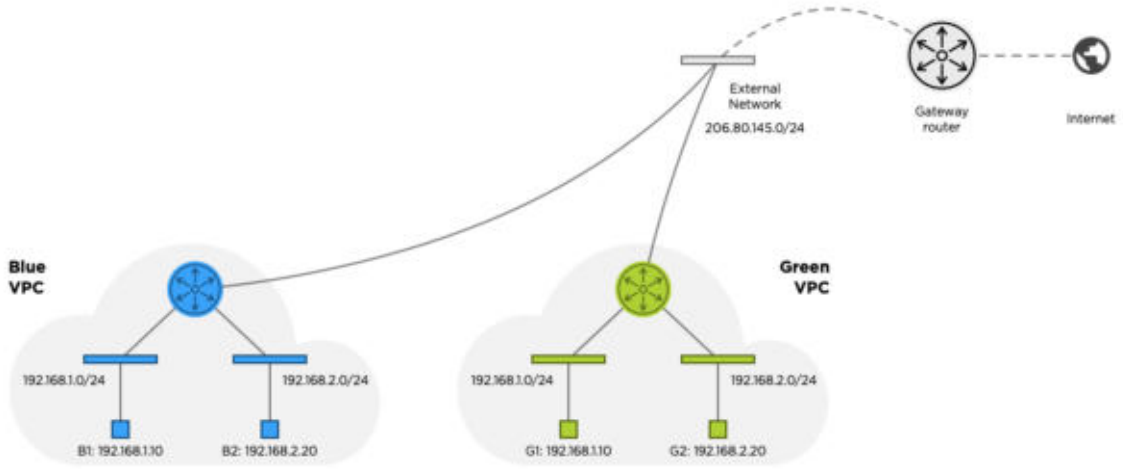


Figure 3: External Communication

External Subnets

Subnets outside a VPC are external subnets. External subnets may be subnets within the deployment but not included in a specific VPC. External subnets may also be subnets that connect to the endpoints outside the deployment such as another deployment or site.

External subnets can be deployed with NAT or without NAT. You can add a maximum of two external subnets - one external subnet with NAT and one external subnet without NAT to a VPC. Both external subnets cannot be of the same type. For example, you cannot add two external subnets, both with NAT. You can update an existing VPC similarly.

Primary and Secondary IP Addresses for VMs

See VM IP Address Management on page 56.

SNAT and Floating IP Address

SNAT and Floating IP addresses are used only when you use NAT for an external subnet.

In Source Network Address Translation (SNAT), the NAT router modifies the IP address of the sender in IP packets. SNAT is commonly used to enable hosts with private addresses to communicate with servers on the public Internet.

For VMs within the VPC to communicate with the rest of the deployment, the VPC must be associated with an external network. In such a case, the VPC is assigned a unique IP address, called the SNAT IP, from the subnet prefix of the external network. When the traffic from a VM needs to be transmitted outside the VPC, the source IP address of

the VM, which is a private IP address, is translated to the SNAT IP address. The reverse translation from SNAT IP to private IP address occurs for the return traffic. Since the SNAT IP is shared by multiple VMs within a VPC, only the VMs within the VPC can initiate connections to endpoints outside the VPC. The NAT gateway allows the return traffic for these connections only. Endpoints outside the VPC cannot initiate connections to VMs within a VPC.

In addition to the SNAT IP address, you can also request a Floating IP address — an IP from the external subnet prefix that is assigned to a VM via the VPC that manages the network of the VM. Unless the floating IP address is assigned to the private IP address (primary or secondary IP address) of the VM, the floating IP address is not reachable. When the VM transmits packets outside the VPC, the private IP of the VM is modified to the Floating IP. The reverse translation occurs on the return traffic. As the VM uses the Floating IP address, an endpoint outside the VPC can also initiate a connection to the VM with the floating IP address.

The translation of the private IP addresses to Floating IP or SNAT IP address, and vice versa, is performed in the hypervisor virtual switch. Therefore, the VM is not aware of this translation. Floating IP translation may be performed on the hypervisor that hosts the VM to which the floating IP is assigned to. However, SNAT translation is typically performed in a centralized manner on a specific host.

NAT Gateway

NAT Gateways are used only when you use NAT for an external subnet.

Network Address Translation (NAT) is a process for modifying the source or destination addresses in the headers of an IP packet while the packet is in transit. In general, the sender and receiver applications are not aware that the IP packets are being manipulated.

A NAT Gateway provides the entities inside an internal network with connectivity to the Internet without exposing the internal network and its entities.

A NAT Gateway is:

- A node or a AHV host. You need a host or a node to implement a NAT Gateway because NAT gateways require operations like load balancing and routing that are automatically performed by Flow Networking.
- Connected to the internal network with an internal subnet based IP address and to the external network with an externally-routable IP address.

  The externally-routable IP address may be an IP address from a private IP address space or an RFC1918 address that is used as a NAT gateway. The NAT Gateway IP address could be a static IP address or a DHCP assigned IP address.

Table 1: NAT Gateway Failover Time

Event	Failover Time
Network controller stops on AHV	Up to 45 seconds.
Node reboot	Up to 45 seconds.
Node power off:	Up to 45 seconds.
When NAT Gateway and network controller MSP worker VMs are not on the same node.	

Event	Failover Time
Node power off: When NAT Gateway and network controller MSP worker VMs are on the same node.	Up to 300 seconds (5 minutes).

Static IP Address

A static IP address is a fixed IP address that is manually assigned to an interface in a network. Static IP addresses provide stable routes that do not have to be updated frequently in the routing table since the static routes generated using static IP addresses do not need to be updated.

Usually in a large IP-based network (a network that uses IP addresses), a Dynamic Host Configuration Protocol or DHCP server assigns IP addresses to interfaces of an entity (using DHCP client service on the entity). However, some entities may require a static IP address that can be reached (manual remote access or via VPN) quickly. A static IP address can be reached quickly because the IP address is fixed, assigned manually and is stored in the routing table for a long duration. For example, a printer in an internal network would need a static IP address so that it can be connected reliably. Static IP addresses can be used to generate static routes which remain unchanged in routing tables, thus providing stable long-term connectivity to the entity that has the static IP address assigned.

Static Route

Static routes are fixed routes that are created manually by the network administrator. Static routes are more suited for small networks or subnets. Irrespective of the size of a network, static routes may be required in a variety of cases. For example, in VPCs where you use virtual private networks (VPNs) or Virtual Tunnel End Point (VTEP) over VxLAN transport connections to manage secure connections, you could use static routes for specific connections such as site-to-site connections for disaster recovery. In such a case it is necessary to have a known reliable route over which the disaster recovery operations can be performed smoothly. Static routes are primarily used for:

- Facilitating the easy maintenance of the routing table in small networks that are not expected to grow.
- Routing to and from other internal route or stub networks. A stub network or an internal route network is a network accessed using a single route and the router has only one neighbor.
- Use as a default or backup route. Such a route is not expected to specifically match any other route in the routing table.

In a network that is not constantly changing, static routes can provide faster and more reliable services by avoiding the network overheads like route advertisement and routing table updates for specific routes.

Overlay networks

You can create an IP-based Overlay subnet for a VPC. An Overlay network is a virtualized network that is configured on top of an underlying virtual or physical network. A special purpose multicast network can be created as an Overlay network within an existing network. A peer-to-peer network or a VPN are also examples of Overlay networks. An important assumption for an Overlay network is that the underlying network is fully connected. Nutanix provides the capability to create Overlay network-based VPCs.

Comparing Overlay with VLAN

See how overlay networks compare with VLAN networks. A virtual local area network or VLAN network is a Layer 2 network that provides virtualized network segmentation solution. VLANs route and balance traffic in a network based on MAC addresses, Protocols such as Ethernet, ports or specific subnets. A VLAN creates a virtual Layer 3 network using Layer 2 addressing by separating broadcast domains virtually or logically. A VLAN configured network behaves as if the network is segmented using a physical layer 2 switch without implementing a layer 3 IP based subnet for the segmentation. VLAN traffic usually cannot traverse outside the VLAN.

The main advantage that VLAN networks provide is that VLAN networks require only layer 2 (L2) connectivity. VLANs do not require any of the layer 3 (L3) Flow Networking features.

Overlay networks can be laid on underlying physical network connections including VLAN networks. Overlay networks provide the following advantages and constraints:

- IP address namespace is decoupled from the physical network.
- You can create, update or delete overlay networks without requiring any configurations on the physical network and powering down the systems.
- You can create overlay networks that can span across multiple clusters.
- VLAN networks are necessary for Bootstrapping of Flow Networking.

  > Note: Nutanix recommends using VLAN0 as the default untagged (also called native) VLAN for a CVM and AHV host. You can create VLANs for user VMs using the Network Configuration page. You can use the Create Virtual Switch dialog box from the Network Configuration page to create virtual switches for the user VM VLANs.

- AHV Networking VLAN and Flow Networking VLAN: VLAN backed subnets for external connectivity are managed by the Flow Networking control plane. Traditional AHV VLAN IPAM networks are managed by Acropolis. Do not configure the same VLAN as both a Flow Networking external network and an AHV IPAM network, as this can lead to IP address conflicts.

Traffic Behavior

Broadcast Traffic

When all the guest VMs belonging to a subnet are in the same AHV: Flow Networking broadcasts the traffic to all guest VMs in the same subnet.

When some VMs belonging to a subnet are in other AHVs: Flow Networking tunnels the traffic to only those AHVs which have endpoints in the same subnet.

In other words, Flow Networking broadcasts traffic to all the guest VMs in the same subnet.

Unicast Traffic

Unicast traffic is traffic transmitted on a one-to-one basis between IP addresses and ports. There is only one sender and one receiver for the traffic. Unicast traffic is usually the most used form of traffic in any LAN network using Ethernet or IP networking. Flow Networking transmits unicast traffic based on the networking policies set.

Unknown Unicast Traffic

Flow Networking always drops unknown unicast traffic. It is not transmitted to any guest VM within or outside the source AHV.

Multicast Traffic

Flow Networking transmits the traffic to the VMs in the multicast group within the same subnet. If the VM is on another AHV, the destination AHV must have an endpoint in the subnet.

Multicast Group

A multicast group is defined by an IP address (called a multicast IP address, usually a Class D IP address) and a port number. Once a host has group membership, the host will receive any data packets that are sent to that group defined by an IP address/port number.

4

PREREQUISITES FOR ENABLING FLOW NETWORKING

Make sure you meet these prerequisites before you enable Flow networking on Prism Central.

Requirements

> Important: Prism Central protection and recovery does not protect or recover Flow networking services.

You must have the following fulfilled to enable Flow networking:

- Ensure that the Prism Central running Flow networking is hosted on an AOS cluster running AHV.

  The network controller has a dependency only on the AHV version.
- Ensure that microservices infrastructure on Prism Central is enabled. See *Prism Central Guide* for information about microservices infrastructure.
- Choose the x-large PC VM size for Flow networking deployments. Small or large PC VMs are not supported for Flow Networking.

  If you are running a small or large Prism Central VMs, upgrade the Prism Central VM resources to x-large PC VM. See *Acropolis Upgrade Guide* for procedure to install an x-large Prism Central deployment.
- Although Flow networking may be enabled on a single-node PC, Nutanix strongly recommends that you deploy a three-node scale-out Prism Central for production deployments. The availability of Flow networking service in Prism Central is critical for performing operations on VMs that are connected to overlay networks. A three-node scale-out Prism Central ensures that Flow networking continues to run even if one of the nodes with a PCVM fails.
- Prism Central VM registration. You cannot unregister the Prism Element cluster that is hosting the Prism Central deployment where you have enabled Flow Networking. You can unregister other clusters being managed by this Prism Central deployment.

- Ensure that Microservices Infrastructure (CMSP) is enabled on Prism Central before you enable Flow Networking. See the *Prism Central Guide* for more information.

  For the procedure to enable Microservices Infrastructure (including enable in dark site), see *Enabling Micro Services Infrastructure* section in the *Prism Central Guide*.

  > Note: When you configure microservices infrastructure, ensure that the DNS name you configure for CMSP does not end with test. Flow networking does not support test as a top level domain. For example, the following are valid domain configurations:
  >
  > - my.cluster.domain
  > - my.test.cluster.test.domain
  >
  > However, the following are examples of domains that Flow networking does not support:
  >
  > - my.cluster.test
  > - my.cluster.domain.test

- Ensure that you have created a virtual IP address (VIP) for Prism Central. The *Acropolis Upgrade Guide* describes how to set the VIP for the Prism Central VM. Once set, do not change this address.
- Ensure connectivity:
  - Between Prism Central and its managed Prism Element clusters.
  - To the Internet for connectivity (not required for dark site) to:
    - ECR for Docker images
    - S3 storage for LCM portal

    > Note: For dark site deployments, Nutanix provides a dark site bundle, which has the Docker images (normally hosted on ECR) and the network controller package (normally hosted on LCM portal). These dark site bundles can be downloaded using an internet-connected system outside the dark site.

- Prism Central backup, restore, and migration. You cannot perform these operations on MSP-enabled Prism Central.
- Nutanix recommends increasing the MTU to 9000 bytes on the virtual switch vs0 and ensure that the physical networking infrastructure supports higher MTU values (jumbo frame support). The recommended MTU range is 1600-9000 bytes.

  Nutanix CVMs use the standard Ethernet MTU (maximum transmission unit) of 1,500 bytes for all the network interfaces by default. The system advertises the MTU of 1442 bytes to guest VMs using DHCP to account for the extra 58 bytes used by Generic Network Virtualization Encapsulation (Geneve). However, some VMs ignore the MTU advertisements in the DHCP response. Therefore, to ensure that Flow networking functions properly with

such VMs, enable jumbo frame support on the physical network and the default virtual switch vs0.

If you cannot increase the MTU of the physical network, decrease the MTU of every VM in a VPC to 1442 bytes in the guest VM console.

> Note: Do not change the MTU of the CVM.

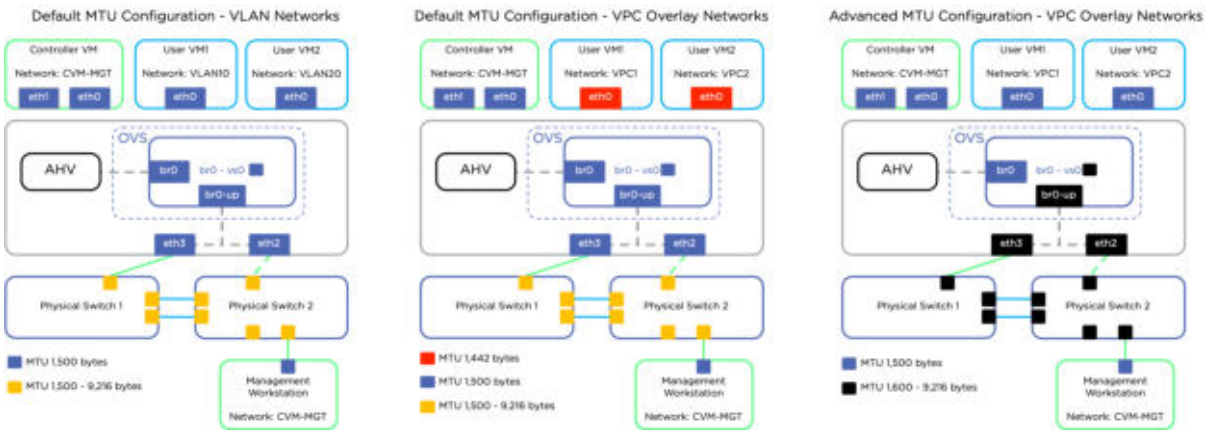


Figure 4: Sample Configurations with and without Higher MTU - VS0, CVM and UVMs

Requirements for Upgrades

The following applies to upgrades of Flow networking network controller (**Advanced Networking** in Prism Control Settings):

- Ensure that the Prism Central host is running an AHV version compatible with the networking controller upgrade version. If necessary, upgrade the AHV version using LCM to the version compatible with the network controller upgrade version.

  > Note:
  >
  > See Compatibility and Interoperability Matrix on the Nutanix Support portal for AOS and Prism Central compatibility.

- Ensure that all the AHV hosts in the AOS cluster are running the version compatible with the network controller upgrade version.

  The network controller upgrade fails if any of the AHV hosts is running an incompatible version.

Limitations

Limitations for Flow networking are as follows.

- Flow networking does not support Flow security for guest VMs.

  You cannot configure rules for Flow security if a guest VM has any NICs connected to VPCs.

- Flow networking is supported only on AHV clusters. It is not supported on ESXi or Hyper-V clusters.

- Flow networking is not enabled on the new PE cluster registering with the Flow networking-enable Prism Central if the Prism Element cluster has an incompatible AHV version.

- Flow networking does not support updating a VLAN-backed subnet as an external subnet.

  You cannot enable the external connectivity option in the Update Subnet dialog box. Therefore, you cannot modify an existing VLAN-backed subnet to add external connectivity.

  VLAN backed subnets for external connectivity are managed by the Flow networking control plane. Traditional AHV VLAN IPAM networks are managed by acropolis.

  > Note: Do not configure the same VLAN as both a Flow networking external network and an AHV IPAM network, as this can lead to IP address conflicts.

- Flow networking cannot be disabled if any external subnets and VPCs are in use. Delete the external subnets and VPCs and then disable Flow Networking.
- Disaster Recovery backup and migration: CMSP-enabled Prism Central does not support disaster recovery backup and migration operations both as a source and target host.

5

FLOW NETWORKING CONFIGURATIONS

Enabling Flow Networking

Before you begin

Ensure tha microservices infrastructure is enabled on Prism Central. See *Enabling Micro Services Infrastructure* section in the *Prism Central Guide*.

About this task

Before you proceed to enable Flow Networking by enabling the **Advanced Networking** option, see Prerequisites for Enabling Flow Networking on page 14.

To enable Advanced Networking, go to **Prism Central Settings** > **Advanced Networking** and do the following.

Procedure

- In the **Advanced Networking** pane, click **Enable**.

  Ensure that the prerequisites specified on the pane are fulfilled.

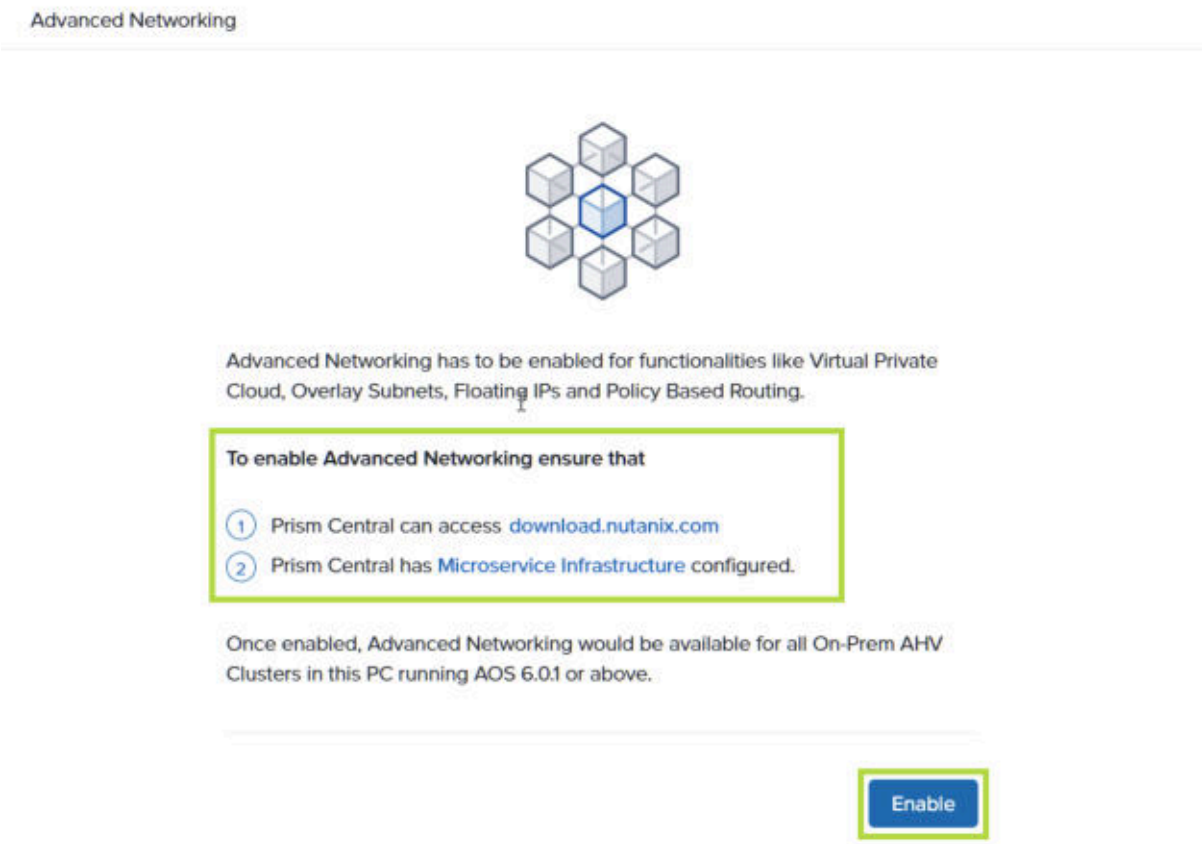


Figure 5: Enabling Flow Networking

Prism Central displays the deployment in-progress.

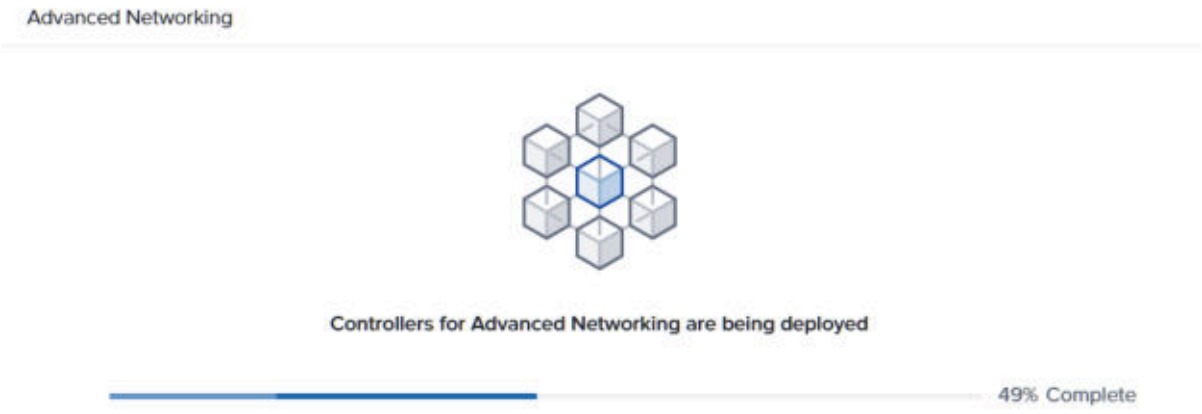


Figure 6: Deployment Progress

- Flow Networking is enabled.

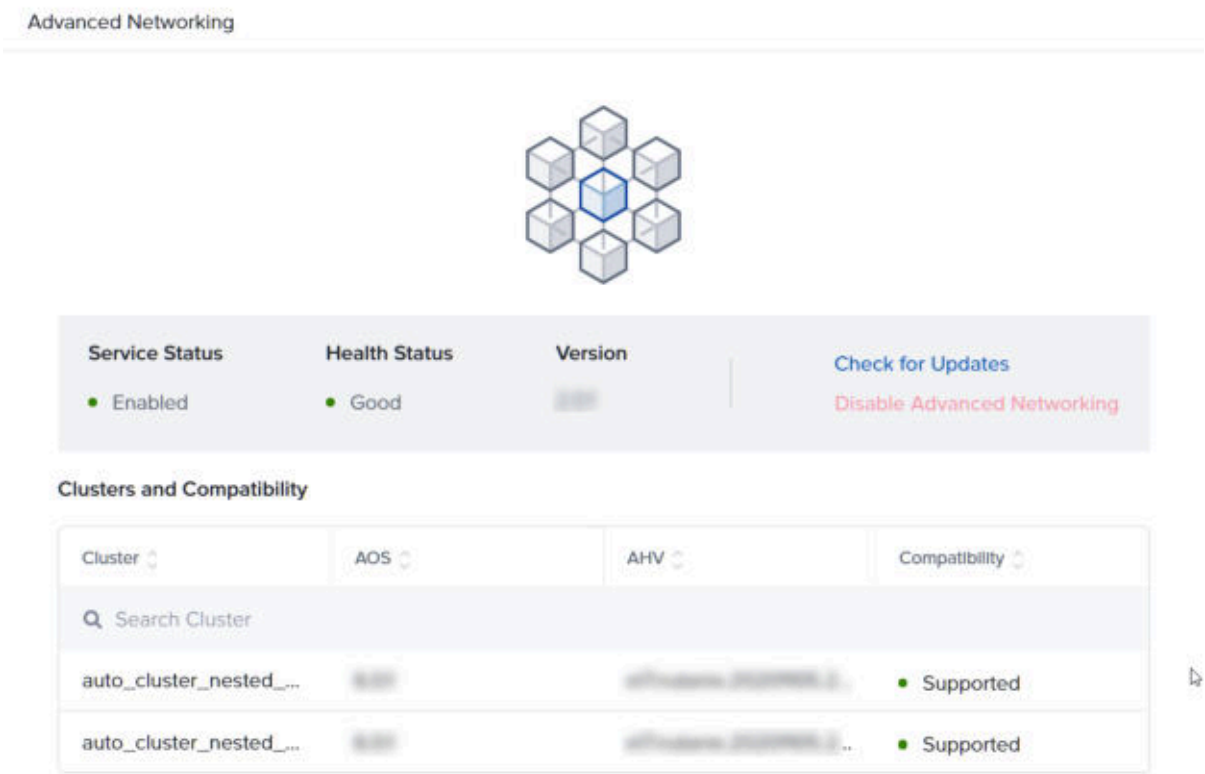


Figure 7: Flow Networking Status

Disabling Flow Networking

About this task

You can disable Flow Networking. However, the network controller cannot be disabled if any external subnets and VPCs are in use. Delete the subnets and VPCs before you disable advanced networking.

> Note:
>
> Flow Networking cannot be disabled if any external subnets and VPCs are in use. Delete the external subnets and VPCs and then disable Flow Networking.

To disable Flow Networking, do the following.

Procedure

1. On the Advanced Networking page, click **Disable**.

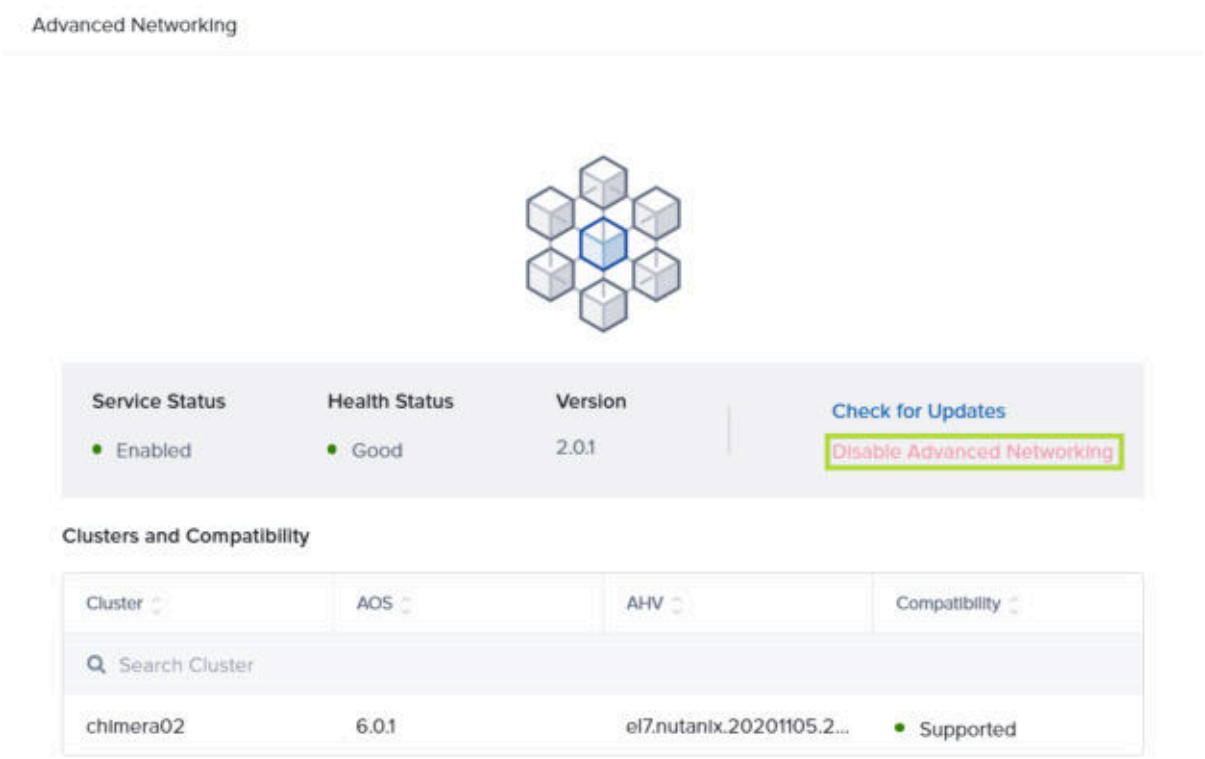


Figure 8:

2. On the confirmation message box, click **Confirm** to confirm disablement.

   To exit without disabling the **Advanced Networking** controller, click **Cancel**.

Unregistering a PE from the PC

Before unregistering a Prism Element from PC, disable Flow Networking on that Prism Element using network controller CLI (or atlas_cli).

About this task

When Flow Networking is enabled on a Prism Central, it propagates the capability to participate in VPC networking to all the registered Prism Elements that are running the required AHV version.

In cases where there are VMs on the Prism Element attached to the VPC network, or if the Prism Element is used to host one or more of the external VLAN networks attached to a VPC, Prism Central alerts you with a prompt. When being alerted about the aforementioned conditions, close the CLI and make adequate configuration to resolve the condition (for example, select a different cluster for the external VLAN network and delete the VMs attached to the VPC network running on the Prism Element). After making such configurations, execute the network controller CLI to disable Flow Networking. If the command goes through successfully, it is safe to unregister the Prism Element.

For example, in a deployment of three Prism Elements - PE1, PE2 and PE3 - registered to the Flow Networking-enabled PC, you want to unregister PE3 from the PC. You must first disable Flow Networking using the following steps:

Procedure

1. SSH to PE3.
2. Run the `ncli cluster info` or `ncli cluster get-params` command to get the cluster parameters.

   Copy the cluster UUID (For example: 017457d3-1012-465c-9c54-aa145f2da7d9) from the displayed cluster parameters.
3. SSH to the Prism Central VM.
4. Open the network controller console by executing the `atlas_cli` command.

   ```
   nutanix@cvm$ atlas_cli
   ```

   ```
   <atlas>
   ```
5. Execute the `config.add_to_excluded_clusters <cluster uuid>` command, providing the cluster UUID that you copied earlier.

   An example of the PC alert, for the condition that PE3 VM is attached to an external network, is as follows:

   ```
   <atlas> config.add_to_excluded_clusters 0005bf8d-2a7f-3b2e-0310-d8e34995511e
   Cluster 0005bf8d-2a7f-3b2e-0310-d8e34995511e has 1 external subnet,
   which will lose connectivity. Are you sure? (yes/no)
   ```

   > Note: To enable Flow Networking on the cluster, execute the `config.remove_from_excluded_clusters <cluster uuid>` command, providing the cluster UUID.

What to do next

To verify if Flow Networking is disabled or enabled, SSH to PE3 and run the `acli atlas_config.get` command.

The output displays the `enable_atlas_networking` parameter as `False` if Flow Networking is disabled and as `True` if Flow Networking is enabled on the Prism Element.

```
nutanix@cvm$ acli atlas_config.get
config {
  anc_domain_name_server_list: “10.10.10.10”
  enable_atlas_networking: False
  logical_timestamp: 19
  minimum_ahv_version: “20190916.101588"
  ovn_cacert_path: “/home/certs/OvnController/ca.pem”
  ovn_certificate_path: “/home/certs/OvnController/OvnController.crt”
  ovn_privkey_path: “/home/certs/OvnController/OvnController.key”
  ovn_remote_address: “ssl:anc-ovn-external.default.anc.aj.domain:6652"
}
```

You can now unregister the PE from the PC.

Upgrading Flow Networking

You can upgrade the Flow networking controller (*Advanced Networking Controller* in *Prism Central Settings*) using Life Cycle Manager (LCM) on Prism Central.

Before you begin

See Prerequisites for Enabling Flow Networking on page 14.

In case of upgrading the Flow networking controller in a dark site, ensure that LCM is configured to reach the local web server that hosts the dark site upgrade bundles.

> Note:
>
> The network controller upgrade fails to start after the pre-check if one or more clusters have Flow Networking enabled and are running an AHV version incompatible with the new network controller upgrade version.

About this task

To upgrade the network controller using LCM, do the following.

Procedure

1. Choose one of the following ways to reach the LCM page:
   - » Go to **Administration** > **LCM** > **Inventory**
   - » Click **Update via LCM** on the Advanced Networking page.

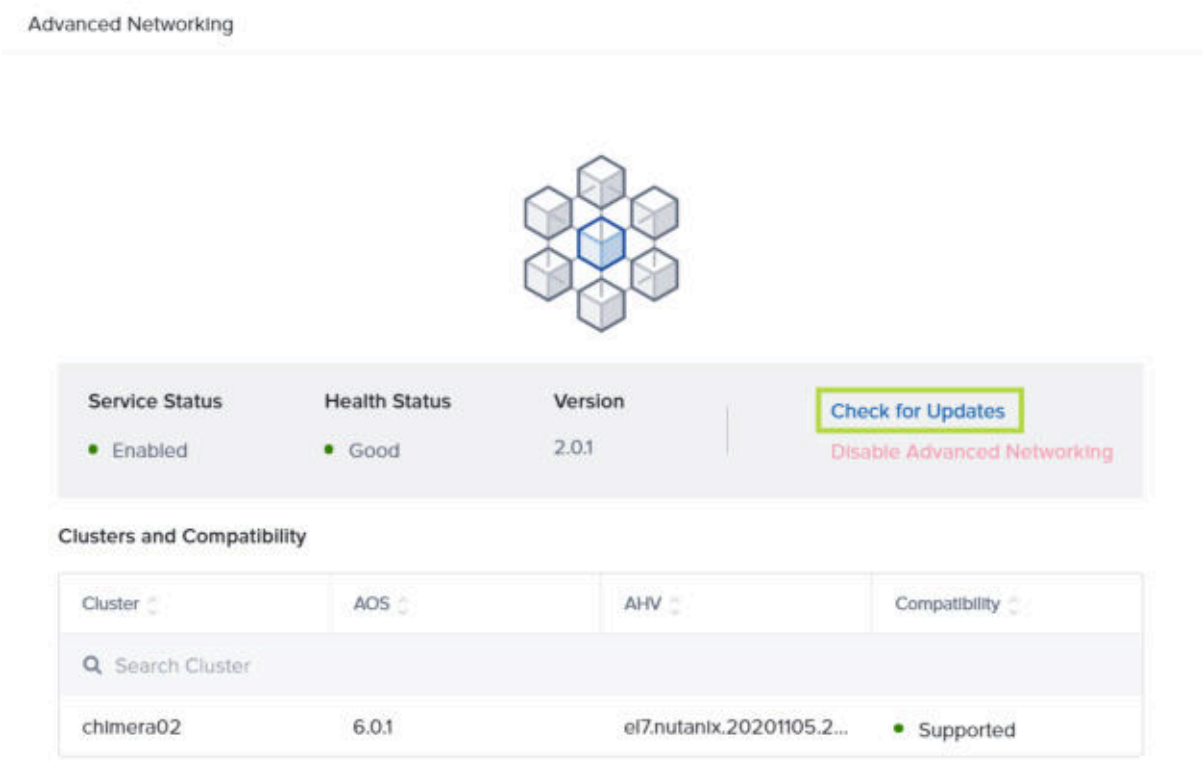


Figure 9: Update via LCM

2. Click **Perform Inventory**.

   When you click **Perform Inventory**, the system scans the registered Prism Central cluster for software versions that are running currently. Then it checks for any available upgrades and displays the information on the LCM page under Software.

3. Go to **Updates** > **Software**. Select the `Advanced Networking Controller` version you want to upgrade to and click **Update**.

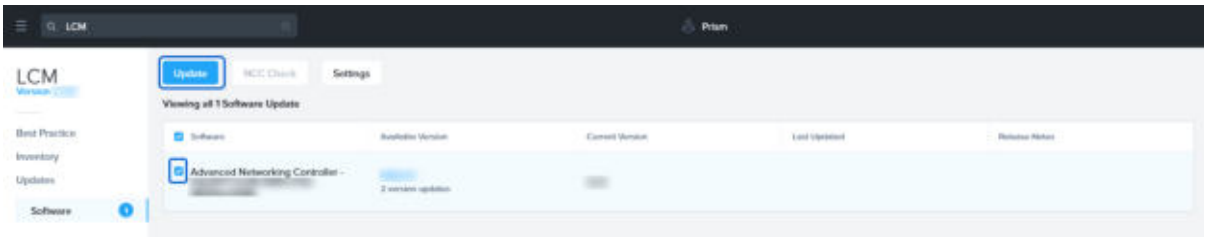


Figure 10: Networking Controller version

Deploying Flow Networking at a Dark Site

About this task

Dark sites are primarily on-premises installations which do not have access to the internet. Such sites are disconnected from the internet for a range of reasons including security. To deploy Flow Networking at such dark sites, you need to deploy the dark site bundle at the site.

This dark site deployment procedure includes downloading and deploying MSP and the network controller bundles.

Before you begin

- See Prerequisites for Enabling Flow Networking on page 14.
- You need access to the Nutanix Portal from an Internet-connected device to download the following dark site bundles:

  > Note: For dark site deployments, Nutanix provides a dark site bundle, which has the Docker images (normally hosted on ECR) and the network controller package (normally hosted on LCM portal). These dark site bundles can be downloaded using an internet-connected system outside the dark site.

  - MSP dark site bundle: https://portal.nutanix.com/page/downloads/list > Microservices Platform (MSP)
  - Flow Networking network controller dark site bundle: See the *Flow Networking Release Notes* for the link to download the dark site bundle.
  - Network Gateway bundle: See the *Flow Networking Release Notes* for the link to download the dark site bundle with checksum text file. Also, see *KB-12393*.

To deploy Flow Networking at a dark site, do the following.

Procedure

1. Upgrade microservices infrastructure.

   See the *Microservices Infrastructure* section in the *Prism Central Guide* for details.

2. Start a web server to host the dark site bundles and act as a source for the LCM downloads, if one is not already created.

   The web server (for example, named **airgap**) can be a virtual machine on a cluster at the dark site. All the Prism Central VMs at the darksite must have access to this web server. This web server is used when you deploy any dark site bundle including the network controller darksite bundle.

   For more information about the server installation, see:

   - Linux web server
   - Windows web server

3. Extract or unpack the networking controller dark site bundle on the web server that you started.

   After unpacking, check if the system shows a directory path that includes the following as per the example: **builds/msp-builds/msp-services/464585393164.dkr.ecr.us-west-2.amazonaws.com/nutanix-msp/atlas-hermes/**.

4. Verify the configuration by running the following command:

   Using the example: `nutanix@cvm$ mspctl controller airgap get`

   Where, *airgap* is the name of the web server. Replace *airgap* with the name of the web server you created.

5. Run the following command after unpacking to ensure that the file permissions are not disrupted during the unpacking:

   » Linux.

   ```
   chmod -R +r builds
   ```

   » Windows NTFS.

   ```
   $> takeown / R / F *
   $> icacls <Build-file-path> /t /grant:F
   ```

   .

6. Enable Flow Networking. See Enabling Flow Networking on page 18.

Troubleshooting Tips

This section provides information to assist troubleshooting of Flow Networking deployments. This is in addition to the information that the "Prism Central Guide" provides.

Audit Logs

Prism Central generates audit logs for all the flow networking activities like it does for other activities on Prism Central. See *Audit Summary View* in the *Prism Central Guide*, for more information about Audit log.

Support Bundle Collection

To support troubleshooting for Flow Networking, you can collect logs.

To collect the logs, run the following commands on the Prism Central VM console:

```
nutanix@cvm$ logbay collect -t msp,anc
```

An example of the command is as follows:

```
nutanix@cvm$ logbay collect -t msp,anc -O
 msp_pod=true,msp_systemd=true,kubectl_cmds=true,persistent=true --duration=-48h0m0s
```

Where:

- `-t` flag indicates the tags to collect
  - *`msp`* tag will collect logs from the services running on MSP pods and persistent log volumes (application-level logs)
  - *`anc`* tag will collect the support bundle, which includes database dumps and OVN state
- `-O` flag adds tag-level options
  - `msp_pod=true` collects logs from MSP service pods

    On the PC, these logs can be found under `/var/log/containers`.
  - `persistent=true` collects persistent log volumes (application-level logs for ANC)

    On the PC, these can be found under /var/log/ctrlog
  - `kubectl_cmds=true` runs `kubectl` commands to get the Kubernetes resource state
- `--duration` sets the duration from the present to collect

The command run generates a zip file at a location, for example: `/home/nutanix/data/logbay/bundles/<filename>.zip`

Unzip the bundle and you'll find the *`anc`* logs under a directory specific to your MSP cluster, the worker VM where the pod is running, and the logging persistent volume of that pod. For example:

`./msp/f9684be8-b4e8-4524-74b4-076ed53ca1fd/10.48.128.185__worker_master_etcd/persistent/default/ovn/anc-ovn_StatefulSet/`

For more information about the task run, see the text file that the command generates at a location, for example:`/home/nutanix/data/logbay/taskdata/<taskID>/collection_result.txt`

For more information about the `logbay collect` command, see the *Logbay Log Collection (Command Line)* topic in the *Nutanix Cluster Check Guide* (NCC Guide).

Layer 2 Virtual Subnet Extension Alert

The `L2StretchLocalIfConflict` alert (Alert with Check ID - 801109) may occur while performing Layer 2 virtual subnet extensions. See KB-10395 for more information about its resolution.

6

NETWORK GATEWAY UPGRADES

Nutanix deployment can detect and install upgrades for the onprem Nutanix Gateways.

For information about identifying the current Nutanix Gateway version, see Identifying the Gateway Version on page 28.

For onprem Nutanix Gateways, the upgrades need to be detected and installed on the respective PC on which each Nutanix Gateway is installed.

For more information, see Detecting Upgrades for Gateways on page 29.

When PC detects the upgrades, it displays a banner on the Gateways tab of the Connectivity page. The banner notifies you that a Gateway upgrade is available after you have run LCM inventory. The table on the Gateways tab also displays an alert (exclamation mark) icon for the network gateways that the upgrade applies to. The hover message for the icon informs you that an upgrade is available for that Gateway.

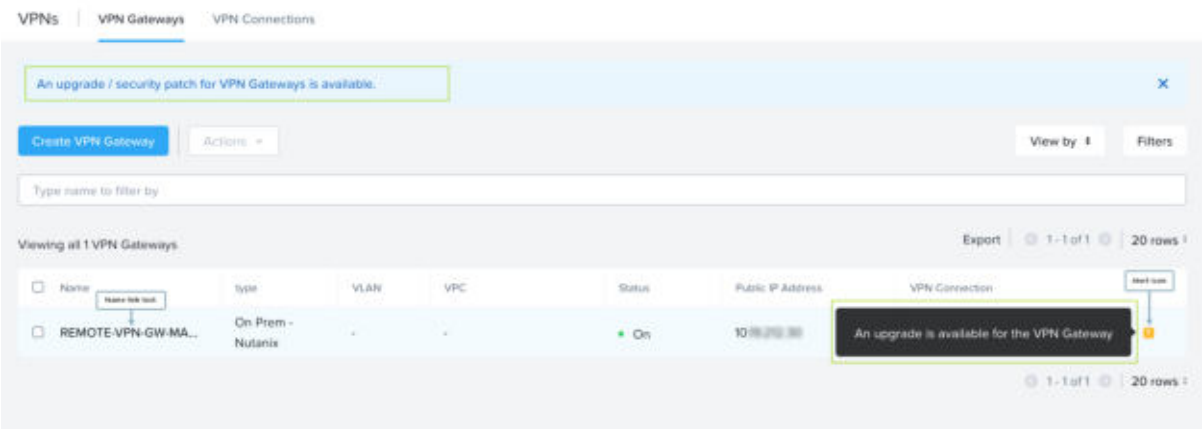


Figure 11: Upgrade Banner

For more information about the upgrade procedure, see Upgrading the PC-managed Onprem Nutanix VPN Gateways on page 29.

Identifying the Gateway Version

About this task

To identify the current Nutanix Gateway version, do the following:

Procedure

- Click the hamburger icon and **Networking & Security** > **Connectivity**.

- On the Gateways tab, click the Gateway name link text to open the Gateway details page.

  In the Gateway table, the VPN Gateway name is a clickable link text.

  The **Gateway Version** is listed in the **Properties** widget.

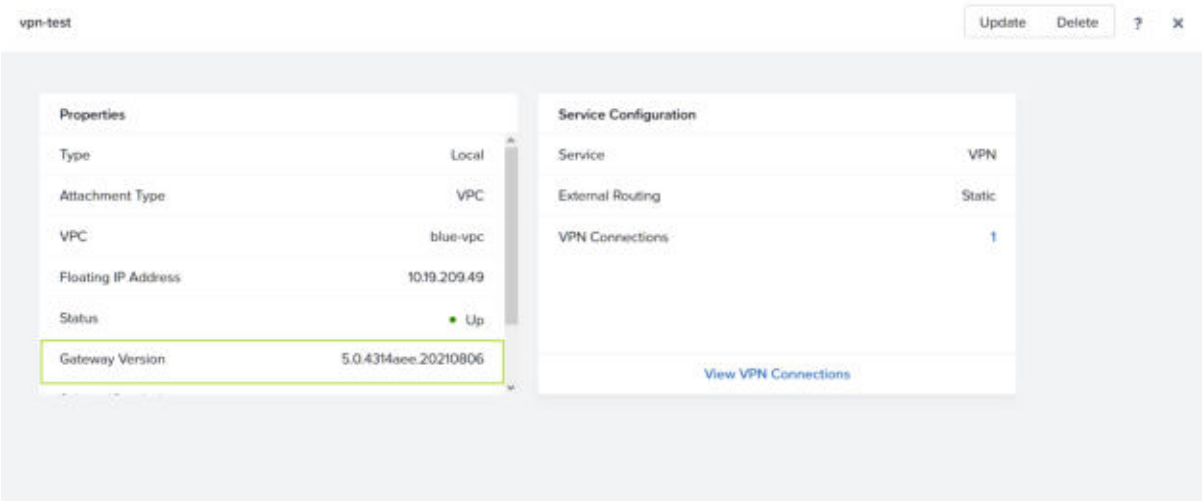


Figure 12: Gateway Version

Detecting Upgrades for Gateways

About this task

Prism Central can detect whether new Gateway upgrades are available, or not, for Nutanix Gateways using LCM. You can then install the upgrade.

Procedure

- Click the hamburger icon of Dashboard.
- Click **Administration** > **LCM** > **Inventory**.
- Click **Perform Inventory**.

  > Note:
  >
  > Nutanix recommends that you select **Enable LCM Auto Inventory** in the LCM page in Prism Central to continuously detect new Gateway upgrades as soon as they are available.

  The upgrade notification banner is displayed on the Gateways page.

Upgrading the PC-managed Onprem Nutanix VPN Gateways

About this task

Perform upgrades of PC-managed Nutanix Gateways using the respective PC on which the Gateway is created.

To upgrade the on-prem Nutanix Gateways, do the following:

Procedure

1. Log on to the Prism Central as the admin user and click the gear icon.
2. Go to Administration > LCM > Inventory.
3. Click **Perform Inventory**.

   When you click **Perform Inventory**, the system scans the registered Prism Central cluster for software versions that are running currently. Then it checks for any available upgrades and displays the information on the LCM page under Software.

   > Note:
   >
   > Skip this step if you have enabled auto-inventory in the LCM page in Prism Central.

4. Go to Updates > Software. Select the Gateway version you want to upgrade to and click Update.

   LCM upgrades the Gateway version. This process takes sometime.

7

NETWORK AND SECURITY VIEW

The **Network and Security** category in the **Entities Menu** expands on-click to display the following networking and security entities that are configured for the registered clusters:

> Note: In

1. Subnets: This dashboard displays the subnets and the operations you can perform on subnets.
2. Virtual Private Clouds: This dashboard displays the VPCs and the operations you can perform on VPCs.
3. Floating IPs: This dashboard displays a list of floating IP addresses that you are using in the network. It allows you to request for floating IP addresses from the free pool of I addresses available to the clusters managed by the Prism Central instance.
4. Connectivity: This dashboard allows you to manage the following networking capabilities:

   - Gateways: This tab provides a list of network Gateways you have created and configured, and the operations you can perform on the network Gateways. You can check and upgrade the Gateway bundle in **Administration** > **LCM** > **Inventory**.
   - VPN Connections: This tab provides a list of VPN connections you have created and configured, and the operations you can perform on VPN connections.
   - Subnet Extensions: This tab provides a list of subnets that you have extended at the Layer 2 level using VPN (point-to-point over Nutanix VPN) or VTEP (point-to-multi-point including third party).
5. Security Policies: This dashboard provides a list of security policies you configured using Flow Segmentation. For information about Security Policies, see the *Flow Microsegmentation Guide*.

See "Network Connections" section for information on how to configure network connections.

Subnets (Overlay IP subnets), Virtual private clouds, floating IPs, and Connectivity are Flow Networking features. These features support flexible app-driven networking that focuses on VMs and applications instead of virtual LANs and network addresses. Flow Networking powers network virtualization to offer a seamless network experience with enhanced security. It is disabled by default. It is a software-defined network virtualization solution providing overlay capabilities for the on-premises AHV clusters.

Security policies drives the Flow Segmentation features for secure communications. See *Flow Microsegmentation Guide*.

Subnets

Manage subnets in the **List** view of `Subnets` dashboard in the `Network and Security` section.

To access the Subnets dashboard, select **Subnets** from the entities menu in Prism Central. The Subnets dashboard allows you to view information about the subnets configured for the registered clusters.

Note: This section describes the information and options that appear in the Network and Security dashboard. See “” () for instructions on how to view and organize that information in a variety of ways.

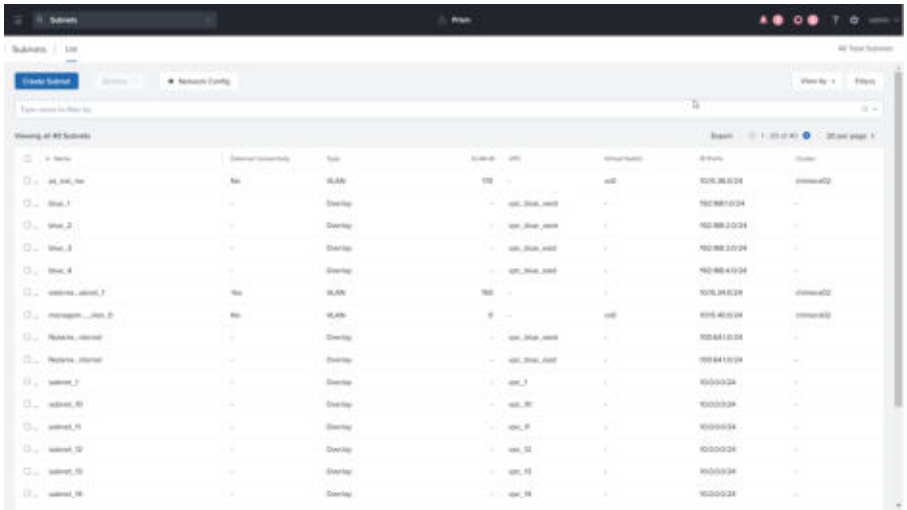


Figure 13: Subnets Dashboard

The following table describes the fields that appear in the subnets list. A dash (-) is displayed in a field when a value is not available or applicable.

Table 2: Subnets Dashboard Fields

Parameter	Description	Values
Name	Displays the subnet name.	(subnet name)
External Connectivity	Displays whether or not the subnet has external connectivity configured.	(Yes/No)
Type	Displays the subnet type.	VLAN
VLAN ID	Displays the VLAN identification number.	(ID number)
VPC	Displays the name of the VPC that the Subnet is used in.	(Name of VPC)
Virtual Switch	Displays the virtual switch that is configured for the VLAN you selected. The default value is the default virtual switch `vs0`. Note: The virtual switch name is displayed only if you add a VLAN ID in the VLAN ID field.	(virtual switch name)
IP Prefix	Displays the IPv4 Address of the network with the prefix.	(IPv4 Address/Prefix)
Cluster	Displays the name of the cluster for which this subnet is configured.	(cluster name)

Parameter	Description	Values
Hypervisor	Displays the hypervisor that the subnet is hosted on.	(Hypervisor)

To filter the list by network name, enter a string in the filter field. (Ignore the **Filters** pane as it is blank.)

To view focused fields in the List, select the focus parameter from the **Focus** drop down list. You can create your own customised focus parameters by selecting **Add custom** from the drop down list and selecting the necessary fields after providing a **Name**, in the Subnet Columns.

There is a **Network Config** action button to configure a new network (see “” ()). The **Actions** menu appears when one or more networks are selected and includes a **Manage Categories** option (see “” ()).

Go to the Subnets list view by clicking **Network and Security** > **Subnets** on the left side-bar.

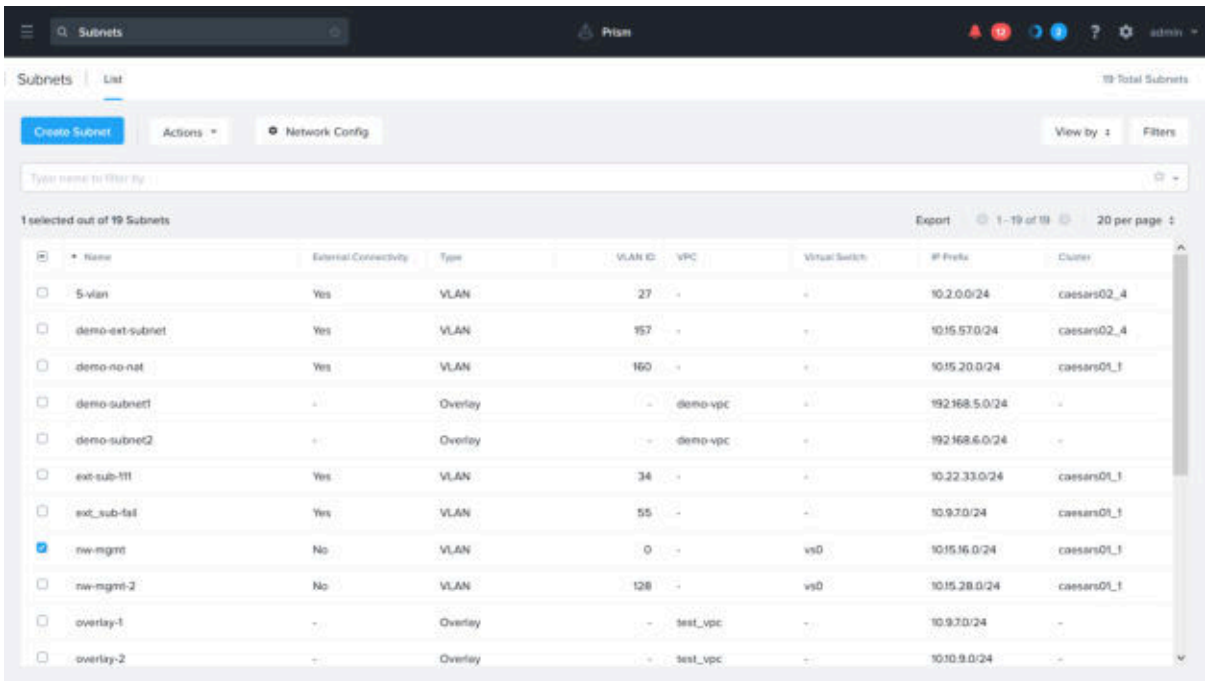


Figure 14: Subnets Page

To view or select actions you can perform on a subnet, select the subnet and click the **Actions** dropdown.

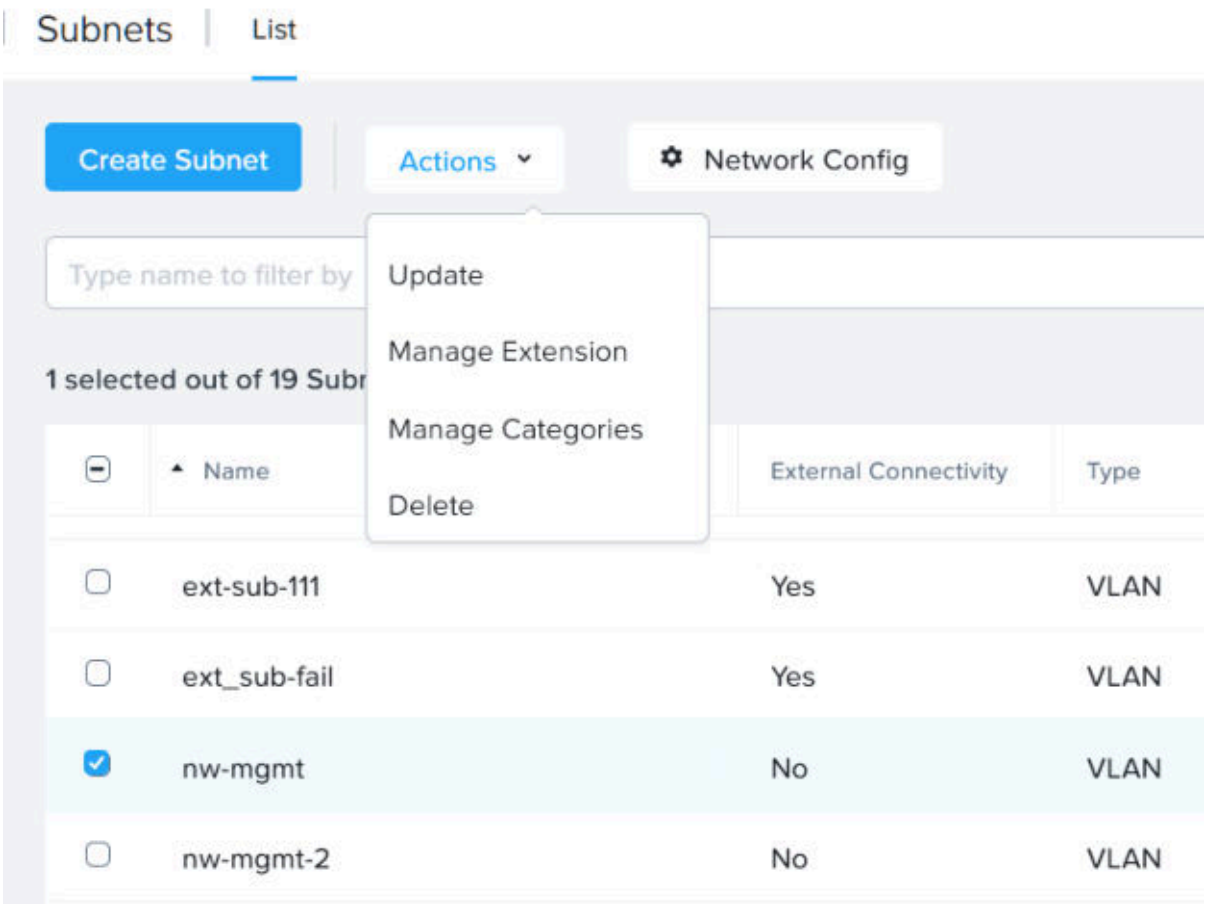


Figure 15: Subnet Actions

Table 3: Subnet Actions

Action	Description
Update	Click this action to update the selected subnet. see Updating a Subnet on page 85.
Manage Extension	Click this action to create a subnet extension. A subnet extension allows VMs to communicate over the same broadcast domain to a remote Xi availability zone (in case of Xi-Leap based disaster recovery) via the extension.
Manage Categories	Click this action to associate the subnet with a category or change the categories that the subnet is associated with.
Delete	Click this action to delete the selected subnet. See Deleting Subnets, Policies or Routes on page 90.

You can also filter the list of subnets by clicking the **Filters** option and selecting the filtering parameters.

Subnet Summary View

View the details of a subnet listed on the Subnets page.

To view the details of a subnet, click the name of the subnet on the subnet list view.

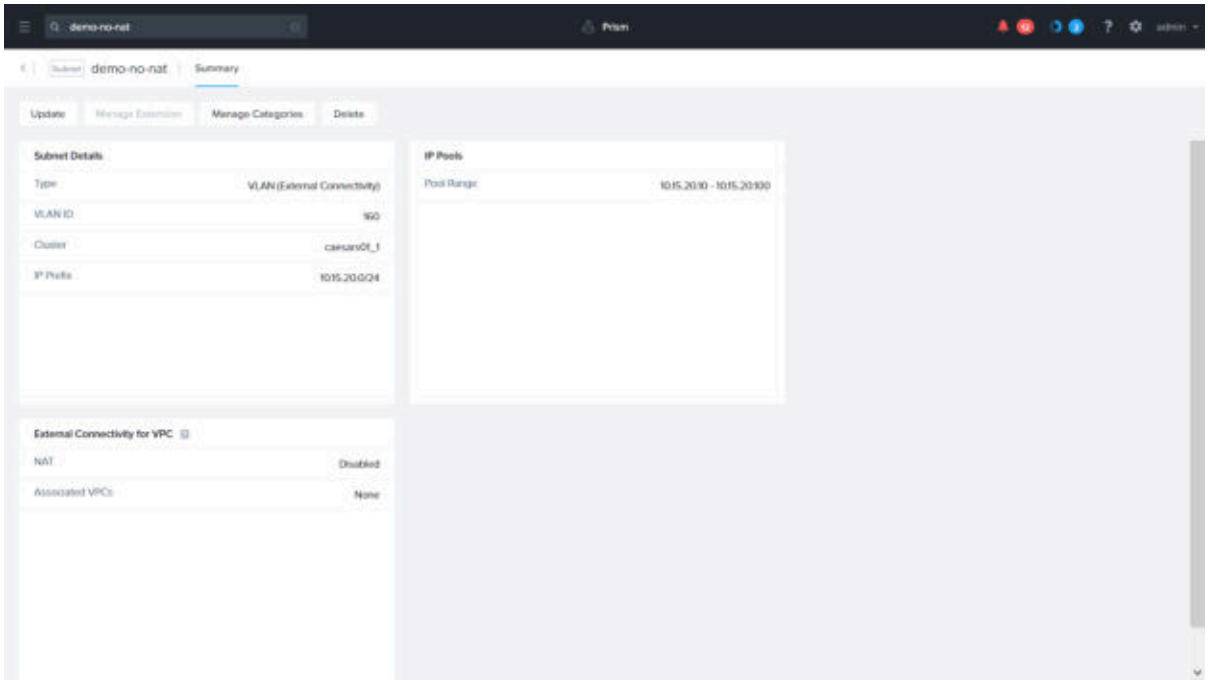


Figure 16: Subnet Summary Page

The Summary page provides buttons for the actions you can perform on the subnet, at the top of the page. Buttons for the following actions are available: **Update**, **Extend**, **Manage Categories**, and **Delete**.

The subnet Summary page has the following widgets:

Widget Name	Information provided
Subnet Details	Provides the following: • Type — Displays the type of network like VLAN or Overlay. • VLAN ID — Displays the VLAN ID. This parameter is displayed only for VLAN networks. • VPC — Displays the VPC name. This parameter is displayed only for Overlay networks. • Cluster — Displays the cluster that the VLAN network is configured on. This parameter is displayed only for VLAN networks. • IP Prefix — Displays the IP address prefix configured for the network. This parameter is displayed for both VLAN and Overlay networks.
IP Pool	Provides the IP address Pool Range assigned to the network.

Widget Name	Information provided
External Connectivity	Provides the following: • NAT — Displays whether NAT is enabled or disabled for VPCs connecting to the network. When you hover on the `Enabled/Disabled` status, the hover message displays details of VPCs connected to the external subnet. • Associated VPCs — Displays the VPCs associated with this external subnet.

Virtual Private Clouds

You can manage Virtual Private Clouds (VPCs) on the `Virtual Private Clouds` dashboard.

Go to the Virtual Private Clouds dashboard by clicking **Network and Security** > **Virtual Private Clouds** on the left side-bar.

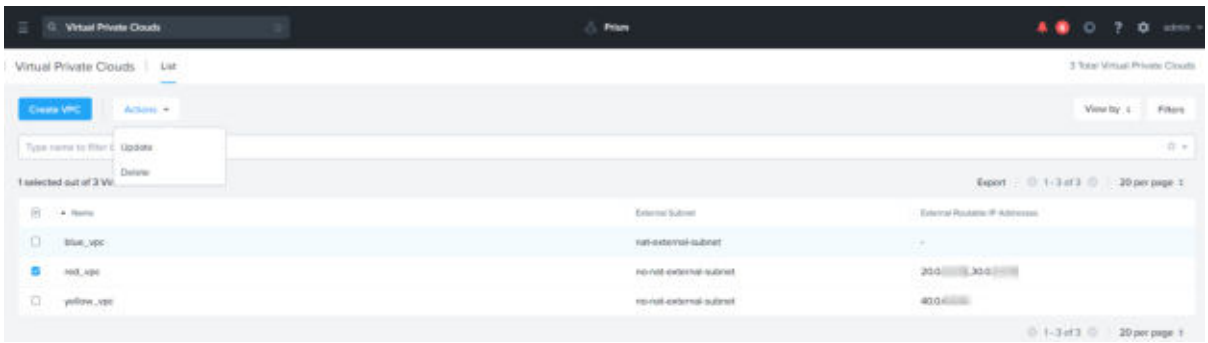


Figure 17: Virtual Private Clouds dashboard

You can configure the table columns for the VPC list table. The available column list includes `Externally Routable IP Addresses` that provides address space within the VPC that is reachable externally without NAT.. For the list of columns that you can add to the list table, see Customizing the VPC List View on page 37.

> Note:
>
> Ensure that the externally routable IP addresses (subnets with external connectivity without NAT) for different VPCs do not overlap.
>
> Configure the routes for the external connectivity subnets with next hop as the Router or SNAT IP address. Also configure the routes on the router for the return traffic to reach the VPC. See **External Connectivity** panel in VPC Details View on page 39.

To view or select actions you can perform on a VPC, select the VPC and click the **Actions** drop down.

You can also filter the list of VPC by clicking the **Filters** option and selecting the filtering parameters.

Customizing the VPC List View

About this task

You can customize the columns in the table. Click the **View by** drop down and select **+ Add custom**.

In the Virtual Network Columns dialog box, do the following.

Procedure

1. Enter a name for the view.

2. Select the columns you want displayed in the table.

   During the column selection, the columns you select are moved under the `Selected Columns` list. The *`Name`* (of the VPC) column is the default column already selected. You can add a maximum of 10 columns (including the *`Name`* column) to the `Selected Column` list.

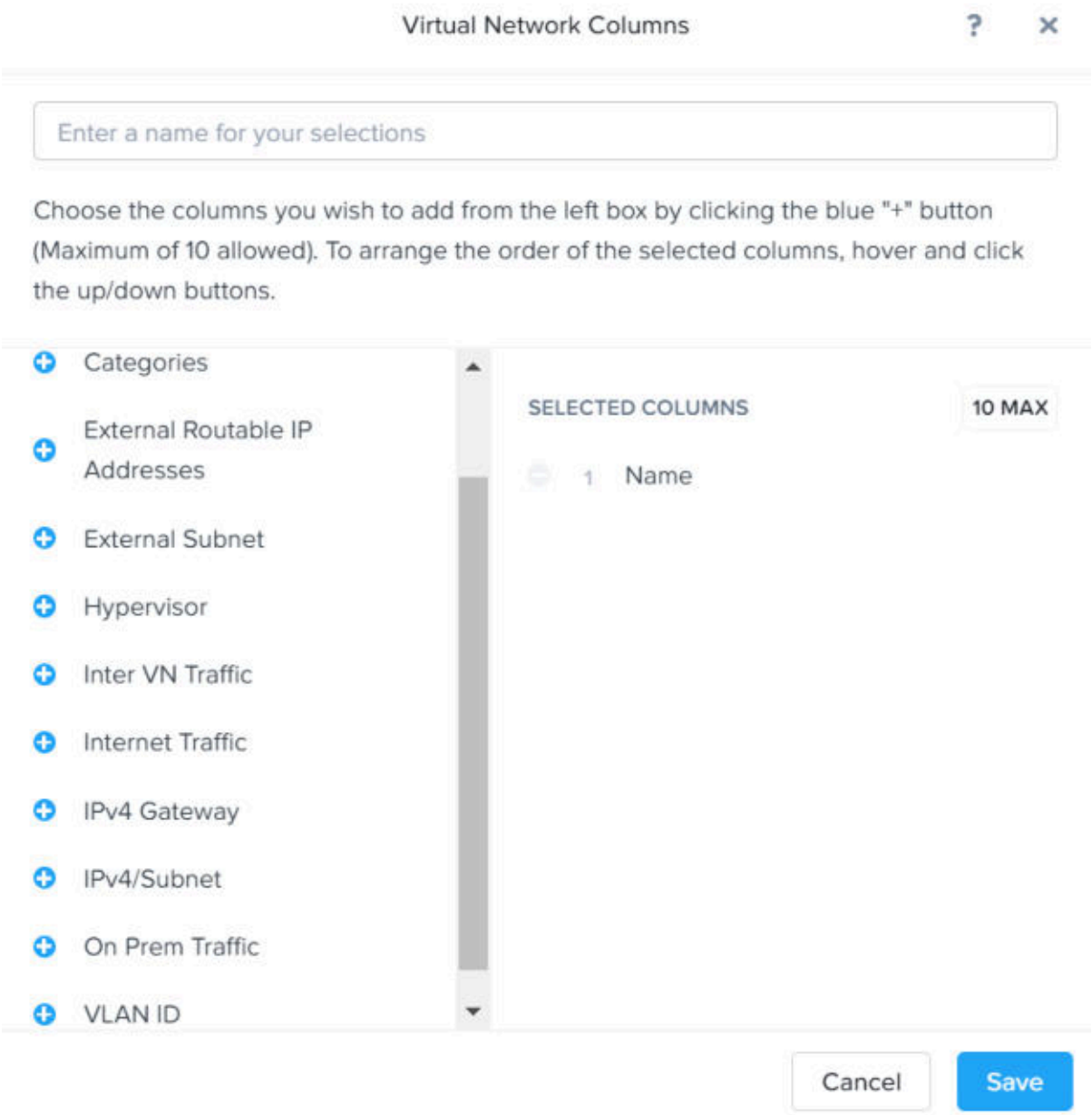


Figure 18: Customizing Columns in VPC View

   To arrange the order of the selected columns, hover on the column name and click the up or down arrow button as appropriate.

3. Click **Save**.

VPC Details View

To view the details of a VPC, click the name of the VPC on the VPC list view.

The VPC details view has the following tabs:

- **Summary**

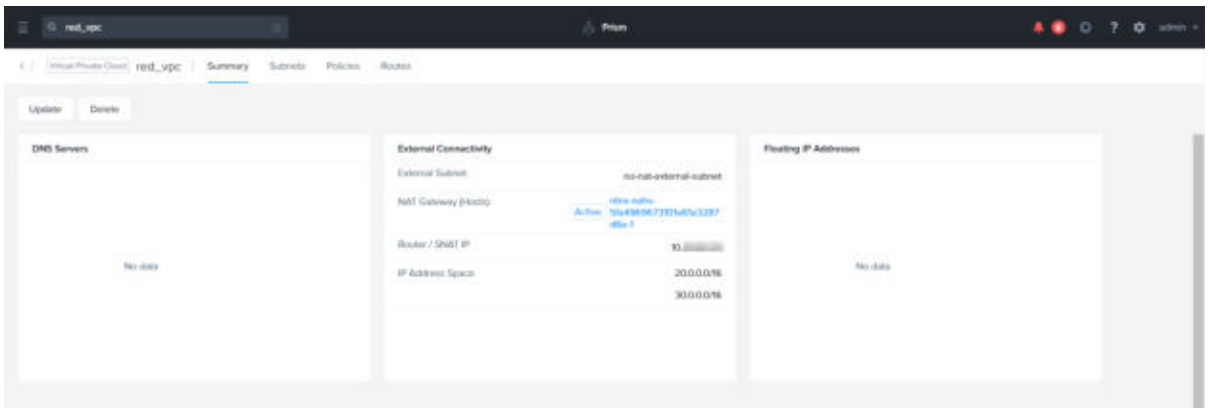


Figure 19: Summary Tab

The Summary tab provides the following panes:

- DNS Servers—Provides more information about the DNS Servers used by the VPC.
- External Connectivity—Provides the name of the external subnet, NAT Gateway host details, router/SNAT IP address and the IP address spaces or ranges configured for the VPC.
- Floating IP Addresses—Provides details of the floating IP addresses that the VPC uses.

- **Subnets**

Figure 20: Subnet Tab

The Subnet tab provides the following information for the subnets:

- Name—Displays the name of the subnet.
- IP Range—Displays the IP address range configured for the subnet.
- DHCP IP Pool—Displays the DHCP IP address pool configured for the subnet.
- Default Gateway IP—Displays the IP address used as the default gateway by the entities in the subnet.
- Actions—Displays the actionable links to **Edit** or **Delete** the subnet.

- **Policies**

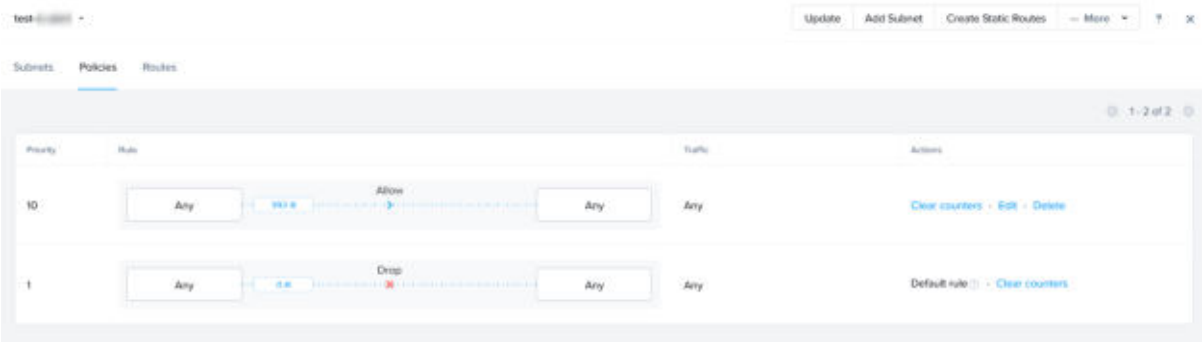


Figure 21: Policies Tab

The Policies tab maps the following information about the security-based traffic shaping policies you configure:

- Priority—The traffic priority.
- Rule—The **Allow** or **Deny** rule set for the priority.
- Traffic—The traffic type that the priority and rule should be applied to.
- Actions—Actions you can take on the policy. You can perform three actions: **Clear counters**, **Edit** the policy or **Delete** the policy.

- **Routes**

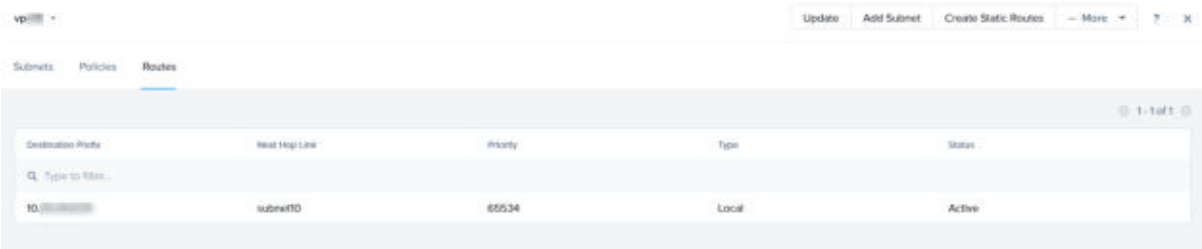


Figure 22: Routes Tab

The Router tab provides the following information about the routes:

The VPC details view has the following configuration options for the VPC:

- **Update**: Use this option to update the VPC. For more information, see Updating Virtual Private Cloud on page 84.
- **Add Subnet**: Use this option to add a subnet to the VPC. For more information, see Creating a Subnet on page 68.
- **Create Static Routes**: Use this option to create a static route. For more information, see Creating Static Routes on page 83.
- **Update Static Routes**: Use this option to update static route configurations that you already created. For more information, see Updating Static Routes on page 89.

- **Create Policy**: Use this option to create traffic policies in addition to the pre-configured default policy. When you create a VPC, there is one default policy that Advanced Networking creates for the VPC. This policy is pre-configured and cannot be edited.

  For more information, see Creating a Policy on page 78.

- **Clear All Counters**: Allows you to clear all the counters for the VPC.
- **Delete**: Allows you to delete the VPC. See Deleting a Virtual Private Cloud on page 90 and Deleting Subnets, Policies or Routes on page 90.

Floating IPs

You can access floating IPs on the **`Floating IPs`** dashboard or list view in the **`Network and Security`** section.

For information about floating IP addresses and their role in Flow Networking, see *SNAT and Floating IP Address* in Essential Concepts on page 8.

Go to the Floating IPs dashboard by clicking **Network and Security** > **Floating IPs** on the left side-bar.

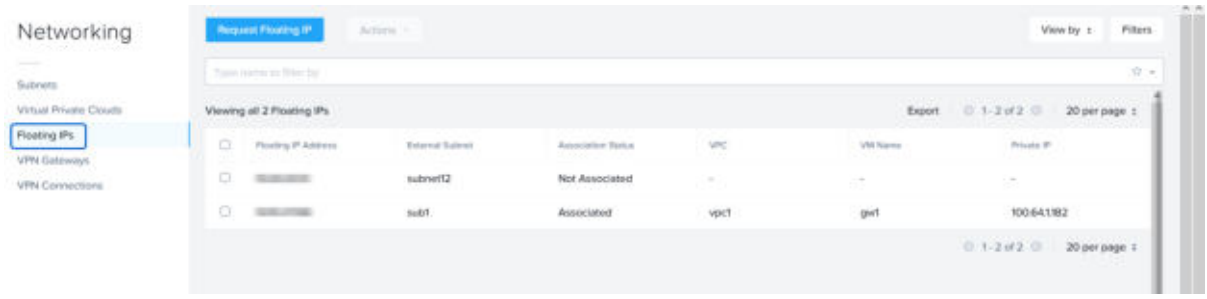


Figure 23: Floating IPs dashboard

To view or select actions you can perform on a floating IP address assigned, select the floating IP address and click the **Actions** drop down. The following actions are available for a selected floating IP address:

- Update—Assign or change the assignment of the floating IP address. You can assign the floating IP address to a IP address such as a private IP address in a VPC or the primary IP address of a VM or a secondary IP address created on a VM.
- Delete—Delete the floating IP address. The deleted IP address returns to the IP address pool as unused. Before you delete, ensure that it is not assigned to a private IP address or a VM. Change the assignment to **`None`** if it is already assigned, using the **Update** action.

> Note: Floating IP addresses are not reachable (Pings fail) unless you associate them to primary or secondary IP addresses of VMs. For information about assigning floating IP addresses to secondary IP addresses of VMs, see Assigning Secondary IP Addresses to Floating IPs on page 62.

To filter the list of floating IP address assignments, click the **Filters** option and select the appropriate filtering parameters.

To request floating IP addresses, see Requesting Floating IPs on page 66.

Connectivity

You can access network Gateways, VPN connections and subnet extensions on the **`Connectivity`** dashboard.

Click **Network & Security** > **Connectivity** to see the Connectivity dashboard.

The Connectivity dashboard opens on the Gateways tab. To see the VPN connections, click the VPN Connections tab. To see the subnets extended across AZs, click the Subnet Extensions tab.

Gateways Summary View

The Connectivity dashboard opens on the Gateways dashboard or summary view.

The **Gateway** dashboard provides a list of gateways created for the clusters managed by the Prism Central.

The **Gateways** dashboard provides a **Create Gateway** dropdown menu that lets you create a **Local** or a **Remote** gateway. You can create a local or remote gateway with **VPN** or **VTEP** service. See Creating a Network Gateway on page 96 for more information.

You can select a gateway from the list (select the checkbox provided for the gateway) and then perform an action provided in the **Actions** dropdown list. The **Actions** dropdown list allows you to **Update** or **Delete** the selected gateway.

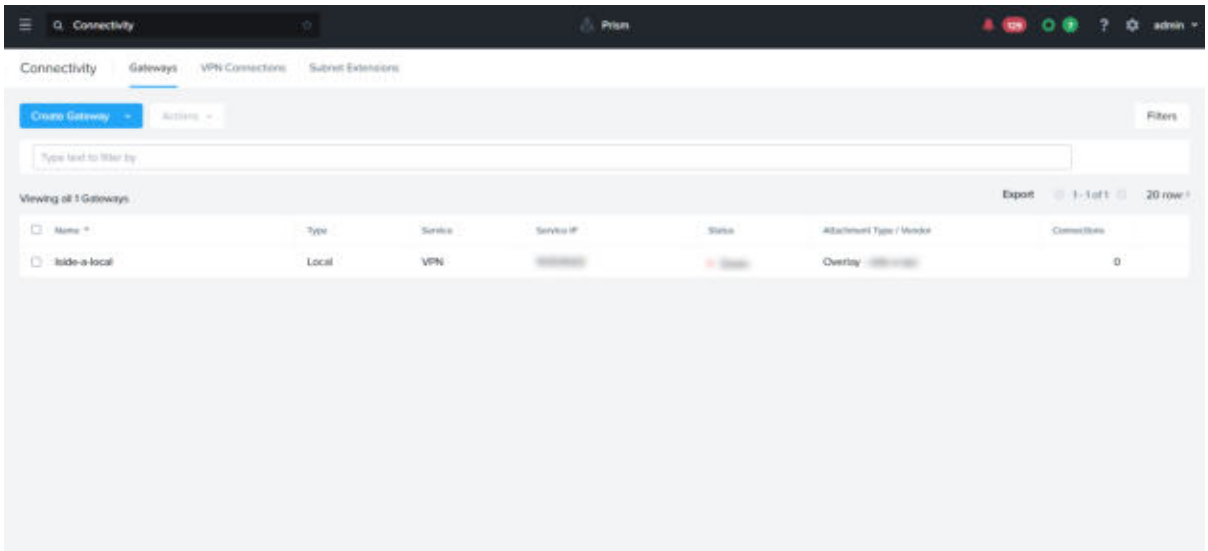


Figure 24: Gateways dashboard

The **Gateway** summary list view provides the following details about the gateway.

Table 4: Gateway List Fields

Parameter	Description	Values
Name	Displays the name of the gateway.	(Name of gateway)
Type	Displays the gateway type.	(Local or Remote)
Service	Displays the service that the gateway uses.	(VPN or VTEP)

Parameter	Description	Values
Service IP	Displays the IP address used by the service.	(IP address)
Status	Displays the operational status of the gateway.	(Up or Down)
Attachment Type/ Vendor	Displays the type of subnet associated with the gateway.	(VLAN or Overlay-VPC name)
Connections	Displays the number of service connections (such as VPN connections) configured and operational on the gateway.	(number)

You can click the name of a gateway to open the gateway details page that presents the information about the gateway in widgets.

Gateway Details View

You can click the name of a gateway in the **Gateway** dashboard list to open the gateway details page that presents the information about the gateway in widgets.

The gateway details page displays the name of the gateway on the top left corner.

- On the top right corner, the close button (X) allows you to close the details page.
- The **Update** button opens the **Update Gateway** page. See *Updating Gateways* for more information.
- The **Delete** button allows you to delete the gateway. See *Deleting Gateways* for more information.

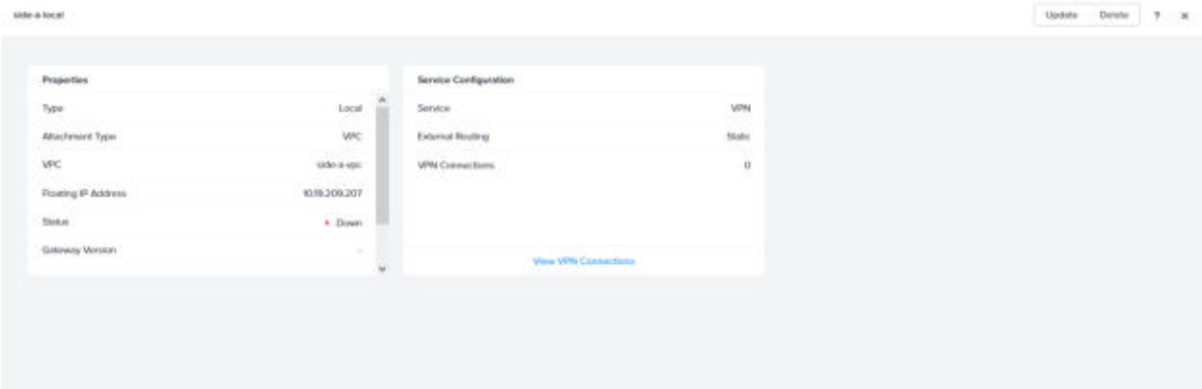


Figure 25: Gateway Details View

The details about the gateway are organized in widgets as follows:

Table 5: Gateway Details

Parameter	Description	Values
Properties widget		
Type	Displays the gateway type.	(Local or Remote)

Parameter	Description	Values
Attachment Type	Displays the network entity like VLAN or VPC that the gateway is attached to.	(VLAN or VPC)
VPC or Subnet (VLAN)	Displays the name of the attached VPC or VLAN subnet.	(Name of VLAN or VPC)
Floating or Private IP Address	Displays the Floating (for VPC) or Private (for VLAN) IP address assigned to the gateway.	(IP Address)
Status	Displays the operational status of the gateway.	(Up or Down)
Gateway Version	Displays the version of the Nutanix gateway appliance deployed.	(Version)
Cluster	Displays the name of the cluster on which the gateway is created.	(Cluster name)
Gateway VM	Displays the name of the VM on which the gateway is created.	(Name of VM - actionable link. Click the name-link to open the VM details page of the gateway VM.)
Service Configuration		
Service	Displays the service used by the gateway.	(VPN or VTEP)
External Routing	Displays the type of routing associated with the gateway for external traffic routing.	(Static or eBGP with ASN)
Internal Routing	Displays the type of routing associated with the gateway for internal traffic routing.	(Static or eBGP with ASN)
VPN Connections	Displays the total number of VPN connections associated with the gateway.	(Number - actionable link. Click the link to open the VPN connection details page for the associated VPN connection.)
View VPN Connections	Click this link to open the VPN Connections tab.	-

VPN Connections Summary View

The Connectivity dashboard allows you to open the VPN Connections dashboard or summary view.

VPN Connection: Represents the VPN IPSec tunnel established between local gateway and remote gateway. When you create a VPN connection, you need to select two gateways between which you want to create the VPN connection.

The **VPN Connections** dashboard provides a list of VPN connections created for the clusters managed by the Prism Central.

The **VPN Connections** dashboard provides a **Create VPN Connection** button that opens the **Create VPN Connection**. See Creating a VPN Connection on page 108 for more information.

You can select a VPN connection from the list (select the checkbox provided for the VPN connection) and then perform an action provided in the **Actions** dropdown list. The **Actions** dropdown list allows you to **Update** or **Delete** the selected VPN connection.

The **VPN Connections** summary list view provides the following details about the VPN connection.

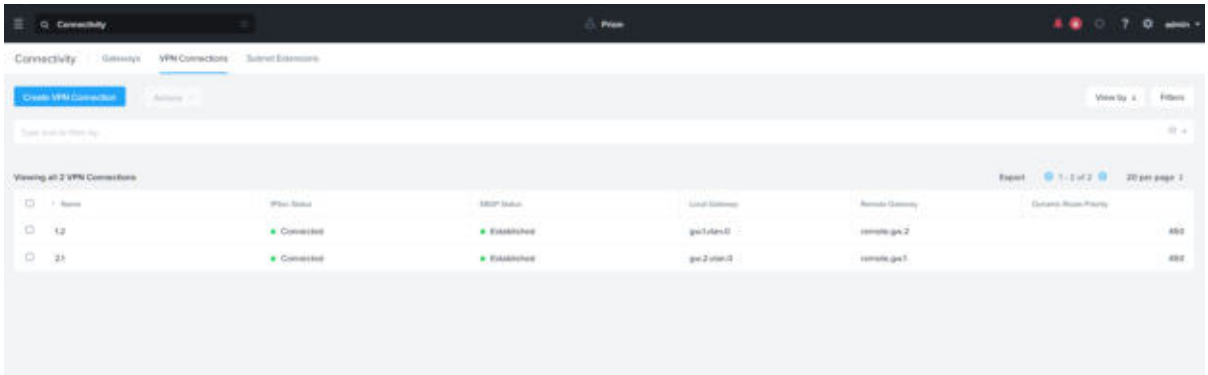


Figure 26: VPN Connections dashboard

Table 6: VPN Connections List Fields

Parameter	Description	Values
Name	Displays the name of the connection.	(gateway name)
IPSec Status	Displays the connection status of IPSec tunnel.	(Connected or Not Connected)
EBGP Status	Displays the status of the EBGP gateway connection.	(Established or Not Established)
Local Gateway	Displays the name of the local gateway used for the connection.	(Name of local gateway)
Remote Gateway	Displays the name of the remote gateway used for the connection.	(Name of remote gateway)
Dynamic Routing Priority	Displays the dynamic routing priority assigned to the connection for throughput management. You can assign any value in the range of 100-1000. Flow networking assigns the first VPN connection the value 500 by default. Thereafter, subsequent VPN connections are assigned values decremented by 50. For example, the first connections is assigned 500, then the second connection is assigned 450, the third one 400 and so on.	(Number in the range of 100-1000. User assigned.)

VPN Connections Details View

You can click the name of a VPN connection in the **VPN Connections** dashboard list to open the VPN connection details page that presents the information about the VPN connection in widgets.

The VPN connection details page displays the name of the VPN connection on the top left corner.

- On the top right corner, the close button (X) allows you to close the details page.
- The **Update** button opens the **Update VPN Connection** page. See Updating VPN Connection on page 109 for more information.
- The **Delete** button allows you to delete the VPN connection. See Deleting a VPN Connection on page 110 for more information.

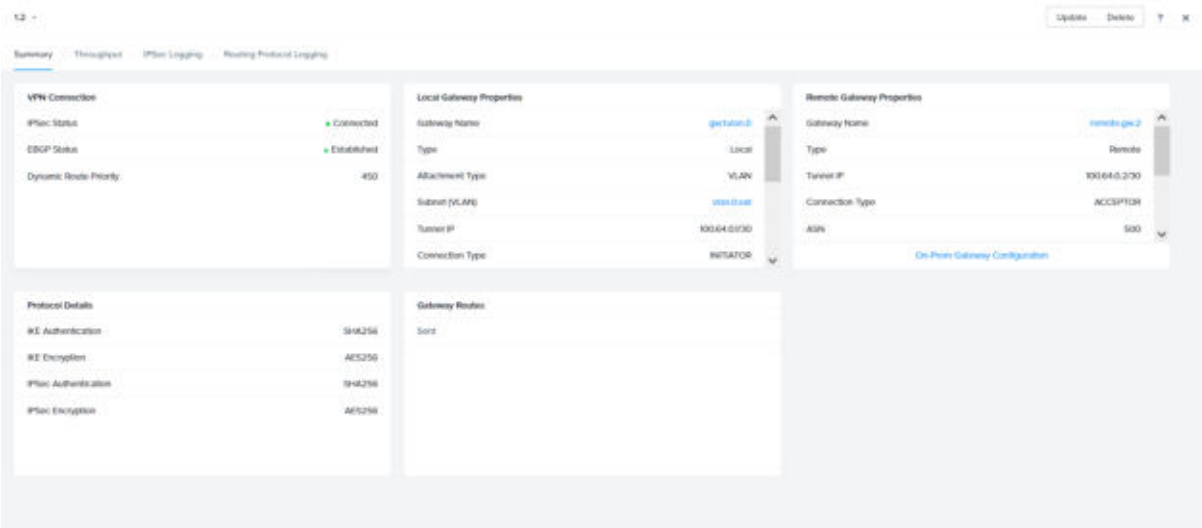


Figure 27: VPN Connection Details

The details about the VPN connection are organized in widgets as follows:

- Summary tab—See the *VPN Connection Summary Tab Details* table below.
- Throughput tab—See the *VPN Connection Throughput Tab Details* table below.
- IPSec Logging tab—Provides logs for the IPSec tunnel.
- Routing Protocol Logging tab—Provides logs for the routing protocol used in the VPN connection.

Table 7: VPN Connection Summary Tab Details

Parameter	Description	Values
VPN Connection widget		
IPSec Status	Displays the connection status of IPSec tunnel.	(Connected or Not Connected)
EBGP Status	Displays the status of the EBGP gateway connection.	(Established or Not Established)

Parameter	Description	Values
Dynamic Routing Priority	Displays the dynamic routing priority assigned to the connection for throughput management. You can assign any value in the range of 100-1000. Flow networking assigns the first VPN connection the value 500 by default. Thereafter, subsequent VPN connections are assigned values decremented by 50. For example, the first connections is assigned 500, then the second connection is assigned 450, the third one 400 and so on.	(Number in the range of 100-1000. User assigned.)
Local Gateway Properties		
Gateway Name	Displays the name of the local gateway used for the connection.	(Name of local gateway)
Type	Displays the type of gateway.	(Local)
Attachment Type	Displays the network entity like VLAN or VPC that the gateway is attached to.	(VLAN or VPC)
VPC or Subnet (VLAN)	Displays the name of the attached VPC or VLAN subnet.	(Name of VLAN or VPC)
Tunnel IP	Displays the Tunnel IP address of the local gateway.	(IP Address)
Connection Type	Displays the connection type you selected while creating the VPN connection. The connection type may be Initiator or Acceptor of a VPN connection between the local and remote gateways. T	(Initiator or Acceptor)
External Routing	Displays the type of routing associated with the gateway for external traffic routing.	(Static or eBGP with ASN)
Internal Routing	Displays the type of routing associated with the gateway for internal traffic routing.	(Static or eBGP with ASN)
Floating or Private IP Address	Displays the Floating (for VPC) or Private (for VLAN) IP address assigned to the gateway.	(IP Address that you assigned to the local gateway with /30 prefix when you configured the VPN connection.)
Status	Displays the operational status of the gateway.	(Up or Down)
Cluster	Displays the name of the cluster on which the gateway is created.	(Cluster name)
Gateway VM	Displays the name of the VM on which the gateway is created.	(Name of VM - actionable link. Click the name-link to open the VM details page of the gateway VM.)

Parameter	Description	Values
Remote Gateway Properties		
Gateway Name	Displays the name of the remote gateway used for the connection.	(Name of remote gateway)
Type	Displays the type of gateway.	(Remote)
Tunnel IP	Displays the Tunnel IP address of the remote gateway.	(IP Address)
Connection Type	Displays the connection type you selected while creating the VPN connection. The connection type may be Initiator or Acceptor of a VPN connection between the local and remote gateways. T	(Initiator or Acceptor)
External Routing	Displays the type of routing associated with the gateway for external traffic routing.	(Static or eBGP with ASN)
ASN	Displays the ASN of the EBGP route. This information is only displayed if you configured EBGP as the External Routing protocol.	(Number)
Vendor	Displays the name of the vendor of the gateway appliance at the remote site.	(Name of vendor of gateway appliance)
External IP	Displays the IP address assigned to remote the gateway.	(IP Address that you assigned to the remote gateway with /30 prefix when you configured the VPN connection.)
Status	Displays the operational status of the gateway.	-
Protocol Details		
Service	Displays the service used by the gateway.	(VPN or VTEP)
Gateway Routes	Displays the status of the routes used by the gateways.	(Sent)

Subnet Extensions Summary View

The Connectivity dashboard opens on the Subnet Extensions dashboard or summary view.

The **Subnet Extensions** dashboard provides a list of subnet extensions created for the clusters managed by the Prism Central.

The **Subnet Extensions** dashboard provides a **Create Subnet Extension** dropdown menu that lets you extend a subnet **Across Availability Zones** or **To a Third Party Data Center**. You can extend a subnet using **VPN** or **VTEP** service. See Layer 2 Virtual Network Extension on page 110 for more information.

You can select a subnet extension from the list (select the checkbox provided for the subnet extension) and then perform an action provided in the **Actions** dropdown list. The **Actions** dropdown list allows you to **Update** or **Delete** the selected subnet extension.

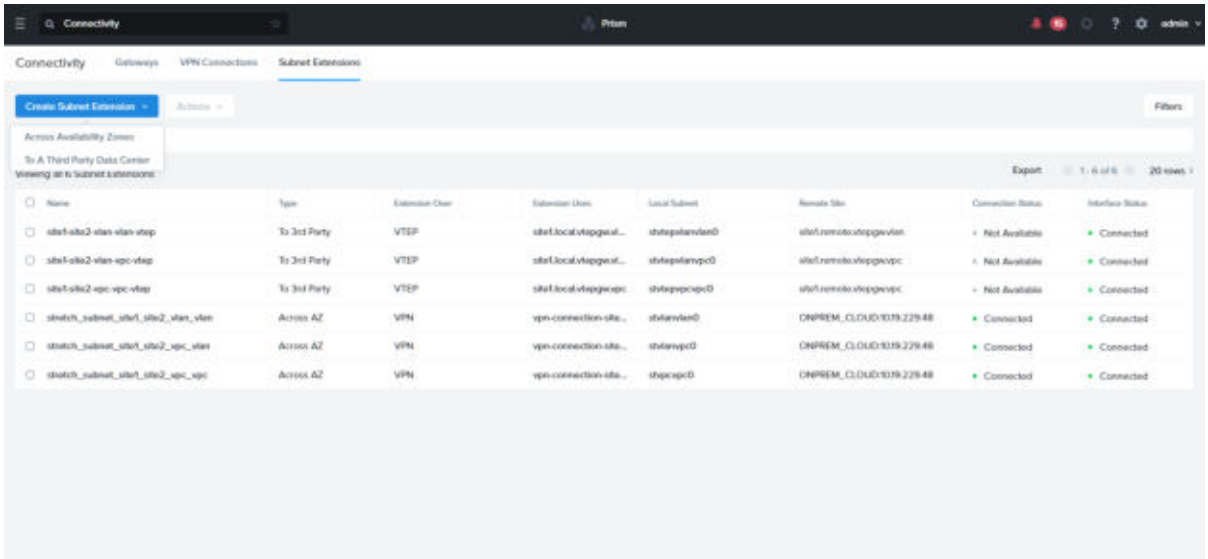


Figure 28: Subnet Extensions dashboard

The **Subnet Extensions** summary list view provides the following details about the gateway.

Table 8: Subnet Extensions List Fields

Parameter	Description	Values
Name	Displays the name of the subnet extension.	(Name of subnet extension)
Type	Displays the subnet extension type.	(**Across Availability Zones** or **To a Third Party Data Center**)
Extension Over	Displays the service that the subnet extension uses.	(VPN or VTEP)
Extension Uses	Displays the name of the local network gateway that the subnet extension uses.	(Name of local network gateway)
Local Subnet	Displays the name of the local subnet that the subnet extension uses.	(Name of local subnet)
Remote Site	Displays the name of the remote network gateway that the subnet extension uses.	(Name of remote network gateway)
Connection Status	Displays the status of the connection that is created by the subnet extension. **Not Available** status indicates that Prism Central is unable to ascertain the status.	(Not Available, Connected, or Disconnected)
Interface Status	Displays the status of the interface that is used by the subnet extension.	(Connected or Down)

You can click the name of a subnet extension to open the subnet extension details page that presents the information about the subnet extension in widgets.

Subnet Extensions Details View

You can click the name of a subnet extension in the **Subnet Extensions** dashboard list to open the subnet extension details page that presents the information about the subnet extension in widgets.

The subnet extension details page displays the name of the subnet extension on the top left corner. It has two tabs - Summary and Address Table. The Summary tab provides the information about the subnet extension in widgets. The Address Table tab provides MAC Address information only when the subnet extension uses VTEP service.

- On the top right corner, the close button (X) allows you to close the details page.
- The **Update** button opens the **Update Subnet Extension** page. See Updating an Extended Subnet on page 127 for more information.
- The **Delete** button allows you to delete the subnet extension. See Removing an Extended Subnet on page 127 for more information.

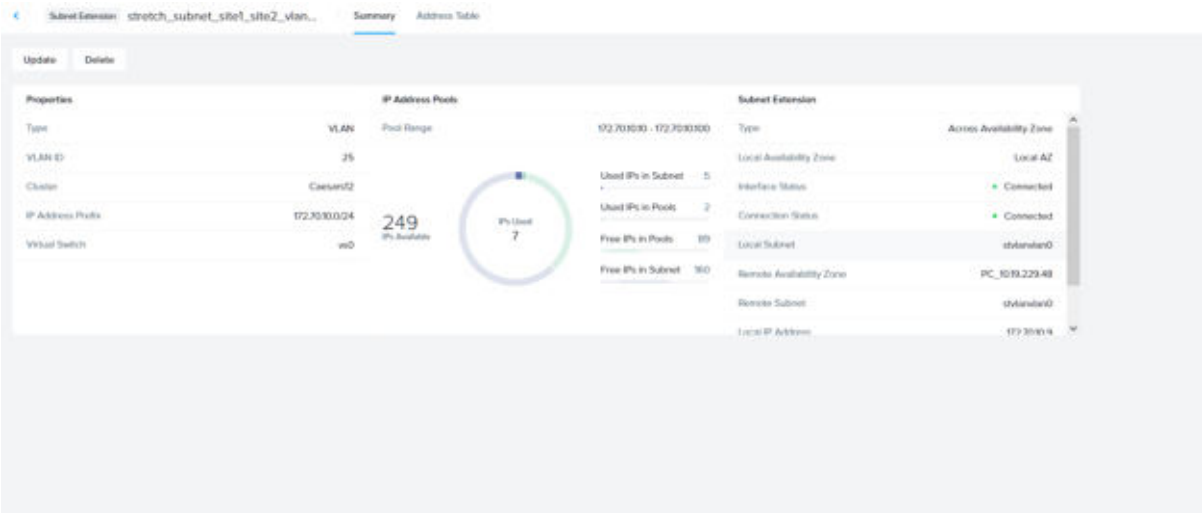


Figure 29: Subnet Extensions Details View - Summary Tab

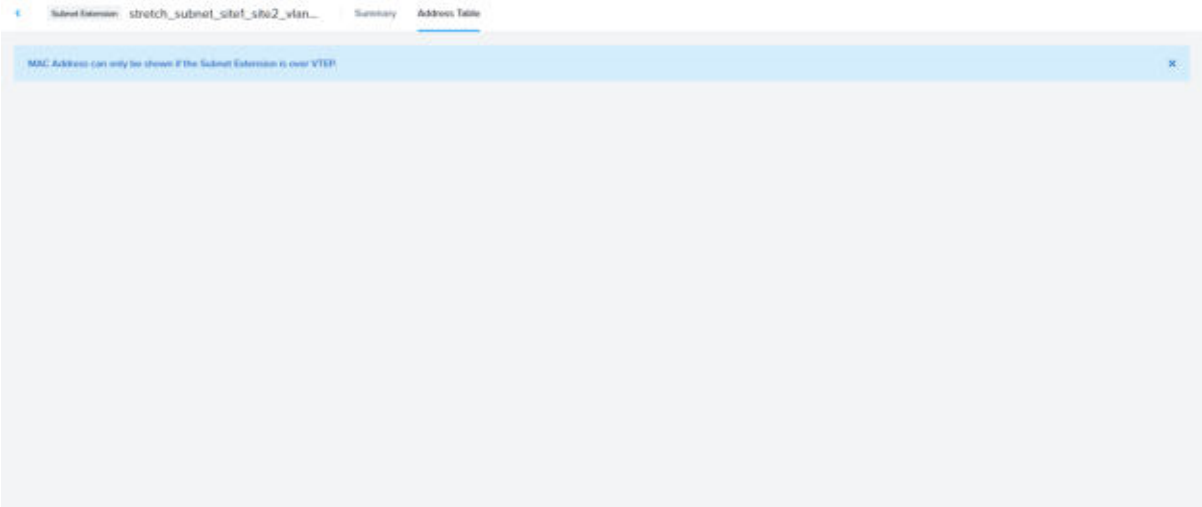


Figure 30: Subnet Extensions Details View - Address Table Tab for VPN-based Extension

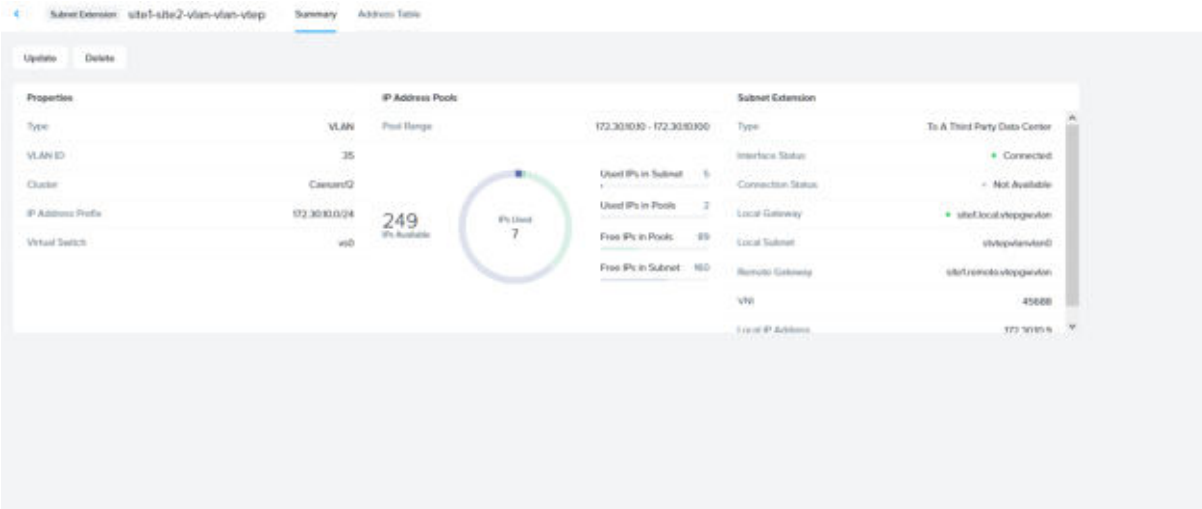


Figure 31: Subnet Extensions Details View - Address Table Tab for VTEP-based Extension

The details about the subnet extension are organized in two tabs. The **Summary** tab organizes the subnet extension details in the extended widget as provided in the table. The **Address Table** tab provides details about the MAC addresses in a list.

Table 9: Subnet Extension Details - Summary Tab Fields

Parameter	Description	Values
Properties		
Type	Displays the subnet type.	(VLAN or Overlay)
VLAN ID	(For VLAN subnets only) Displays the VLAN ID of the VLAN subnet that is extended.	(VLAN ID number)
VPC	(For Overlay subnets only) Displays the name of the VPC subnet that is extended.	(Name of VPC)
Cluster	(For VLAN subnets only) Displays the cluster that the VLAN subnet belongs to.	(Name of cluster)
IP Address Prefix	Displays the network IP address with prefix, of the VLAN subnet that is extended.	(IP Address with prefix)
Virtual Switch	(For VLAN subnets only) Displays the virtual switch on which the VLAN subnet is configured.	(Virtual Switch name such as vs0 or vs1)
IP Address Pools		
Pool Range	Displays the range of IP addresses in the pool configured in the subnet that is extended.	(IP address range)

Parameter	Description	Values
(Interactive Graphic Pie Chart)	Displays a dynamic pie chart that displays the statistic you hover on. Displays the following IP address statistics outside the pie chart, that you can hover on: • Total number of IP addresses available. • Used IP addresses in the subnets • Used IP addresses in the IP address pools • Free IP addresses in the subnets • Free IP addresses in the IP address pools	(IP Address statistics)
Subnet Extension		
Subnet Extension (properties) - Common		
Type	Displays the subnet extension type.	(**Across Availability Zones** or **To a Third Party Data Center**)
Interface Status	Displays the status of the interface that is used by the subnet extension.	(Connected or Down)
Connection Status	Displays the status of the connection that is created by the subnet extension. **Not Available** status indicates that Prism Central is unable to ascertain the status.	(Not Available, Connected, or Disconnected)
Local IP Address	Displays the IP address that you entered in the **Local IP Address** field while creating the subnet extension.	(IP Address)
Local Subnet	Displays the name of the local subnet that the subnet extension uses.	(Name of local subnet)
Subnet Extension (properties) - (Only for **Across Availability Zones** type)		
Local Availability Zone	(Only for **Across Availability Zones** type) Displays the name of the local AZ that is hosting the subnet that is extended.	(Name of the local Availability Zone)
Remote Availability Zone	(Only for **Across Availability Zones** type) Displays the name of the remote AZ that the subnet is extended to.	(Name of the remote Availability Zone)
Remote Subnet	(Only for **Across Availability Zones** type) Displays the name of the remote subnet that the subnet extension connects to.	(Name of remote subnet)
Remote IP Address	(Only for **Across Availability Zones** type) Displays the IP address that you entered in the **Remote IP Address** field while creating the subnet extension.	(IP Address)
Subnet Extension (properties) - (Only for **To a Third Party Data Center** type)		

Parameter	Description	Values
Local Gateway	(Only for **To a Third Party Data Center** type) Displays the name of the local gateway used for the subnet extension.	(Name of local gateway)
Remote Gateway	(Only for **To a Third Party Data Center** type) Displays the name of the remote gateway used for the subnet extension.	(Name of remote gateway)

Security Policies Summary View

To access the security policies dashboard, select **Policies** > **Security Policies** from the entities menu (see *Entities Menu* in the *Prism Central Guide* (see Prism Central Guide)). The security policies dashboard allows you to view summary information about defined security policies.

> Note: This section describes the information and options that appear in the security policies dashboard.
>
> - See *Entity Exploring* in the *Prism Central Guide* (Prism) for instructions on how to view and organize that information in a variety of ways.
> - See *Flow Microsegmentation Guide* for information about how to create and apply security policies.

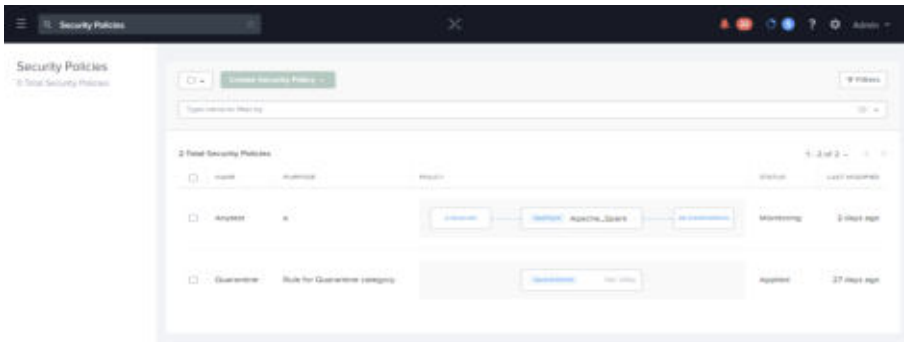


Figure 32: Security Policies Dashboard

The following table describes the fields that appear in the security policies list. A dash (-) is displayed in a field when a value is not available or applicable.

Table 10: Security Policies List Fields

Parameter	Description	Values
Name	Displays the policy name. The policy is one of three types: application, quarantine, or isolation.	(name), Application, Quarantine, Isolation
Purpose	Describes (briefly) the policy's purpose.	(text string)
Policy	Displays (high level) what the policy does.	(boxed text)

Parameter	Description	Values
Status	Displays the current status of the policy (either applied currently or in monitoring mode).	Applied, Monitoring
Last Modified	Displays the date the policy was last modified (or the creation date if the policy has never been modified).	(date)

You can filter the security polices list based on several parameter values. The following table describes the filter options available when you open the Security Policies view **Filter** pane. To apply a filter, select a parameter and check the box of the desired value (or multiple values) you want to use as a filter. You can apply filters across multiple parameters.

Table 11: Filter Pane Fields

Parameter	Description	Values
Name	Filters on the item name. Select a condition from the pull-down list (**Contains**, **Doesn't contain**, **Starts with**, **Ends with**, or **Equal to**) and enter a string in the field. It will return a list of security policies that satisfy the name condition/string.	(policy name string)
Type	Filters on the policy type. Check the box for one or more of the policy types (application, quarantine, isolation). It will limit the list to just those policy types.	Application, Quarantine, Isolation
Status	Filters on the policy status. Check the box for applied or monitoring.	Applied, Monitoring

The security policies dashboard includes a **Create Security Policy** action button with a drop-down list to **Secure an Application** or **Isolation Environments**.

The **Actions** menu appears when one or more policies are selected. It includes options to update, apply, monitor, and delete. The available actions appear in bold; other actions are grayed out. (For grayed out options, a tool tip explaining the reason is provided.)

Security Policy Details View

To access the details page for a security policy, click on the desired security policy name in the list (see Security Policies Summary View on page 53). The Security Policy details page includes the following:

- The policy name appears in the upper left. You can switch from one policy to another by selecting the policy name from the pull-down list.
- The rule status appears below the name and indicates whether the policy is being applied currently or is in monitoring mode.
- Three columns appear that specify the Inbound policy (on the left), the affected entities (in the middle), and the Outbound policy (on the right).

- There are three action buttons (upper right).
  - Click the appropriate button to update, apply, monitor, or delete the policy (see *Nutanix Security Guide* for details). The available actions appear in bold; other actions are grayed out. (For grayed out options, a tool tip explaining the reason is provided.)
  - Click the question mark icon to open a help page in a separate tab or window.
  - Click the X icon to close the details page.

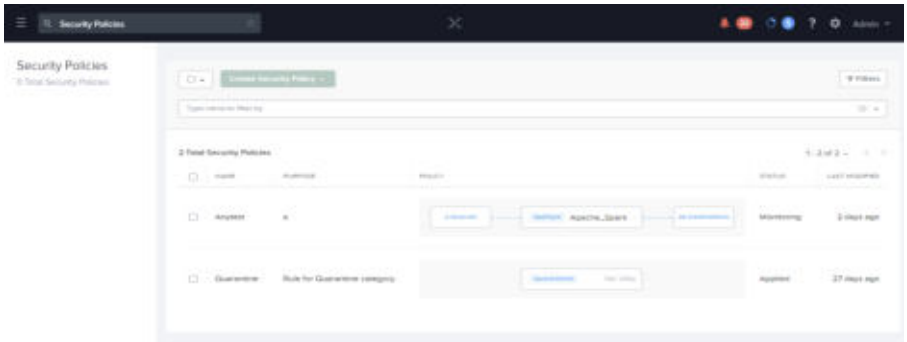


Figure 33: Security Policy Details View: Monitoring Rule Example

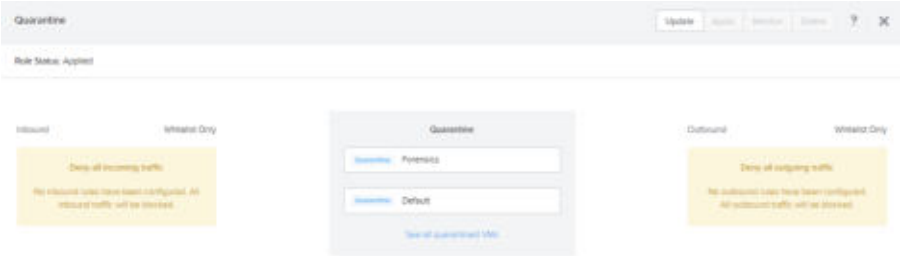


Figure 34: Security Policy Details View: Applied Rule Example

For more information about Security Policies, see the *Flow Microsegmentation Guide*.

8

VIRTUAL PRIVATE CLOUD

A Virtual Private Cloud (VPC) is an independent and isolated IP address space that functions as a logically isolated virtual network. A VPC could be made up of one or more subnets that are connected through a logical or virtual router. The IP addresses within a VPC must be unique. However, IP addresses may overlap across VPCs. As VPCs are provisioned on top of another IP-based infrastructure (connecting AHV nodes), they are often referred to as the overlay networks. Tenants may spin up VMs and connect them to one or more subnets within a VPC.

Virtual Private Cloud (VPC) is a virtualized network of resources that are specifically isolated from the rest of the resource pool. VPC allows you to manage the isolated and secure virtual network with enhanced automation and scaling. The isolation is done using network namespace techniques like IP-based subnets or VLAN based networking.

AHV provides the framework to deploy VPC on on-premises clusters using the following.

- Advanced Networking subnets and DHCP management
- Multiple uplink and bridge management via virtual switch (VS)
- Virtual Private Network (VPN) gateways and connections

Flow Networking simplifies the deployment and configuration of overlay-based VPCs. It allows you to quickly:

- Create, update and delete VPCs.
- Create, update and delete subnets within VPCs.

  Note: Create subnets as necessary when you create VPCs.

- Add network security policies and services.
- Configure hybrid cloud connectivity with VPNs.

This section covers the concepts and procedures necessary to implement VPCs in the network.

VM IP Address Management

Primary Address

During initialization, the cluster provides any virtual NIC (NIC) with only one IP address - the primary IP address. The IP address is a static IP address that you assign or a dynamic IP address that the DHCP server assigns from the DHCP pool.

Secondary IP Addresses (Overlay Networks only)

For your deployment, you may need to configure multiple IP addresses to a single NIC. These IP addresses (other than the primary IP address) are secondary IP addresses. A secondary IP address can be permanently associated with a specific NIC or be changed to any other NIC. The NIC ownership of a secondary IP address is important for security routing policies.

Note: You can configure secondary IP addresses only for VMs in an Overlay network.

You can configure secondary IP addresses to a NIC when you want to:

- When you must associate multiple floating IP addresses with one VM without creating multiple NICs (each with one primary IP address) for the VM. You can assign one floating IP address to one secondary IP address that you create for the single NIC.
- Run appliances, such as load balancers, that have multiple IP addresses on each interface.
- Host applications in a High Availability (HA) configuration where the ownership of IP address moves from the active entity to the standby entity when the active entity goes down.
- Host applications in a clustered configuration where the ownership of IP address follows the leader.
- Host Nutanix Files service in a VPC as a case of clustered application.

Note:

In applications that use secondary IP addresses as virtual IP addresses and the NIC ownership of the secondary IP address changes dynamically from one NIC to another, configure the application to incorporate the ownership change in its settings or configuration. If the applications do not incorporate these ownership changes, the VPCs configured for such applications fail.

For information about configuring secondary IP addresses, see Creating Secondary IP Addresses on page 57.

IP Address Information

You can view the IP addresses configured on a VM by clicking the **See More** link in the `IP Address` column in the VM details view to open the IP Address Information box.

Note: The **See More** link in the `IP Address` column in the VM details view and the IP Address Information box are available only if the VM has any secondary IP addresses configured.

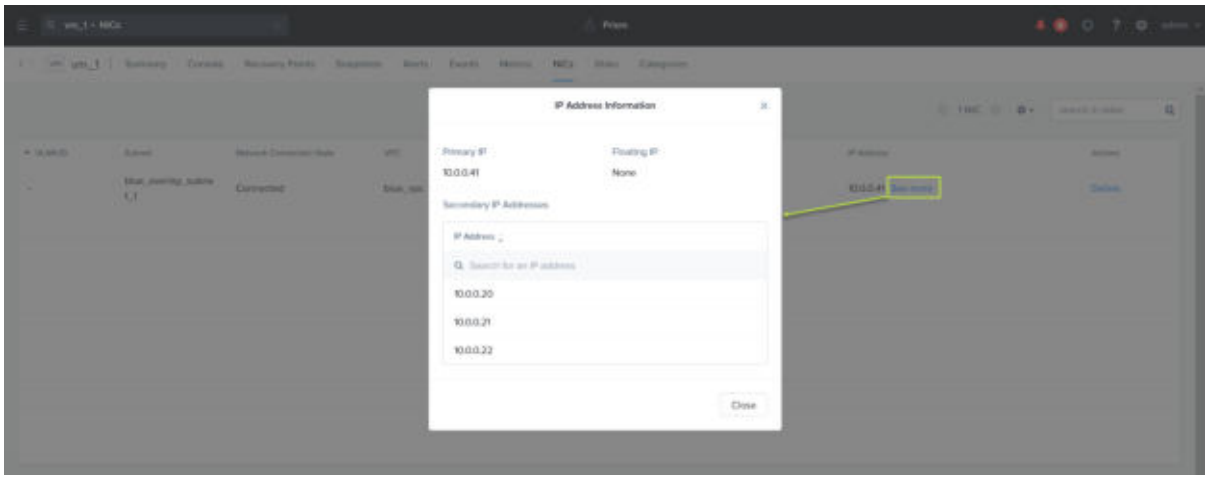


Figure 35: IP Address Information

Creating Secondary IP Addresses

You can assign multiple secondary IP addresses to a single vNIC.

About this task

You can add multiple secondary IP addresses to the vNIC configured on a VM. Add the secondary IP addresses to the vNIC in the Create VM or the Update VM page.

Procedure

1. Go to the **Networks** section.
2. Click the **Edit** icon for the subnet that you want to add the secondary IP addresses from. The Update NIC dialog box opens.

3. Check the **Add Secondary IPs** check box in the Update NIC dialog box.

Update NIC

Subnet

-vpc

Network Connection State

Connected

Private IP Assignment

Network Address / Prefix

10.1.1.0/

Assignment Type

Assign Static IP

IP Address

10.1.1.

Add Secondary IPs

Secondary IP Addresses

e.g. 10.0.0.2, 10.0.0.3, 10.0.0.10

Enter comma separated IP addresses list.

Public IP Assignment

Assignment Type

No Public IP

Floating IP Address

Select Floating IP

4. Add a comma-separated list of the secondary IP addresses that you want to add to the vNIC of the VM.

> Note:
>
> Ensure that the secondary IP addresses are within the same subnet that the primary IP address of the NIC is from. The subnets are displayed in the **Private IP Assignment** section in the Update NIC dialog box.
>
> Ensure that the secondary IP address is not the same as the IP address provided in the **Private IP Assignment** field.

5. Click **Save**.
6. Click **Next** on the Resources and the Management tabs of the Update VM page.

   If you need to make any other changes on the Resources and the Management tabs for any configurations other than adding secondary IP addresses, make the changes and then click **Next** on these tabs.
7. Click **Launch VM** on the Review tab after you review

What to do next

You can view the secondary IP addresses configured on the VM in the IP Address Information box.

Assigning Secondary IP Addresses to Interfaces

Assign the secondary IP addresses to interfaces or subinterfaces on the VM.

About this task

To assign the secondary IP addresses to virtual interfaces on the VM, do the following on the VM details page:

Procedure

1. Click Console.
2. Log in as a root user.
3. Run the `ifconfig` command as follows:

   ```
   root@host$ ifconfig <interface> <secondary ip address> <network mask>
   ```

   Provide the following in the command:

Parameter	Description
<interface>	The interface of the VM such as eth0. You can provide subinterfaces such as eth0:1 and eth0:2.
<secondary IP address>	The secondary IP address that you created and want to associate with the interface.

Parameter	Description
<network mask>	The network mask that is an expansion of the network prefix of the network that the secondary IP address belongs to. For example, if the secondary IP address belongs to 10.0.0.0/24 then the network mask is 255.255.255.0.

4. Repeat the aforementioned steps for all the secondary IP addresses you want to associate with interfaces on the VM.
5. Exit from the Console.

Assigning Secondary IP Addresses to Floating IPs

Assign the secondary IP addresses to floating IP addresses on the VM.

About this task

After you assign secondary IP addresses to interfaces or subinterfaces on the VM, you can assign the secondary IP addresses to floating IP addresses that may be used for external connectivity.

Do one of the following:

Procedure

- Assign floating IP addresses when you request floating IP addresses in the **Assign Floating IPs** section of the Request Floating IP dialog box.

  To assign floating IP addresses while requesting for them, you must have the secondary IP addresses configured and ready when you are requesting the floating IP addresses.
- Select the floating IP address you want to assign, in the Floating IPs dashboard. Click the **Update** option in the **Actions** drop-down menu.

  Assign the secondary IP addresses you configured to the floating IP addresses you have.

VPC Workflow

A virtual private cloud (VPC) can be deployed on Nutanix cluster infrastructure to manage the internal and external networking requirements using Flow Networking. The workflow to create a complete network based on VPC is described below.

1. Create a VPC—See Creating Virtual Private Cloud on page 62. See Updating Virtual Private Cloud on page 84 to update a VPC you created.
2. Add Subnets to the VPC—See Creating a Subnet on page 68 to create a Subnet. See Updating a Subnet on page 85 to update a subnet.
3. Attach the Subnet to VMs—See Attaching a Subnet to a Virtual Machine on page 75.

VPC Management

Creating Virtual Private Cloud

About this task

You can create VPCs on the Virtual Private Clouds page. Go to the Virtual Private Clouds page by clicking **Virtual Infrastructure** > **Networking** > **Virtual Private Clouds**.

To create a VPC, do the following.

Procedure

1. On the VPC dashboard, click **Create VPC**.

   See Network and Security View on page 31 for more information about the VPC dashboard.

   The Create Virtual Private Cloud (VPC) dialog box opens.

Create VPC

Name

Enter VPC Name

External Connectivity

VLAN Subnets with external connectivity are required to be associated to a VPC to send traffic to a destination outside of it.

External Subnets

Select VLAN Subnets with External Connectivity

Maximum of 1 NAT and 1 No-NAT Subnet allowed

Externally Routable IP Addresses

(optional)

Non-overlapping (with other VPCs) address spaces between /16 and /28 netmasks

Domain Name Servers (DNS)

Enter IP Address or FQDN

Cancel Create

Figure 37: Create Virtual Private Cloud

2. Provide the necessary values in respective fields in the Create Virtual Private Cloud (VPC) dialog box.

Fields	Description and Values
Name	Provide a name for the VPC.
External Connectivity	This section takes you through configuration of the parameters necessary for connectivity to the Internet or clusters outside the VPC. A subnet with external connectivity (External Subnet) is required if the VPC needs to send traffic to a destination outside of the VPC. Note: You can add a maximum of two external subnets - one external subnet with NAT and one external subnet without NAT to a VPC. Both external subnets cannot be of the same type. For example, you cannot add two external subnets, both with NAT. You can update an existing VPC similarly. Network address translation (NAT) Gateways perform the required IP-address translations required for external routing. You can also have external connectivity without NAT.
External Subnet	Select an external subnet from the drop down list. By associating the VPC with the external subnet you can provide external connectivity to the VPC. Note: Ensure that the externally routable IP addresses (subnets with external connectivity without NAT) for different VPCs do not overlap. Configure the routes for the external connectivity subnets with next hop as the Router or SNAT IP address. Also configure the routes on the router for the return traffic to reach the VPC. See **External Connectivity** panel in VPC Details View on page 39.
Externally Routable IP Addresses	Provide IP addresses that are externally routable. Externally routable IP addresses are IP addresses that within the VPC which can communicate externally without NAT. These IP addresses are used when an external subnet without NAT is used.
Domain Name Servers (DNS)	(Optional) DNS is advertised to Guest VMs via DHCP. This can be overridden in the subnet configuration. Click **+ Server IP** to add DNS server IPs under **IP Address** and click the check mark. You can **Edit** or **Delete** an IP address you added using the options under **Actions**.

3. Click **Save**.

Requesting Floating IPs

About this task

Each VPN gateway requires a floating IP. If you do not provide one during the VPN gateway creation, then Flow Networking automatically allocates a floating IP to a VPN gateway. To provide floating IP during the VPN gateway creation, you can request floating IPs and assign them to VMs.

You can view the allocated floating IPs on the Floating IPs page. Click **Networking** > > **Floating IPs**.

To request a floating IP, do the following.

Procedure

1. Click the **Request Floating IP** button on the Floating IPs page.

2. On the Request Floating IP dialog box, provide the information in the respective fields.

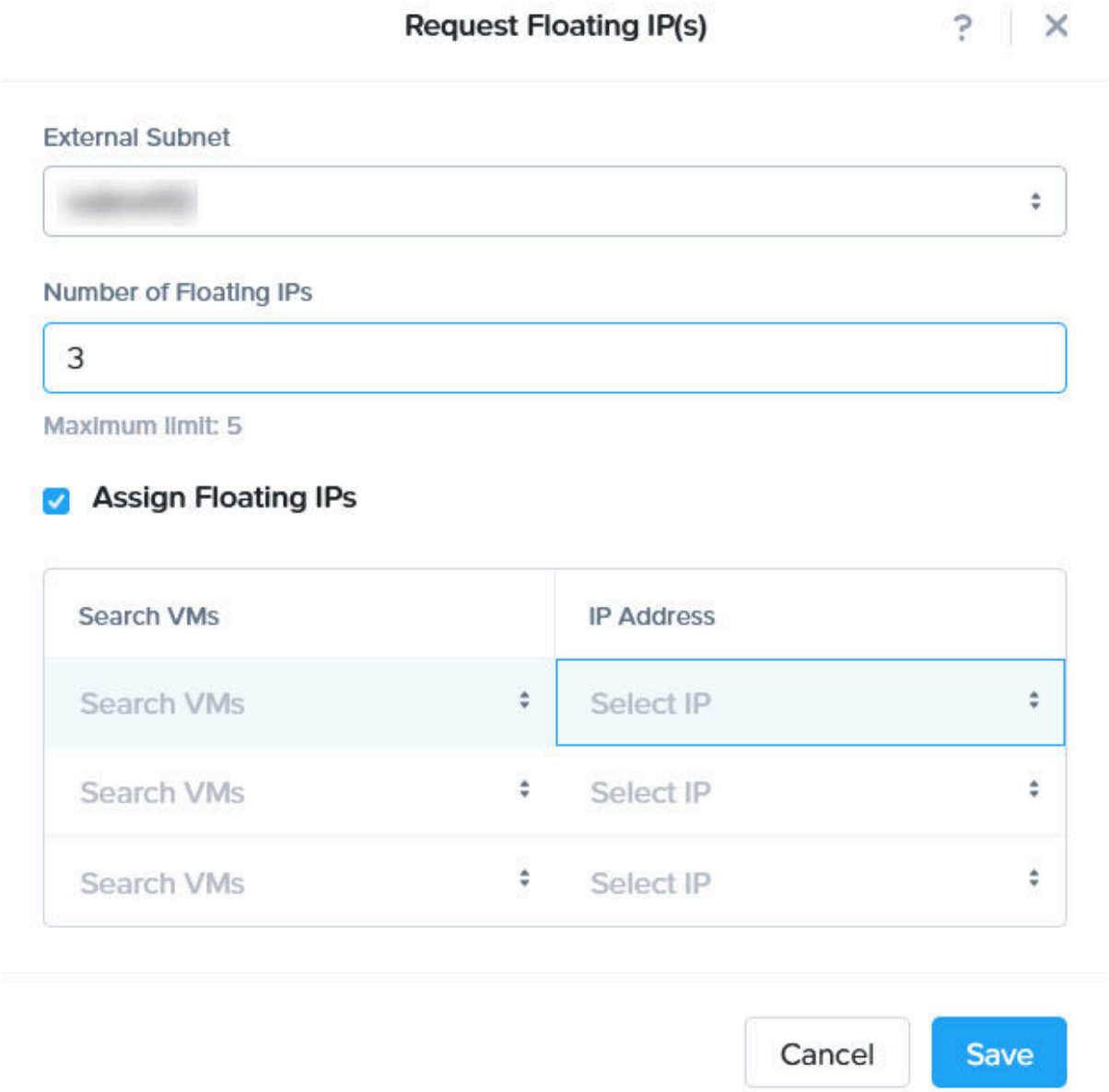


Figure 38: Request and Assign Floating IPs

> Note:
>
> Uncheck the **Assign Floating IPs** box if you want to assign the requested IP addresses after you receive it.
>
> See Floating IPs on page 41 for more information.

Fields	Description and Values
External Subnet	Select a subnet that you configured with external connectivity.

Fields	Description and Values
Number of Floating IPs	Enter the number of floating IPs you want. You can request a maximum of 5 floating IP addresses.
Assign Floating IPs	Select this check box if you want to assign the floating IPs to specific VMs in the table. Based on the number you entered in the **Number of Floating IPs** field, the system provides an equivalent number of rows of `Search VMs` and `IP Address` in the table. Under **Search VMs**, select the VM to which you want to assign a floating IP address. Under **IP Address**, select the IP address on the VM (primary or secondary IP address) to which you want to assign the floating IP. You can assign multiple floating IP addresses to multiple secondary IP addresses that you can create on the NIC of the VM. For information about configuring secondary IP addresses, see Creating Secondary IP Addresses on page 57.

Note:

3. Click **Save**.

What to do next

When you receive the floating IP address you requested, you can see it, assign it (if not already assigned while requesting) or delete it in the Floating IPs view.

Creating a Subnet

About this task

You can create subnets on the Subnets page. Go to the Subnets page by clicking **Virtual Infrastructure** > **Networking** and open the Create Subnet dialog box.

You can also open the Create Subnet dialog box from the VPC details view by clicking the **Add Subnet** option.

To create a subnet, do the following.

Procedure

1. Click **Create Subnet**.
   The Create Subnet dialog box opens. The following figure displays the Create Subnet dialog box with all the options. These options are displayed based on the values you select in the **Type** field.

Create Subnet

Name

Subnet name

Type

VLAN ID

Cluster

Select Cluster

External Connectivity

IP Address Management

Network IP Prefix

0.0.0.0/24

Gateway IP Address

0.0.0.0

DHCP Settings

Domain Name Servers

(Comma Separated): 0.0.0.0

Domain Search

(Comma Separated): nutanix.com,nutanix.eng.com

Domain Name

nutanix.com

TFTP Server Name

Boot File Name

IP Address Pools

Create Subnet

Name

Subnet name

Type

VLAN

VLAN ID

Cluster

Select Cluster

External Connectivity for VPCs

NAT

IP Address Management

Network IP Prefix

0.0.0.0/24

Gateway IP Address

0.0.0.0

IP Address Pools

+ Add IP Pool

Start Address	End Address	Actions

Specify at least one IP address pool. IP Addresses will be used for assigning External IPs to VPCs. These External IPs could additionally be consumed as SNAT and Floating IPs.

Cancel Save

Fields	Description and Values
Name	Provide a name for the subnet.
Type	Select the type of subnet you want to create. You can create a `VLAN` subnet or an `Overlay` subnet.
VLAN ID	(VLAN subnet only) Enter the number of the `VLAN`. Enter just the number in this field, for example 1 or 27. Enter 0 for the native VLAN. The value is displayed as vlan.1 or vlan.27 in the View pages. Note: Provision any single VLAN ID either in the AHV network stack or in the Flow Networking (brAtlas) networking stack. Do not use the same VLAN ID in both the stacks.
IP Address management	(Mandatory for Overlay type subnets) This section provides the ***`Network IP Prefix`*** and ***`Gateway IP`*** fields for the subnet. (Optional for VLAN type subnet) Check this box to display the ***`Network IP Prefix`*** and ***`Gateway IP`*** fields and configure the IP address details. Unchecking this box hides these fields. In this case, it is assumed that this virtual LAN is managed outside the cluster. Note: The **DHCP Settings** option is only available for VLAN subnets if you select this option.
DHCP Settings	(Optional for both VLAN and Overlay subnets) Check this box to display fields for defining a domain. Checking this box displays fields to specify DNS servers and domains. Unchecking this box hides those fields. See Settings the DHCP Options on page 73 for more information.
Cluster (VLAN subnet only)	(VLAN subnet only) This option is available only for `VLAN` subnet configuration. Select the cluster that you want to assign to the subnet.
External Connectivity	(VLAN subnet only) Turn on this toggle switch if you want use this `VLAN` subnet for external connectivity.
NAT	(Option under **External Connectivity**) If you turn on the **External Connectivity** toggle switch, then you can choose whether to connect to external networks with or without enabling NAT. Check the **NAT** check box to enable NAT for external connectivity for VPCs.
Virtual Switch	(VLAN subnet only) Select the virtual switch that is configured for the VLAN you selected. The default value is the default virtual switch vs0. This option is displayed only if you add a VLAN ID in the VLAN ID field.
VPC (Overlay subnet only)	Select the `Virtual Private Cloud (VPC)` that you want to assign to the subnet from the drop down list. You can create VPCs and assign them to `Overlay` subnets.

Fields	Description and Values
IP Address Pool	Defines a range of addresses for automatic assignment to virtual NICs. This field is optional for both `VLAN` and `Overlay`. For `VLAN`, this field is displayed only if you select the **IP Address Management** option. Note: Configure this field for `VLAN` or `Overlay` to complete the creation of the VPC, if you do not need external connectivity for this subnet. You must configure this field only if you need external connectivity for this subnet. Click the **Create Pool** button and enter the following in the Add IP Pool page: • Enter the starting IP address of the range in the **Start Address** field. • Enter the ending IP address of the range in the **End Address** field. • Under **Actions**, click the check mark to submit the starting and ending IP addresses you entered. Click the X mark to remove the entries.
Override DHCP Server	(VLAN subnet only) To configure a DHCP server, check the **Override DHCP Server** box and enter an IP address in the **DHCP Server IP Address** field. See *Override DHCP Server (VLAN Only)* in Settings the DHCP Options on page 73 for information about this option.

2. Click **Save**.

Settings the DHCP Options

About this task

DHCP Settings configures the DHCP options for the VMs within the subnet. When the DHCP Settings are configured for a VM in a subnet and the VM is powered on, Flow Networking configures these options on the VM automatically. If you do not configure the DHCP Settings, then these options are not available on the VM automatically when you power it on.

You can enable DHCP Settings when you create a subnet and configure the DHCP Settings for the new subnet. You could also update the DHCP Settings for an existing subnet.

DHCP Settings is common to and is available on both the Create Subnet and the Update Subnet dialog boxes.

To configure the DHCP Settings, do the following in the Create Subnet or the Update Subnet dialog box:

Procedure

- Provide the information in the DHCP Settings fields.

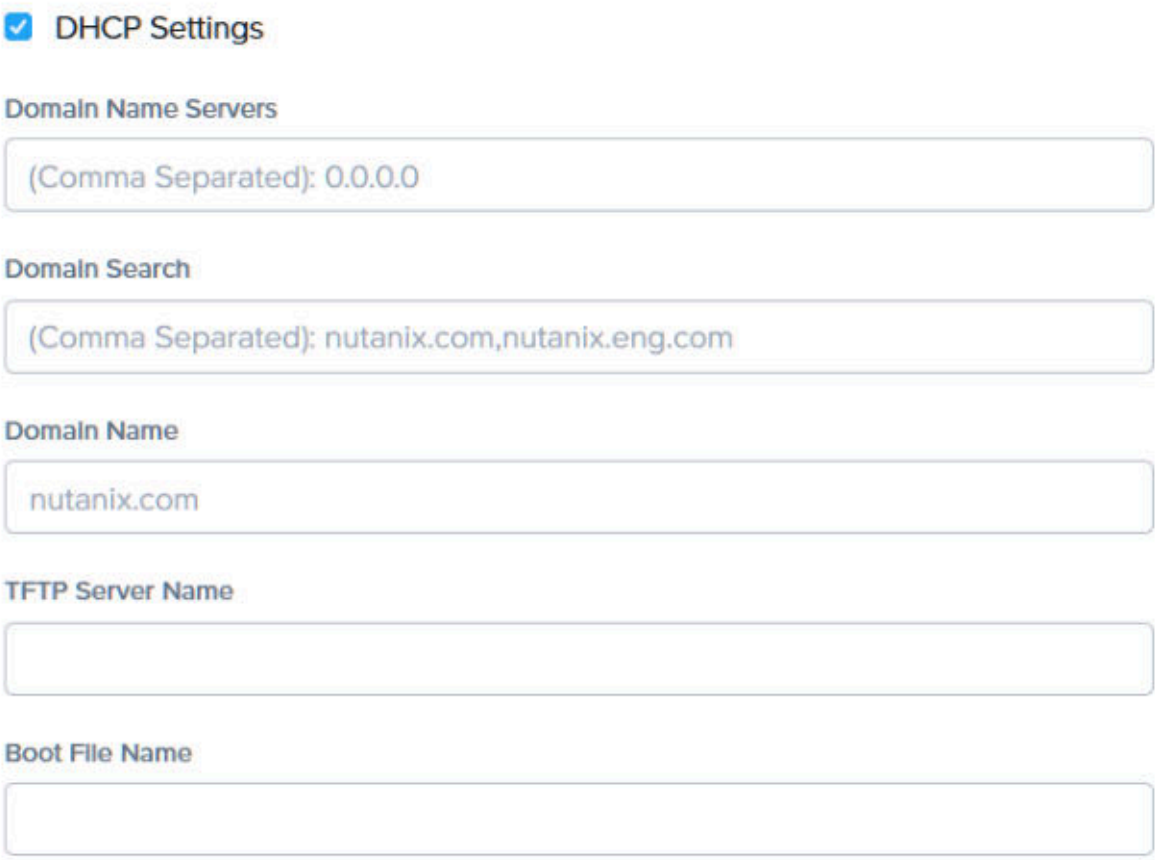


Figure 41: DHCP Settings

Fields	Description and Values
Domain Name Servers	Provide a comma-separated list of DNS IP addresses. Example: 8.8.8.8, or 9.9.9.9
Domain Search	Enter the VLAN domain name. Use only the domain name format. Example: nutanix.com
TFTP Server Name	Enter a valid TFTP host server name of the TFTP server where you host the host boot file. The IP address of the TFTP server must be accessible to the virtual machines to download a boot file. Example: tftp_vlan103
Boot File Name	The name of the boot file that the VMs need to download from the TFTP host server. Example: boot_ahv202010

- (Optional and for VLAN networks only) Check the **Override DHCP Server** dialog box and enter an IP address in the **DHCP Server IP Address** field.

  You can configure a DHCP server using the **Override DHCP Server** option only in case of VLAN networks.

  The DHCP Server IP address (reserved IP address for the Acropolis DHCP server) is visible only to VMs on this network and responds only to DHCP requests. If this box is not checked, the DHCP Server IP Address field is not displayed and the DHCP server IP address is generated automatically. The automatically generated address is `network_IP_address_subnet.254`, or if the default gateway is using that address, `network_IP_address_subnet.253`.

  Usually the default DHCP server IP is configured as the last usable IP in the subnet (For eg., its 10.0.0.254 for 10.0.0.0/24 subnet). If you want to use a different IP address in the subnet as the DHCP server IP, use the override option.

Attaching a Subnet to a Virtual Machine

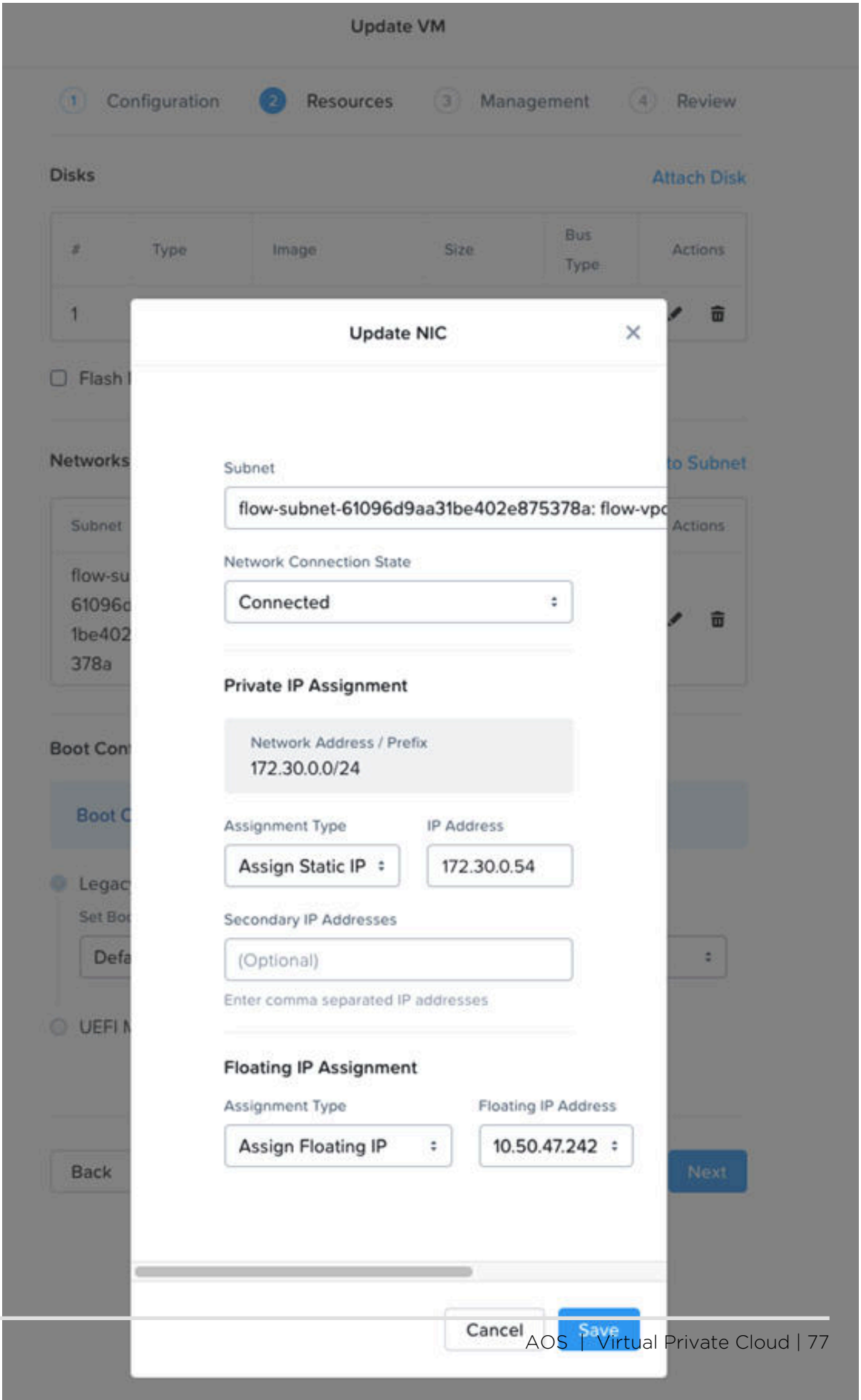
About this task

To attach a subnet to a VM, go to the **Virtual Infrastructure** > **VM** > **List view** in Prism Central and do the following.

Procedure

1. Select the VM you want to attach a subnet to. Click **Actions** > > **Update**.

2. In the Update VM dialog box, click **Add NIC**.


Update VM
1 Configuration
2 Resources
3 Management
4 Review
Disks
Attach Disk
Type Image Size Bus Type Actions
1
Flash
Networks
to Subnet
Subnet Actions
flow-su
61096d
1be402
378a
Boot Con
Boot C
Legac
Set Bo
Defa
UEFI
Back
Next
Update NIC
Subnet
flow-subnet-61096d9aa31be402e875378a: flow-vpc
Network Connection State
Connected
Private IP Assignment
Network Address / Prefix
172.30.0.0/24
Assignment Type
IP Address
Assign Static IP
172.30.0.54
Secondary IP Addresses
(Optional)
Enter comma separated IP addresses
Floating IP Assignment
Assignment Type
Floating IP Address
Assign Floating IP
10.50.47.242
Cancel
Save

3. Provide the necessary information in the indicated fields in the Create NIC dialog box.
   a. Select the **Subnet Name** from the drop down list.
   b. Select the **Network Connection State** as `Connected` or `Disconnected`.

      The **Network Connection State** selection defines the state of the connection after the NIC configuration is implemented.
   c. Select the **Assignment Type**.

      You can select `Assign with DHCP` to assign a DHCP based IP address to the VM.

      You can select `Assign Static IP` to assign a static IP address to the VM to reach the VM quickly from any endpoint in the network such as a laptop.
   d. Click **Add**.
4. Click **Save** on the Update VM dialog box.

Creating a Policy

About this task

For Policy-based routing you need to create policies that route the traffic in the network.

When you create a VPC, there is one default policy that Flow Networking creates for the VPC. This policy is pre-configured with the Priority 1 and other default values to `Deny` traffic flow and service (see the table of field descriptions and values for this dialog box).

> Note: You cannot update or delete the default policy.

- Policies control the traffic flowing between subnets (inter-subnet traffic).
- Policies control the traffic flowing in and out of the VPC.
- Policies do not control the traffic within a subnet (intra-subnet traffic).

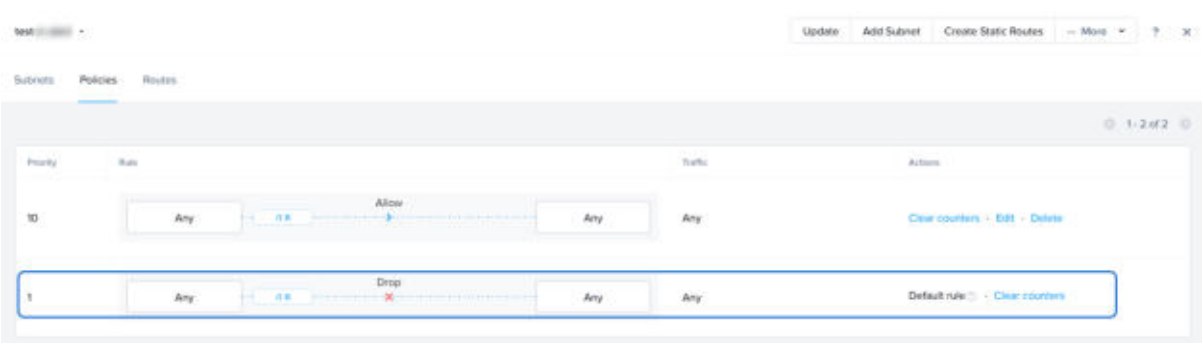


Figure 43: Policy Tab

You can create a traffic policy using the Create Policy dialog box. You can open the Create Policy dialog box either from the VPC list view or the VPC list view.

- On the VPC list view, select the VPC you want to update and click Create Policy in the **Actions** drop down menu.
- On the VPC details view, click the Create Policy option in the **More** drop down menu.

To create a policy, do the following in the Create Policy dialog box.

Procedure

1. Provide the necessary values in the respective fields.

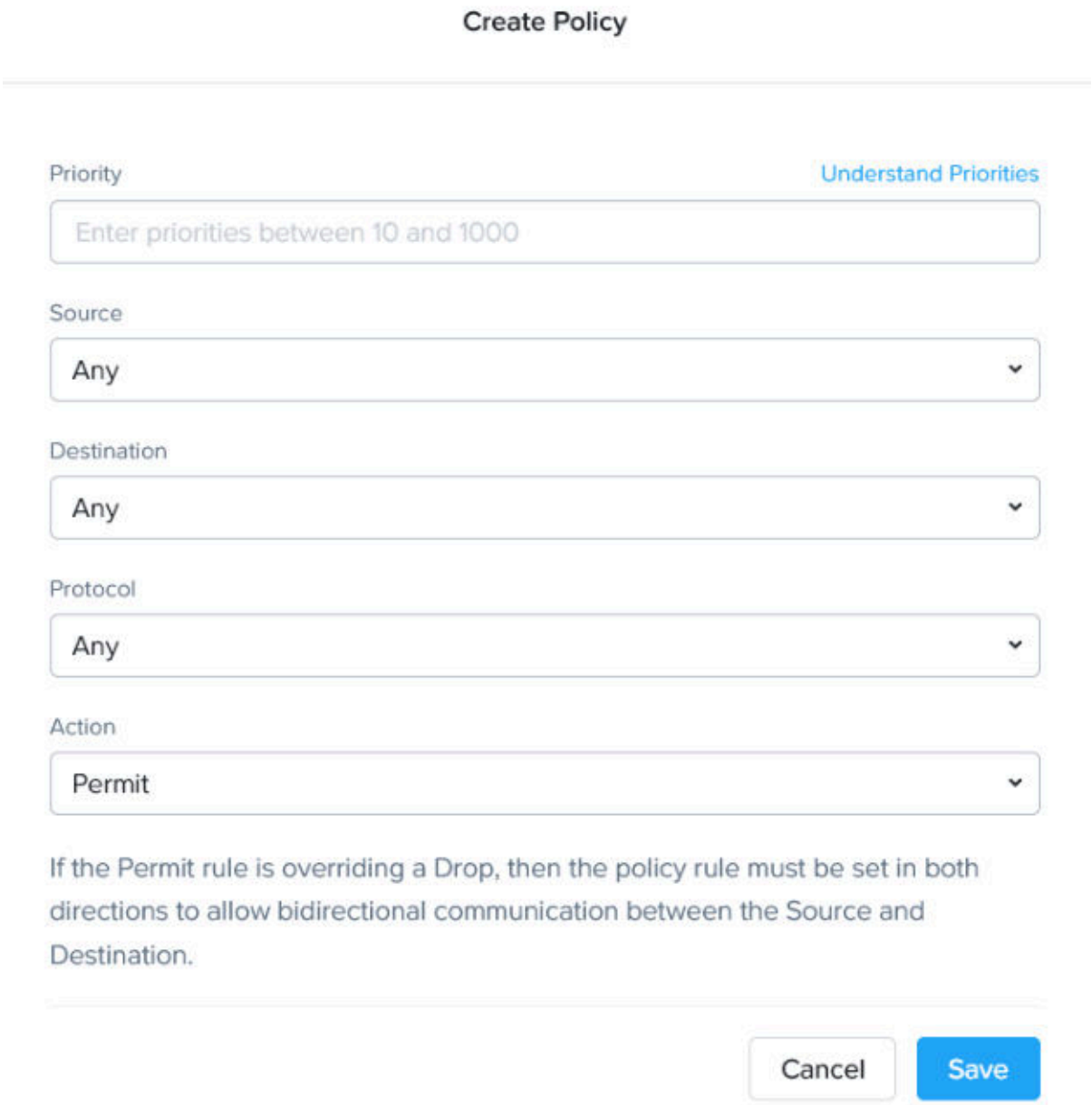


Figure 44: Create Policy

Fields	Description and Values	Value in Default Policy
Priority	The priority of the access list (ACL) determines which ACL is processed first. Priority is indicated by an integer number. A higher priority number indicates a higher priority.For example, if two ACLs have priority numbers 100 and 70 respectively, the ACL with priority 100 takes precedence over the ACl with priority 70. Note: • Click the `Understand Priorities` link to see the `Understand Priorities` information box (see the image of this box below this table).	1
Source	The source indicates the source IP or subnet for which you want to manage traffic. Source can be: • `Any`: Indicates any IP address. • `External`: Indicates an IP address that is outside the subnets configured for the VPC. • `Custom`: You can provide a specific **Source Subnet IP** with prefix.	`Any`
Source Subnet IP	Only required if you selected the **Source** as `Custom`. Provide the subnet IP and prefix that you want to designate as the source for the policy. Use the CIDR notation format to provide the subnet IP. For example, 10.10.10.0/24.	None
Destination	The destination is the destination IP or subnet for which you want to set the priority. Destination can be: • `Any`: Indicates any IP address. • `External`: Indicates an IP address that is outside the subnets configured for the VPC. • `Custom`: You can provide a specific **Destination Subnet IP** with prefix.	`Any`
Destination Subnet IP	Only required if you selected the **Destination** as `Custom`.	None

Fields	Description and Values	Value in Default Policy
Protocol	You can also set the priority configure policy for certain protocols. You can select one of the following options: • `Any`: Indicates any IP address. • `Protocol Number`: Provide an integer number that indicates the protocol you want to prioritize. Provide the appropriate value in the **Protocol Number** field. • `TCP` • `UDP` • `ICMP`	
Protocol Number	This field is displayed only if you select **Protocol Number** as the value in the **Protocol** field. The number you provide must be the IANA designated number that indicates respective protocol. See IANA Protocol Numbers.	None

Fields	Description and Values	Value in Default Policy
Action	Assign the appropriate action for implementation of the policy. • `Permit`: Permits traffic and services based on the parameters set. If the Permit rule is set to override a Drop rule, then the Permit rule must be set in both the directions to allow bidirectional communication between the **Source** and **Destination**. • `Deny`: Denies traffic and service based on the parameters set. • `Re-route`:Sends matching traffic to the next-hop IP address specified by the `Reroute IP`. In case of reroute, you need to provide an IP address that the traffic needs to be re-routed to, in the `Reroute IP` field.	`Permit`

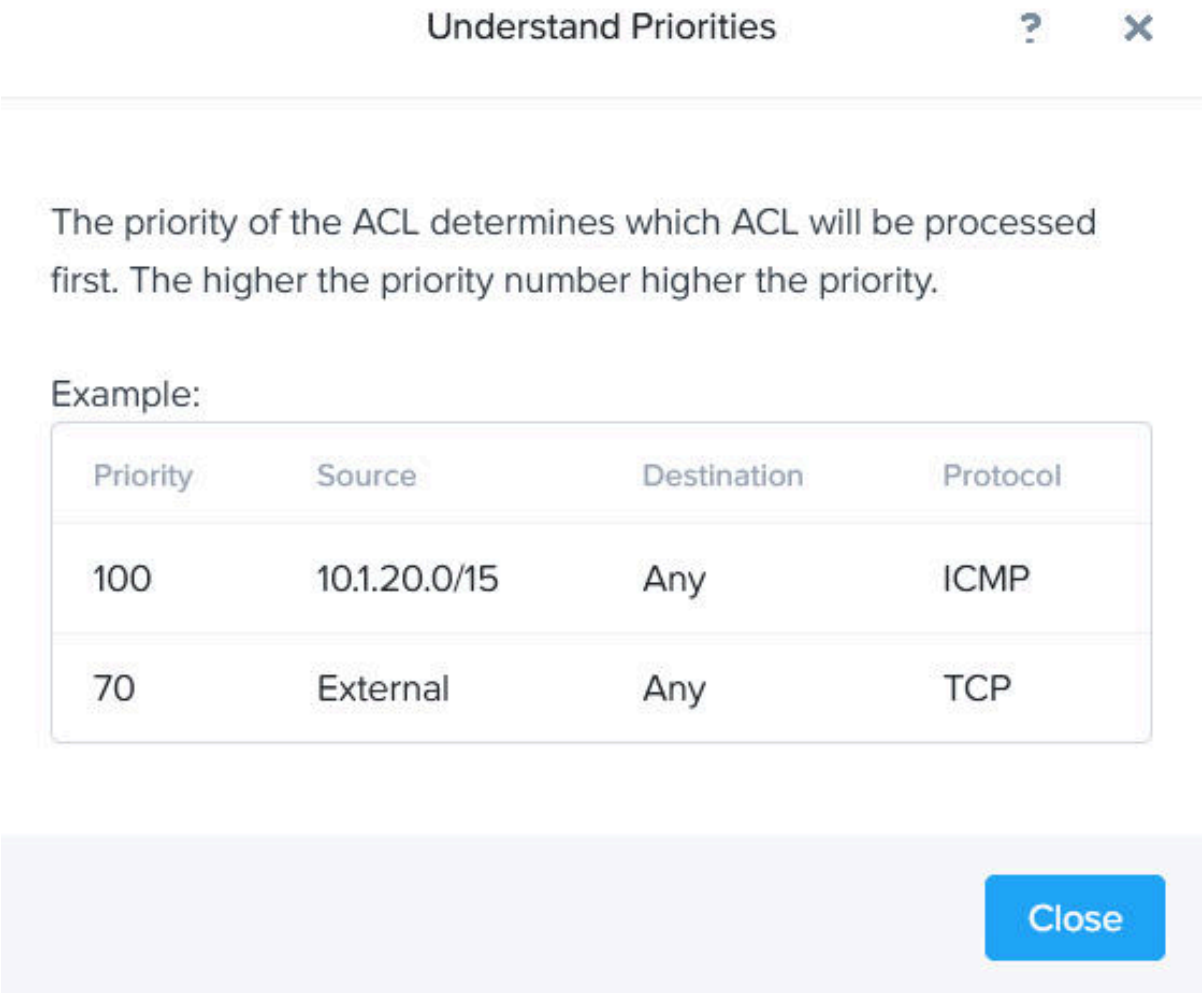


Figure 45: Understanding Priorities

2. Click **Save**.

Creating Static Routes

About this task

You can create a static route using the Create Static Routes dialog box. You can open the Create Static Routes dialog box either from the VPC list view or the VPC details view.

- On the VPC list view, select the VPC and click **Create Static Routes** in the Actions drop down menu.
- On the VPC details view, click the **Create Static Routes** option in the **More** drop down menu.

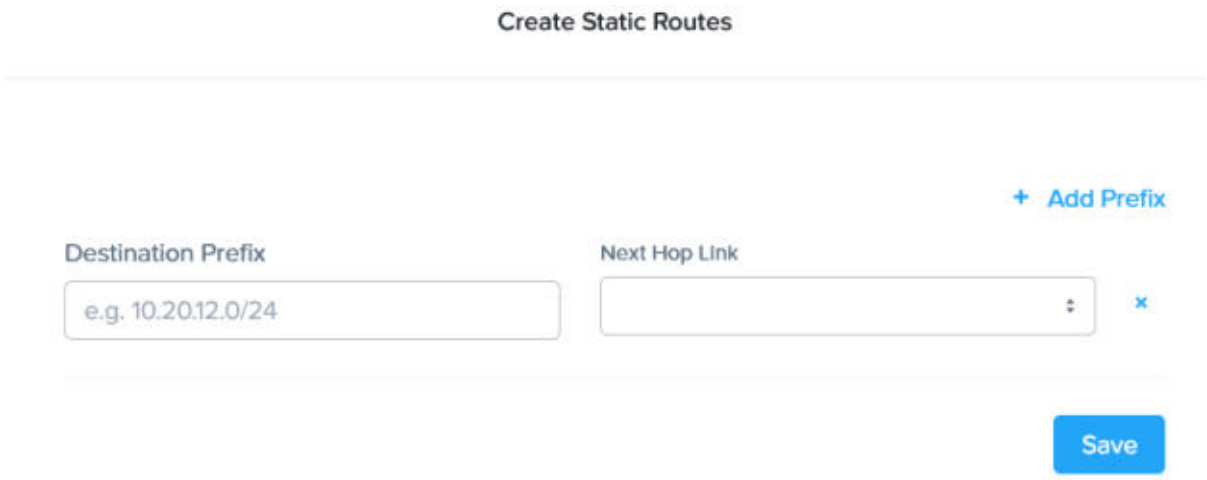


Figure 46: Create Static Routes

To create static route, do the following in the Create Static Routes dialog box:

Procedure

1. Provide the necessary values in the respective fields.

Fields	Description and Values
Destination Prefix	Provide the IP address with prefix of the destination subnet.
Next Hop Link	Select the next hop link from the drop down list. The next hop link is the IP address that the traffic must be sent for the static route you are configuring.
Add Prefix	You can create multiple static routes using this option. Click this link to add another set of **Destination Prefix** and **Next Hop Link** to configure another static route.

2. Click **Save**.

Updating Virtual Private Cloud

About this task

You can update a VPC using the Update Virtual Private Cloud (VPC) dialog box. You can open the **Update Virtual Private Cloud (VPC)** dialog box either from the VPC list view or the VPC details view.

- On the VPC list view, select the VPC you want to update and click **Update** in the **Actions** drop down menu.
- On the VPC details view, click the **Update** option.

The Update Virtual Private Cloud (VPC) dialog box is identical to the Create Virtual Private Cloud (VPC) dialog box.

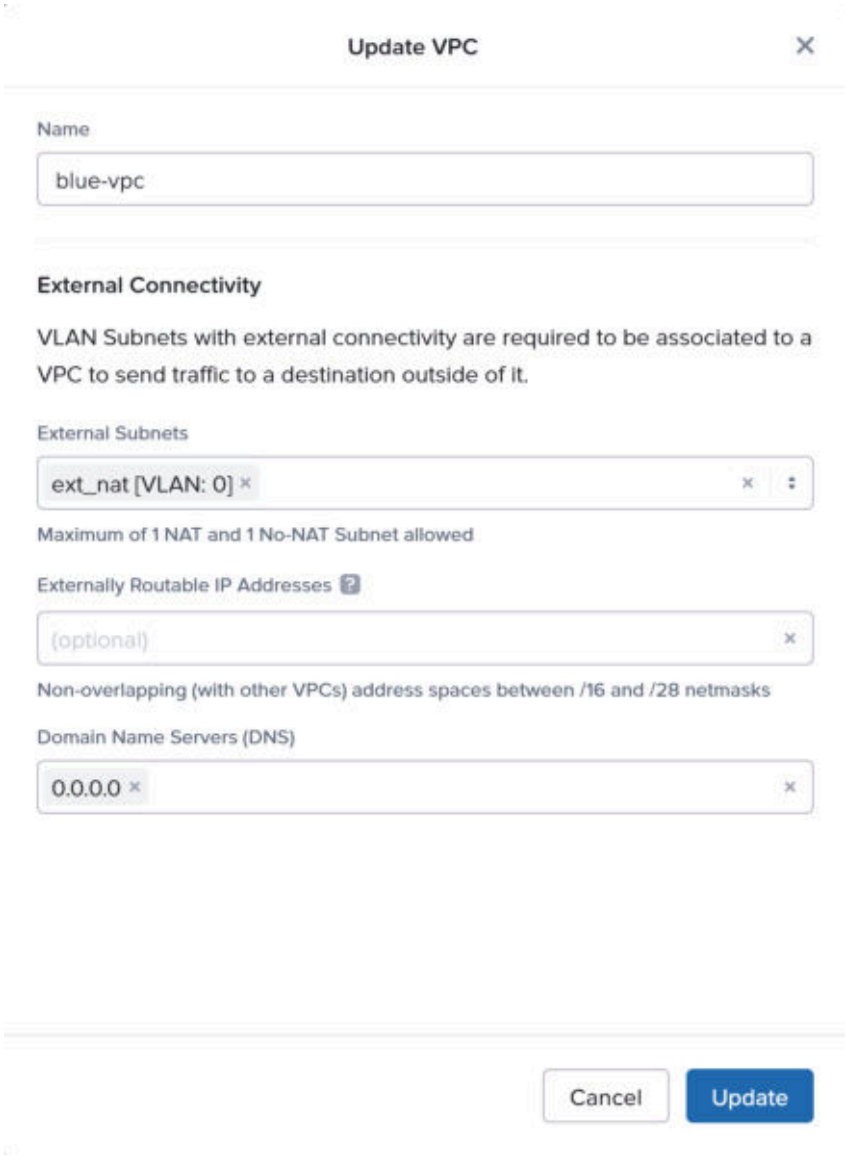


Figure 47: Update VPC

For details about the parameters that you can update in the Update Virtual Private Cloud (VPC) dialog box, see Creating Virtual Private Cloud on page 62.

Procedure

- Update the parameters in the Update Virtual Private Cloud (VPC) dialog box.
- Click **Save**.

Updating a Subnet

About this task

You can update a subnet displayed on the Subnets page. Go to the Subnets page by clicking **Virtual Infrastructure** > **Networking** > **Subnets** and open the Update Subnet dialog box.

You can also open the Update Subnet dialog box from the VPC dashboard for a specific VPC. Click the **Edit** option for the subnet listed on the Subnets tab of the VPC dashboard.

The fields in the Update Subnet and the Create Subnet dialog boxes are the same.

> Note: You cannot edit or update the subnet type. For example, if the subnet type is already configured as ***VLAN***, you cannot modify it to an ***Overlay*** type subnet.

To update a subnets, do the following.

Procedure

1. Select the subnet you want to update. Select **Actions** > **Update Subnet**.

2. Update the necessary values in the respective fields in the Update Subnet dialog box.

Update Subnet

Name

net117

Type

VLAN

VLAN ID

127

Cluster

tagaytay0201

External Connectivity

Virtual Switch

vs0

IP Address Management

Network IP Prefix

10.[redacted]

Gateway IP Address

10.[redacted]

DHCP Settings

Domain Name Servers

10.[redacted],10.4[redacted]

Domain Search

(Comma Separated): nutanix.com,nutanix.eng.com

Domain Name

nutanix.com

TFTP Server Name

Boot File Name

IP Address Pools

+ Create Pool

The Update Subnet dialog box has the same fields as the Create Subnet dialog box. For details about the fields and the values that can be updated in the Update Subnet dialog box, see Creating a Subnet on page 68.

3. Click **Save** to ensure that the updates are saved in the configuration.

Category Management

A category is a key-value pair that groups similar entities. Associating a policy with a category ensures that the policy applies to all the entities in the group regardless of how the group scales with time. For example, you can associate a group of VMs with the Department: Marketing category, where Department is a category that includes a value Marketing along with other values such as Engineering and Sales.

Currently, you can associate only VMs with a category. Categories are implemented in the same way on on-premises Prism Central instances and in Xi Cloud Services. For information about configuring categories, see the *Prism Central Guide*.

Updating a Policy

About this task

You can update a policy using the Update Policy dialog box. You can open the Update Policy dialog box in two ways in the VPC details view.

- On the VPC details view, select the VPC you want to update and click the **Update** option in the top menu.
- On the VPC details view, click the **Edit** option provided in the **Actions** menu for the selected VPC.

> Note: You cannot update or delete the default policy.

The Update Policy dialog box has the same parameters as the Create Policy dialog box.

For details about the parameters that you can update in the Update Policy dialog box, see Creating a Policy on page 78.

Procedure

- Update the parameters in the Update Policy dialog box.
- Click **Save**.

Updating Static Routes

About this task

You can update a static route using the Update Static Routes dialog box. You can open the Update Static Routes dialog box either from the VPC list view or the VPC details view.

> Note: You must configure the default route (0.0.0.0/0) to the external subnet as the next hop for connectivity outside the cluster (north-south connectivity).

- On the VPC details view, select the VPC you want to update and click the **Update** option in the top menu.

- On the VPC details view, click the **Edit** option provided in the **Actions** menu for the selected VPC.

The Update Static Routes dialog box has the same parameters as the Create Static Routes dialog box.

For details about the parameters that you can update in the Update Static Routes dialog box, see Creating Static Routes on page 83.

Procedure

- Update the parameters in the Update Static Routes dialog box.
- Click **Save**.

Deleting a Virtual Private Cloud

About this task

Prism Central does not allow you to delete a VPC if the VPC is associated with any subnets and/or VPNs. After you remove all the subnets or VPN associations from the VPC, delete the VPC.

You can delete a VPC from the VPC list view or the VPC details view.

Procedure

- Do one of the following.
  - » To delete a VPC from the VPC list view, select the VPC you want to delete and click **Delete** in the **Actions** drop down menu.
  - » To delete a VPC from the VPC details view, click the VPC name to go to the VPC details view and click the **Delete** option in the **More** drop down menu.
- In the confirmation dialog box, do the following.
  - » Click **Delete** to delete the VPC.
  - » Click **Cancel** to exit without deleting the VPC.

Deleting Subnets, Policies or Routes

About this task

You can delete VPC entities such as subnets, policies or routes from the VPC details page.

> Note: You cannot update or delete the default policy.

Do the following.

Procedure

1. Open the VPC details page and go to the respective tab like Subnets, Policies or Routes.
2. Click the **Delete** option provided for the selected entity (subnet, policy or route respectively).

3. In the confirmation dialog box, do the following.
   - » Click **Delete** to delete the entity.
   - » Click **Cancel** to exit without deleting the entity.

9

CONNECTIONS MANAGEMENT

This section covers the management of network gateways, VPN connections and Subnet Extensions including operations like create, update and delete network gateways and VPN connections, and extending subnets.

Virtual Network Connections

Virtual Private Network

You can use the Nutanix VPN solution to set up VPN between your on-prem clusters, which exist in distinct routing domains that are not directly connected. These distinct routing domains could either be VPCs within the same cluster or remote clusters or sites.

If you need to connect one Nutanix deployment in one site to another deployment in a different site, you can create a VPN endpoint in each of the sites. A VPN endpoint consists of a local VPN gateway, remote VPN gateway and VPN connection. Local VPN gateway can be instantiated in a VPC context or a legacy VLAN context. Launching the VPN gateway within a VPC allows stretching of the VPC. For example, in the figure, the Blue VPC is stretched between two sites with a VPN.

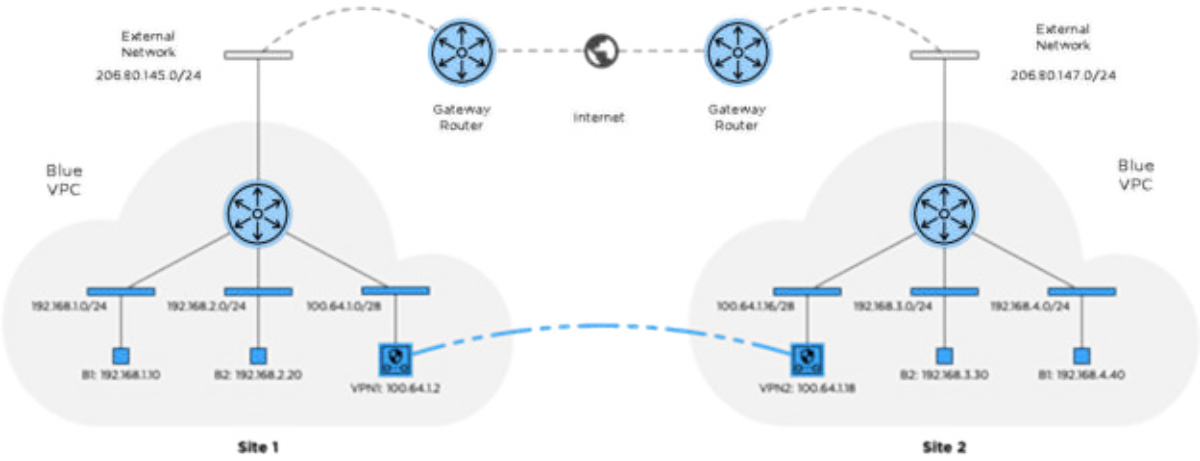


Figure 49: VPN Working

VPN connections are useful in connecting two points. You can connect two VPCs in the same cluster using a VPN or VPCs in different clusters in the same site. However, VPN connection can connect only one endpoint to another endpoint. Flow networking based VPN service allows you to only connect two endpoints that use Nutanix VPN based gateway service.

Virtual Tunnel End Points Based Network Extensions

To connect one endpoint to multiple endpoints or third party (non Nutanix) networks, use Virtual Tunnel End Point (VTEP) service based subnet extensions. For more information about VTEP, see .

VPN Workflow

If you need to connect one Nutanix deployment in one site to another deployment in a different site, you can create a VPN endpoint in each of the sites. A VPN endpoint consists of a local VPN gateway, remote VPN gateway and VPN connection. You can configure multiple VPN endpoints for a site.

Each endpoint must have configurations for a local VPN gateway, remote VPN gateway (pointer information for the peer local VPN in the remote site endpoint) and a VPN connection (connecting the two endpoints). Then, based on the VPN connection configuration as initiator or acceptor, one endpoint initiates a tunnel and the endpoint at the other end accepts the tunnel connection and, thus, establishes the VPN tunnel.

1. Gateways: Every VPN endpoint for each site consists of two VPN gateway configurations - Local and Remote.

   Local gateway is a VM that runs the VPN protocols (IKEv2, IPSec) and routing (BGP and OSPF). Remote gateway is a pointer - database entry - that provides information about the peer remote VPN endpoint. One of the key information contained in the remote gateway is the source IP of the remote VPN endpoint. For security reasons, the local VPN gateway will accept IKEv2 packets originating only from this Source IP.

   VPN gateways are of the following types:

   - On premises Nutanix VPN Gateway: Represents the VPN gateway appliance at your on-premises local or remote site if you are using the Nutanix VPN solution.
   - On premises Third Party Gateway: Represents the VPN gateway appliance at your on-prem site if you are using your own VPN solution (provided by a third-party vendor).

     To configure third party VPN Gateways, see the relevant third party documentation.
2. VPN Connection: Represents the VPN IPSec tunnel established between local gateway and remote gateway. When you create a VPN connection, you need to select two gateways between which you want to create the VPN connection.

VPN appliances perform the following:

1. Implementation of IKEv2 and IPSec protocols.
2. Routing: Between remote sites, Flow Networking advertises prefixes using eBGP. Optionally it uses Static routing. Within a site, Flow Networking uses iBGP or OSPF to share prefixes between the Nutanix VPN appliance and the edge router.

Prerequisites for VPN Configurations

General Requirements

- Ensure that you have enabled Flow Networking with microservices Infrastructure.
- Ensure that you have floating IP addresses when you create VPN gateways.

  Flow Networking automatically allocates a floating IP to a VPN gateway if you do not provide one during the VPN gateway creation. To provide floating IP during the VPN gateway creation, you can request floating IPs. See Requesting Floating IPs on page 66.

- Ensure that you have one of the following, depending on whether you are using iBGP or OSPF:
  - Peer IP (for iBGP): The IP address of the router to exchange routes with the VPN gateway VM.
  - Area ID (for OSPF): The OSPF area ID for the VPN gateway in the IP address format.
- Ensure that you have the following details for the deployment of the VPN gateway VM:
  - Public IP address of the VPN Gateway Device: A public WAN IP address that you want the on-prem gateway to use to communicate with the Xi VPN gateway appliance.
  - Static IP Address: A static IP address that you want to allocate to the VPN gateway VM. Use a floating IP address requested as the static IP address.
  - IP Prefix Length: The subnet mask in CIDR format of the subnet on which you want to install the VPN gateway VM. You can use an overlay subnet used for a VPC and assigned to the VM that you are using for the VPN gateway.
  - Default Gateway IP: The gateway IP address for the on-premise VPN gateway appliance.
  - Gateway ASN: ASN must not be the same as any of your on-prem BGP ASNs. If you already have a BGP environment in your on-prem site, the customer gateway is the ASN for your organization. If you do not have a BGP environment in your on-prem site, you can choose any number. For example, you can choose a number in the 65000 range.

Ports and Protocols

Nutanix deploys a number of ports and protocols in its software. To see the ports and protocols used for Disaster Recovery, see Ports Reference.

Configure rules for ports in your on-premises firewall depending on your deployment scenario. Enable the following in the on-prem VPN Gateway:

- Enable the business application policies to `Allow` the commonly-used business application ports.
- Open UDP ports 500 and 4500 and TCP port 179 in both directions in the firewall policies.

Ports 1024 through 1034 are ephemeral ports used by the CVMs. Ports 1024-1034 are used as load balancer routes in disaster recovery settings with the Prism Central subnets as the source.

Endpoints and Terminations

The following endpoints and terminations occur in the course of Flow networking based connections. For information about creating, updating or deleting VPN connections, see Connections Management on page 92.

> Note: In a VPN connection do not configure both the gateways (local gateway and remote gateway) in an endpoint as Initiators or as Acceptors. If you configure the local gateway as Initiator then configure the remote gateway as Acceptor in one endpoint and vice-versa in the (other) remote endpoint.

VPN Endpoint Behind a Network Address Translation or Firewall Device

In this scenario, the IPSec tunnel terminates behind a network address translation (NAT) or firewall device. For NAT to work, open UDP ports 500 and 4500 in both directions.

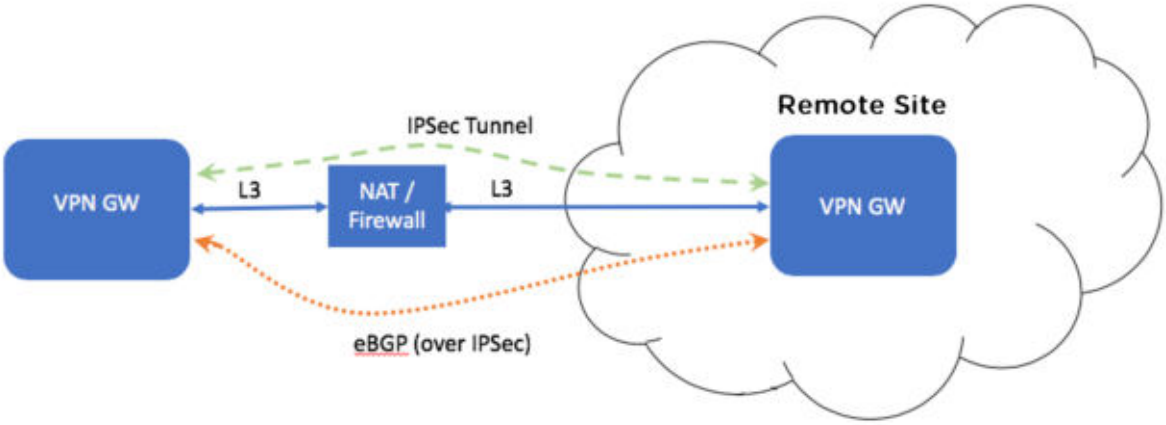


Figure 50: VPN Endpoint Behind NAT or Firewall

Things to do in NAT	Things to do in on-prem VPN GW
Open UDP ports 500 and 4500 on both directions	Enable the business application policies to `Allow` the commonly-used business application ports.

IPSec Terminates on the Firewall Device

In this scenario, you do not need to open the ports for NAT (500 and 4500).

However, enable the on-prem VPN gateway to allow the traffic from the PC subnet to the advertised load balancer route where the Source port is any and the Destination port may be in the range of 1024-1034.

The PC subnet refers to the subnet where your Prism Central is running.

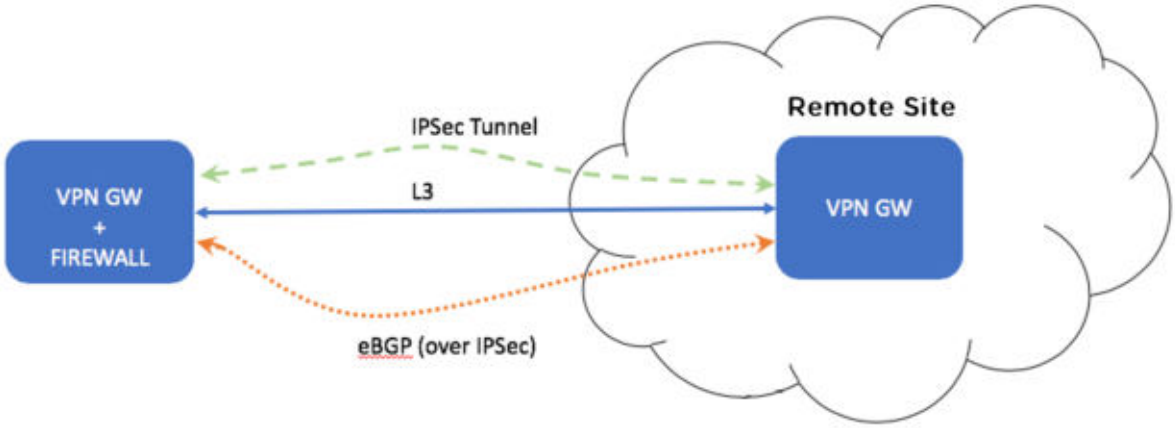


Figure 51: Tunnel Terminates on NAT or Firewall

Creating a Network Gateway

About this task

VPN or s connect two networks together, and can be used in both VLAN and VPC networks on AHV. In other words, you can extend the routing domain of a VLAN network or that of a VPC using a VPN. Accordingly, VPN gateways can be configured using VLANs or VPCs. You need VPN gateways on clusters to provide a gateway to the traffic between on-premise clusters or remote sites.

You can create multiple VPN gateways for a VPC. Since a VPC is configured only on a PC, the VPC is available to all the clusters registered to that PC.

A VPN gateway may be defined as a Local gateway or a Remote gateway based on where the traffic needs to be routed.

To create a VPN gateway, do the following on the **Networking & Security** > **Connectivity** > **Gateways** page.

Procedure

1. Select **Local** or **Remote** in the **Create Gateway** drop-down menu.
   If you select **Local** in the drop-down menu, the Create Local Gateway dialog box opens. If you select **Remote** in the drop-down menu, the Create Remote Gateway dialog box opens.