Technical Report

NetApp HCI 2-Node Storage Cluster

Technical Overview

Stephen Carl, NetApp
April 2020 | TR-4823

Abstract

This document is an overview and description of the NetApp® HCI 2-node storage cluster.

NetApp®

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 NetApp HCI 2-Node Storage Cluster

1.1 Use Cases

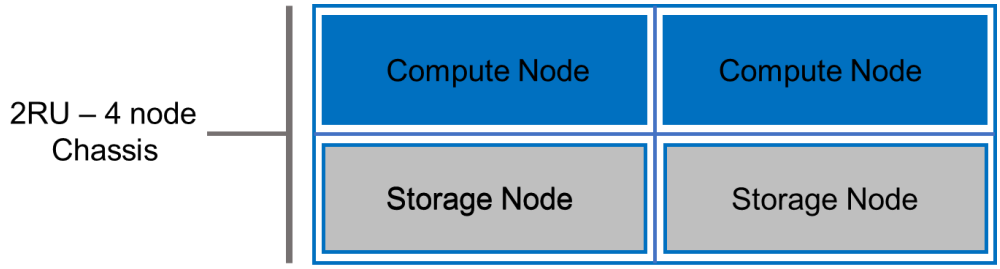
NetApp® HCI provides a foundation for simple scaling, performance, quality of service (QoS) for workloads and multitenancy, and the ability to integrate into the data fabric. For environments that need a smaller footprint enterprise-grade hybrid cloud infrastructure configuration, NetApp HCI can now be deployed at the entry level with only two storage nodes.Customers can design application workloads to take advantage of the performance and simplicity of this 2-node cluster, while retaining the ability to scale when more capacity or bandwith is needed to meet business requirements.

With a 2-node cluster configuration, there are some differences in deployment and expectations as compared to a 3-node, 4-node, or larger NetApp HCI configuration. This report presents technical information about a 2-node cluster and how the differences can influence and aid design considerations.

1.2 Configuration Information

NetApp HCI is available in a range of configuration options for both compute and storage node types. As shown in Figure 1, the configuration for a 2-node cluster must have two storage nodes and two compute nodes. NetApp HCI 2-node cluster is available in HCI version 1.8 running NetApp Element software version 12.0.

Figure 1) NetApp HCI 2-node configuration example in 4U chassis.



The 2-node configuration is available for specific storage nodes for the first release:

- Compute nodes can be H410C 1RU half-width model that fit into the 4U HCI chassis.
- Compute nodes can be any NetApp HCI available 1U model (does not fit into a 4U chassis).
- Storage nodes must be the same model type when deploying 2-node. (Only homogeneous nodes can form a 2-node storage cluster.)

Storage Node Types Supported for 2-Node Clusters

The NetApp HCI 2-node cluster release supports models of the H410 and H610 series of storage nodes. Table 1 describes supported models.

- H410S-0 (small storage)
- H410S-1 (medium storage)
- H610S-1 (small 1U form factor storage)

Table 1) NetApp HCI 2-node storage cluster node models.

Model	Form Factor	Performance (4K IOPS)	Drive Size (GB)	Raw Block Capacity Per Node (TB)	Usable Capacity Per Node (GB)	Effective Capacity (4:1) Per Node (TB)
H410S-11110	1RU, half width	50K	480	1.92	892.80	3.57
H410S-21010	1RU, half width	50K	960	3.84	1785.60	7.14
H610S-1	1RU, full width	100K	960	9.60	4464	17.85

Table 1 uses error threshold to calculate the effective capacity. All 2-node and 3-node cluster configuration storage nodes should be the same model and capacity.

1.3 Node and Cluster Sizing

2-node sizing recommendations are limited to the following models:

- H410S-11110
- H410S-21010
- H610S-1

3-node sizing recommendations are limited to the following models:

- H410S-11110
- H410S-21010
- H610S-1

4-node and larger sizing recommendations allow all NetApp HCI available storage and compute models.

All configurations should use the rated IOPS and cluster capacity to guide the best choice for customer environment workloads.

Compute Node Types Supported for 2-Node

Compute nodes for a 2-node cluster can be any available NetApp HCI model. Only H410C models can be installed in the 4U chassis for half-width nodes. The H610C and H615C are full-width 1U models. For information about available models, see the NetApp HCI Data Sheet.

2 NetApp HCI 2-Node Cluster Architecture and Design Information

2.1 Overview

A NetApp HCI 2-node cluster is created with only two storage nodes; however, the cluster actually contains four nodes: two storage nodes and two witness nodes.

When there are only two storage nodes in the cluster, during a node failure condition there is no method to break a tie when the cluster needs to nominate a node to serve as the new cluster master. To enable cluster redundancy in the event of a node failure, an additional node is required in the cluster. For NetApp HCI, these nodes are called witness nodes.

Witness nodes cannot be deployed on physical storage node drives. The NetApp Deployment Engine (NDE) deploys two witness nodes during NetApp HCI installation, one on each of any two separate compute nodes in a configuration. Each witness node is a virtual machine (VM) that is created by VMware and runs the same Element software as storage nodes, but hosts a minimal set of services.

Node Roles

Element software defines these node roles for a cluster regardless of whether it's a traditional 4-node or larger cluster, or a 2- node cluster with witness nodes. In either scenario, the minimum number of compute nodes remains two for a NetApp HCI deployment. Here are the four types of node roles:

- Management
- Storage
- Compute
- Witness

A management node interacts with a storage cluster to perform management actions, but it is not a member of the storage cluster. Management nodes periodically collect information about the cluster through API calls and report this information. Management nodes are also responsible for coordinating software upgrades of cluster nodes.

A storage node may or may not have any drives and can be a member of a storage cluster. Examples of storage nodes are iSCSI storage nodes that contain slice and block drives.

A compute node provides the compute services in an HCI environment. Compute nodes are not members of a storage cluster.

A witness node does not have any drives, and it can be a member of a storage cluster. Witness nodes run a master service and a cluster ensemble service. These services enable the storage cluster to have enough nodes to form a cluster ensemble.

Cluster Services

Both traditional HCI clusters and the new NetApp 2-node cluster run the same services. These are defined as:

- Master service. The master service manages each node. Each node runs one master service.
- Slice service. The slice service manages the metadata for slices. It also runs the transport service and manages all client I/O. A node runs one slice service if its metadata drive or drives have been added to the cluster.
- Transport service. The transport service manages data transfers with iSCSI channel initiators. This service runs in the same process as the slice service, so it exists only if the metadata drive or drives in the node have been added.
- Block service. The block service manages block data. A node runs one block service for each block drive participating in the cluster.

Cluster Quorum

Element software creates clusters, called ensembles, that store information in a replicated database on storage nodes. Examples include information for the cluster, nodes, volumes, and accounts. There are only two ensemble node roles in an Element cluster:

- Ensemble member
- Ensemble leader

In an optimal Element cluster, physical storage nodes are used for the cluster ensemble members. An ensemble requires a minimum of three nodes or a maximum of five nodes to reach a quorum for cluster resiliency. A cluster cannot be fully node-failure tolerant with fewer than three nodes. To meet the 3-node quorum requirement for an optimal Element cluster in a 2-node cluster, two physical storage nodes and a witness node are required.

Witness Nodes

A witness node is a VM that is integrated into the storage cluster as a member of the ensemble. Element software uses this VM to keep configuration information about the storage cluster in a distrubuted fashion.

Commonality with Physical Storage Nodes

Witness nodes share many aspects of storage nodes, such as:

- Witness nodes can be upgraded or Returned to Factory Install (RTFI) state.
- Witness nodes can be reset using the ResetNode API.
- Witness nodes can be an ensemble member or leader.
- Witness nodes can be added to or removed from a cluster.
- Support bundles can be collected from witness nodes.
- Witness nodes are monitored by the management node.

Limitations and Differences

In a 2-node cluster, the main functions that witness nodes perform are as ensemble members and as Element cluster members. There are some differences from physical nodes that limit the scope of what a witness node is capable of in the cluster:

- Witness nodes are never promoted to Element cluster master.
- Witness nodes never host slice or block services.
- A maximum of 4 witness nodes can exist in an Element cluster at any given time.
- The 2-node cluster is created with a minimum of 2 witness nodes, which are deployed for redundancy in the event of a witness node failure.

For details about Element storage capability and attributes, see the current NetApp Element release notes on the SolidFire All-Flash Storage Documentation Resources page.

2.2 Networking

The 2-node cluster uses the same networks as a 4-node or larger storage node cluster. Additional networks are created in VMware to handle the data path and management for the witness node. These networks are created by the NetApp Deployment Engine during installation of NetApp HCI. For more information about these networks, see section 4, Installation, and section 5, Integration with VMware.

2.3 Security

The 2-node cluster has the same security features and functionality that all NetApp HCI and Element storage products offer. Encryption at rest works the same for a 2-node cluster and has the same failure tolerances, whether or not the cluster has a witness node. For details, see the Manage SolidFire Storage section on the SolidFire All-Flash Storage Documentation Resources page.

2.4 Cluster Behavior Considerations

A 2-node cluster is deployed with different NetApp Element software auto-healing architecture capabilities than a larger cluster. The complete auto-healing capability of a 3-node or larger cluster automatically handles events that result from an unplanned hardware failure, including rebalancing any data that was on the component that failed to other nodes in the cluster. The 2-node cluster cannot rebalance from a physical node failure in this manner because only one physical node remains. A minimum of three physical storage nodes must be available to the cluster for auto-healing. This fundamental difference in 2-node cluster auto-healing resiliency must be considered in planning and design for workload performance of applications.

The 2-node cluster has less capacity compared to a 4-node or larger storage cluster. The performance IOPS and capacity of a 2-node cluster are optimally designed for workloads that do not depend on larger capacity requirements and maximum IOPS ratings of the storage nodes. Application workloads that are read heavy and that do not exceed the maximum IOPS rating of a single node can tolerate node failures with limited performance impact.

2.5 How a 2-Node Cluster Is Different from a Traditional HCI Cluster

Node Ensemble

In general, witness nodes are used to ensure that enough cluster storage nodes exist to form a valid ensemble quorum. Ensemble creation and reconfiguration prefer physical storage nodes over witness nodes; only a physical storage node can be an Element cluster master.

Ensemble Management with Witness Nodes

The following examples show expected mangement behaviors when witness nodes are present. The ensemble is configured or reconfigured to achieve this behavior whenever the cluster is created or a node is added or removed. Some of the examples reflect more physical nodes after expanding from a 2-node cluster.

- Two storage nodes and one witness; 3-node ensemble using physical storage nodes and witness node
- Three storage nodes and two witnesses; 3-node ensemble using only physical storage nodes
- Four storage nodes and two witnesses; 3-node ensemble using only physical storage nodes
- Five storage nodes and two witnesses; 5-node ensemble using only physical storage nodes
- Five+ storage nodes and two+ witnesses; 5-node ensemble using only physical storage nodes

2-Node Cluster Ensemble Management with Physical Storage Node Failure

One storage node online, one storage node offline, and two witness nodes. The cluster is in a degraded state. Only one witness node is active in the ensemble. This condition won’t add the second witness node to the ensemble as its role is a witness node backup. The degraded cluster is fixed when the offline storage node returns to an online state, or a replacement node joins the cluster.

Figure 2 and Figure 3 show this difference of physical nodes and the witness node for a 4-node storage cluster and a 2-node storage cluster. Each cluster contains the minimum three nodes necessary to create an ensemble quorum. ZooKeeper is used to store and manage the ensemble and database.

Figure 2) Traditional 4-node storage cluster ensemble.

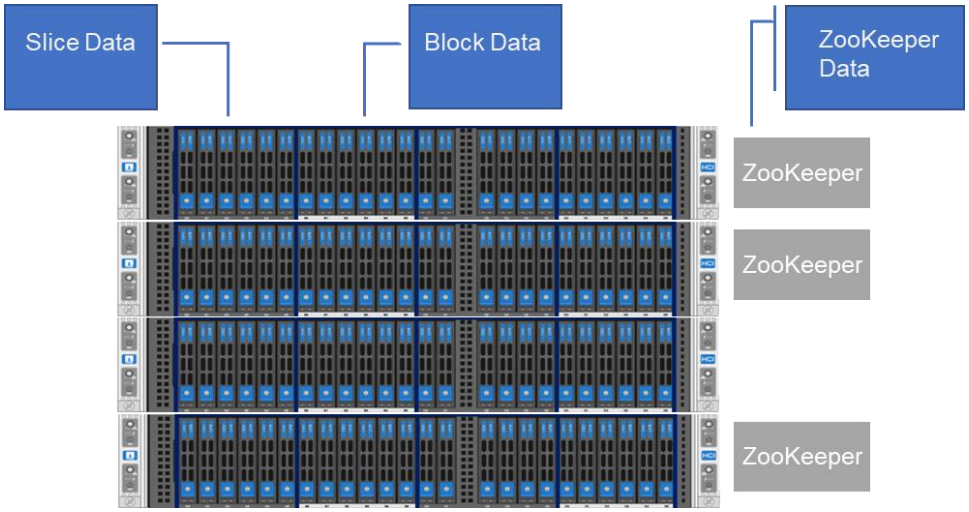
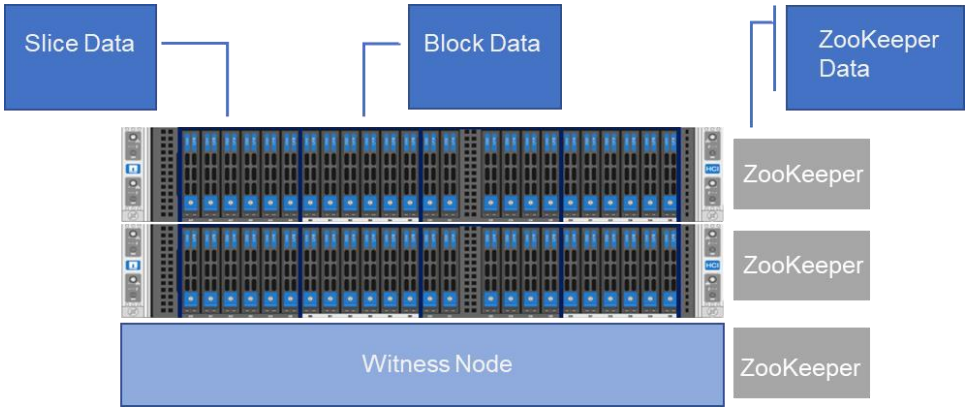


Figure 3) 2-node storage cluster ensemble.



shows that the witness node has ensemble membership capabilites in the cluster, because it stores ZooKeeper database data. This satisfies the ensemble requirement of a minimum of 3 members and acts as a tie breaker to promote a new cluster master vote if a physical node fails. Only a physical node can be a cluster master.

2.6 API Changes for 2-Node

Element software API changes were made to accommodate the 2-node cluster release. The following list summarizes the changes. See the NetApp SolidFire API Guide for details.

- CreateCluster
- CheckProposedCluster
- CheckProposedNodeAdditions
- GetClusterFullThreshold
- ModifyClusterFullThreshold
- GetLimits

- ListAllNodes
- ListActiveNodes
- ListPendingNodes
- ListPendingActiveNodes
- ListAvailableNodes

2.7 Node and Drive Failure Behaviors

When a storage node fails, the new 2-node cluster behaves differently than a larger cluster. A 2-node cluster does not require regeneration of a second copy of data when a single node fails. New writes are replicated for block data in the remaining active storage node. When the failed node is replaced and rejoins the cluster, the data is rebalanced between the two physical storage nodes.

Node Failure

In the event of a node failure, a 2-node cluster defaults to running a 2-node ensemble of a physical node and a witness node, which triggers an ensemble degraded fault. The failed node contains 50% of the cluster data; this condition triggers the ensembleDegraded, blocksDegraded, and volumesDegraded faults. If space allows, all new writes are replicated in the remaining storage node on two separate block drives. Once the failed node is repaired or replaced and the cluster is returned to optimal, the data is rebalanced between the two storage nodes and the triggered faults are resolved.

Witness Node Failure

If a witness node fails, the default is for the remaining running active witness node to join the ensemble to reach the cluster quorum of three. A new witness node can be deployed in VMware to replace the failed witness node. During NDE installation of a 2-node cluster, a witness node template is created for the purpose of creating a new witness node. The new witness node is automatically created from the correct datastore for witness nodes on the compute node.

To join the storage cluster, the new witness node must be configured. This configuration can be accomplished by accessing the VM remote console of the new witness node and editing the teminal user interface (TUI). The cluster and network options in the menu have fields to edit, such as node name, IP address for net0 and net1 networks, gateway, and cluster name to join. The new witness node VM should have configuration settings similar to existing witness nodes from the original NDE installation.

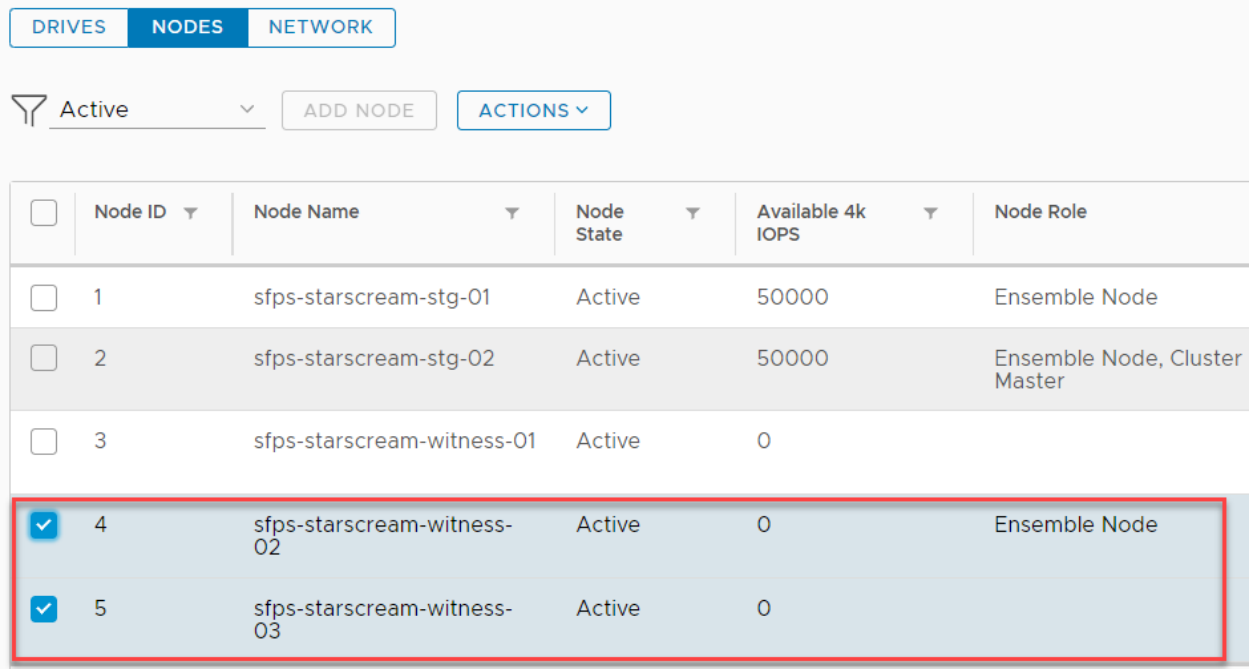
Figure 4 shows the TUI screen capture for the new witness nodes cluster settings that need to be edited. After editing, the net0 and net1 interfaces join the required virtual networks already confingured for the existing NetApp HCI system. The user has to set the Hostname and Cluster fields to add the node to the cluster.

Figure 4) New witness node cluster information.

After all fields are edited and saved, the new node needs to be accessed in the Element UI or vSphere Element Configuration Management UI, where the node is presented as a pending node. Select the new node and add it to the cluster. The VM becomes available as an active witness node.

Figure 5 shows the new witness node has an ID of 5. In this example, the failed witness node is node ID 3. It is still displayed as an active member of the cluster until it is removed manually by using the UI or APIs. After it is removed it can be deleted in VMware.

Figure 5) New witness node in cluster.



Node ID	Node Name	Node State	Available 4k IOPS	Node Role
1	sfps-starscream-stg-01	Active	50000	Ensemble Node
2	sfps-starscream-stg-02	Active	50000	Ensemble Node, Cluster Master
3	sfps-starscream-witness-01	Active	0	
4	sfps-starscream-witness-02	Active	0	Ensemble Node
5	sfps-starscream-witness-03	Active	0	

Drive Failure

In the event of a drive failure, the normal Element software failed drive fault is triggered. If the drive type is a metadata drive, the volumesDegraded fault is also triggered.

- Regardless of capacity, only a single metadata drive failure can ever be tolerated.
- For block drives, if there is enough capacity, the cluster can continue to have cascading block drive failures (leaving time between failures to auto-heal).

Single Storage Node Failure

- No loss of access to data.
- If capacity allows, new writes are replicated for two available copies on the remaining node. Otherwise, new writes are singly written.
- The cluster does not auto-heal. The user must restore or replace the failed node.

Single Storage Node Failure Plus Block Drive Failure

Same as current Element software behavior, data unavailability.

Single Storage Node Failure Plus Slice Drive Failure

Both metadata copies are lost, restore from backup.

2.8 Node Failure Performance impacts

The impact of a node failure scales up as the cluster size scales down. If highly available performance is a requirement for any protection domain size (n), that cluster should not exceed the performance rating of n-1 domains. (A domain is a node.)

For clusters that exceed the performance rating of n-1 domains, the impact of a failure increases as the number protection of domains decreases.

Because a base 2-node cluster has the smallest number of domains, the impact of a node failure when running at performance capacity is more extreme than with clusters of four or more storage nodes. When running workloads on a 2-node cluster at performance capacity, a single node failure limits 50% of the performance resources of the cluster. It is a best practice to size the workloads deployed appropriately to meet performance expectations during the short time it takes to repair a failed node in a 2-node cluster.

When designing a 2-node cluster, it is a best practice to use IOPS ratings of storage nodes to avoid workload performance that exceeds the maximum of one node’s IOP ratings. A node failure using one node’s rating results in minimal impact to performance.

2.9 Node Failover Recovery Impacts

A 2-node cluster exhibits quick failover and recovery time of the failed node. This is a result of the cluster doing less work because the failed data is not synced to other nodes during the failure/recovery process. Although the process is substantially faster, the cluster is less resilient to further failures compared to a larger cluster that has synced out data and auto-healed to full replication. The 2-node node cluster does not have the same auto-healing features as larger clusters. This imposes a tradeoff: node failures are resolved quickly, but the cluster is placed at an elevated risk to data integrity in the case of future failures during longer outages.

2.10 Drive Failure Performance Impacts

The impact of a slice drive failure is significant when running at the performance capacity of the cluster. A block drive failure on a 2-node cluster has similar impact as on a 4-node cluster.

The impact of drive failures when running an n-1 IOPS rate (50K) is considered minimal.

2-Node Cluster Fullness Thresholds

All Element software clusters have three levels of cluster fullness: warning, error, and critical. All three are displayed in the NetApp Element UI or the NetApp Element Management UI in vSphere.

Element software APIs support modifying block and metadata warning threshold separately. Quality of service (QoS) proportionally throttles write IOPS between max and min IOPS as metadata fullness moves from error to critical. Read IOPS remain unaffected.

For a 2-node storage cluster there are some differences in these threshold levels and recommendations for resolution.

2-Node Block Cluster Thresholds

The system uses the Block Cluster Full error code to warn about cluster block storage fullness. The cluster fullness severity levels can be viewed from the Alerts tab of the Element UI or the NetApp Element Management UI in vSphere.

In general, a storage cluster fault is triggered when any singular slice (block) drive reaches 100% full capacity. The following list describes the Block Cluster Full severity levels:

- **Warning:** Set at 3% under the error level. Free up capacity or add more capacity by scaling to a 4-node cluster minimum. Expansion to a 3-node cluster is supported but does not add capacity.
- **Error (2-node calculation changes):**
  - Block Cluster Full error threshold is (N-1)/N - 3% where N is the number of drives in the node with the fewest drives.

    For example: For storage node type H410S with six drives, one drive in each storage node is allocated as a slice drive (metadata), leaving the block drive count at five.
  - Calculation: (5-1)/5 - 3%, or 77%, is the block cluster full threshold for this 2-node cluster.
  - Free up capacity or add more capacity by scaling to a 4-node cluster minimum. Expansion to a 3-node cluster is supported but does not add capacity.
- **Critical:** The cluster is 100% consumed, in a read-only state. Free up capacity or add more capacity by scaling to a 4-node cluster minimum. Expansion to a 3-node cluster is supported but does not add capacity.

2-Node Metadata Cluster Thresholds

The following list includes information about the metadata cluster full severity levels:

- **Warning:** The warning threshold is configurable; the default is 3% below warning threshold. If the cluster is in this state and a node fails, the cluster is using only a single node in a degraded state.
- **Error:** The threshold is 20% of total cluster capacity. Free up capacity or add more capacity by scaling to a 4-node cluster minimum. Expansion to a 3-node cluster is supported but does no add capacity.
- **Critical:** The cluster is 100% consumed, in a read-only state. You must free up capacity or add more capacity, added by scaling to a 4-node cluster minimum. Expansion to a 3-node cluster is supported but does not add capacity.

3-Node Cluster

A 3-node cluster is a valid configuration. However, there are design considerations and limitations. The 3-node storage cluster only adds resiliency to enable auto-heal for a 2-node storage cluster. Added capacity to the storage cluster from the addition of a third node is minimal. The amount of available cluster IOPS increases slightly with the third node. These limitations should be considered during

configuration design. The 3-node storage cluster also deploys a witness node to make sure that the 3-node cluster quorum of three is maintained if a storage node fails or is removed.

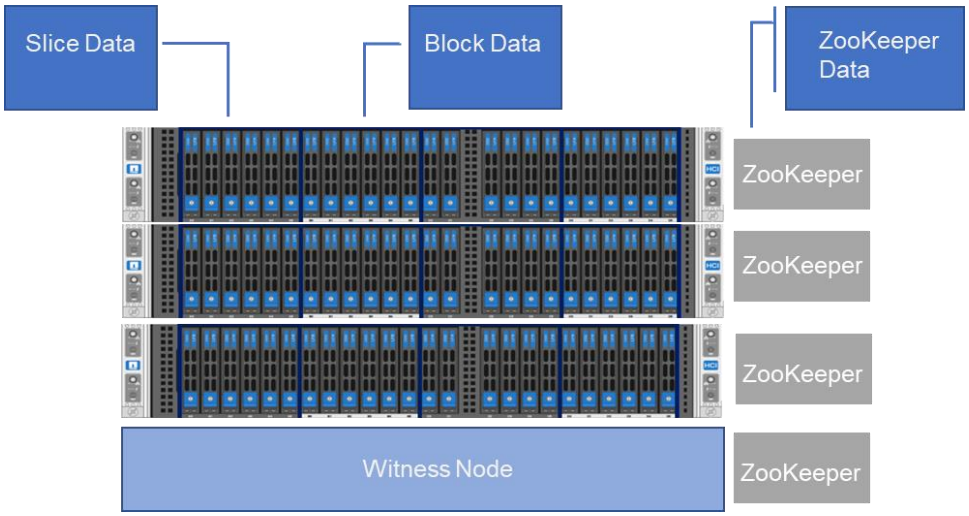
3-Node Cluster Differences and Limitations

The following list describes the differences and limitations of a 3-node cluster:

- No special behaviors from a cluster that has only three nodes.
- When additional storage nodes are required, the nodes (and their associated drives) can be added with no special steps.
- With a single-node failure, the cluster remains operational and fault tolerant.
- If there is enough available capacity to rebuild from a single-node failure, when a single node fails, the block and slice data heal back to a state where the data has single-node fault tolerance.
- When a single node fails, the ensemble remains operational with a fault that indicates that the ensemble is degraded.

Figure 6 shows a 3-node cluster. The three physical storage nodes are the ensemble members; the witness node is a standby node in the event of a physical storage node failure.

Figure 6) 3-node storage cluster ensemble.



3 Scalability

3.1 Scaling Behavior

NetApp HCI supports independant scaling of storage nodes and compute nodes. The scale-out architecture allows a storage cluster to grow when additional capacity or performance is needed. Similarly, a cluster can shrink when power, cooling, or space needs are reduced. Clusters are scaled by adding nodes to or removing nodes from the cluster; scaling is nondisruptive to client applications. Storage clusters cannot be created with fewer than two storage nodes or fewer than thre total nodes (two storage nodes and one witness node).

A 2-node storage cluster can be scaled to a storage cluster of four nodes or more. In this upgrade the new storage nodes take the place of the wtiness nodes. The witness nodes can then be removed and retired out of the cluster.

3.2 Scaling a 2-Node Storage Cluster

The 2-node or 3-node cluster can be expanded to a storage cluster of four or more nodes. The expansion nodes should be the same type and model as the 2-node or 3-node cluster storage until a minimum of four is reached. This enables more resiliency for auto-healing, adds capacity, and increases cluster IOPS. After the minimum of four is reached, additional expansion storage nodes can be added, including other storage node model types.

A storage node for scale that has less new capacity than the capacity of the largest node in the 2-node cluster is not allowed. It is a best practice to scale storage with nodes of the same type and capacity. Larger-capacity nodes can be used to scale; however, there will be some stranded capacity.

Scalability Options

Use these cluster scaling options:

- Scales from 2 compute + 2 storage nodes up to 64 compute + 40 storage nodes.
- Scaling from 2 to 3 storage nodes only increases resiliency, not additional capacity.
- Recommended scaling from two to a minimum of four storage nodes.

When scaling a 2-node cluster, NetApp recommends upgrading to a storage cluster of 4 nodes or more to add additional capacity and resiliency featuring Element cluster auto-healing capabilities. When a cluster is upgraded to four nodes, the physical node additions are added to the ensemble and the witness nodes are no longer required to reach a cluster quorum. Witness node VMs can be deleted manually, but deletion is not necessary.

When scaling from a 2-node to a 3-node storage cluster, capacity can be stranded. This upgrade only adds auto-healing resiliency for node and drive failures; it does not add capacity. The NDE scale process shows warnings about stranded capacity before installation.

For full information about scaling NetApp HCI, see the “Deploying NetApp HCI” section of the NetApp HCI Documentation Center.

3.3 Compute Nodes

Compute nodes can be expanded to any amount and model or type. The 2-node cluster refers only to the storage nodes running Element software. Scalability limits for compute nodes are the same as the limits of a 4-storage node cluster. When scaling compute nodes in a 2-node cluster environment, the witness node is not deployed to the new compute node. The witness node template is deployed to enable the user to create a witness node, if necessary.

3.4 Witness Nodes

Witness nodes can be added to or removed from a cluster if certain conditions apply. When a 2-node cluster is scaled to four storage nodes or more:

- Witness node is removed from the cluster.
- VM is deleted/removed from vCenter.
- Ensemble is configured to the max ensemble (3/5) size using physical storage nodes.
- When a storage cluster expands and the witness node is decommissioned, a message is sent to NetApp Active IQ®.
- Witness nodes can be installed from the witness node template, which is installed during NDE installation.
- Scaling to add compute nodes by NDE; the witness node is not deployed on the compute node. The witness node template is deployed by NDE for manual installation by the user.

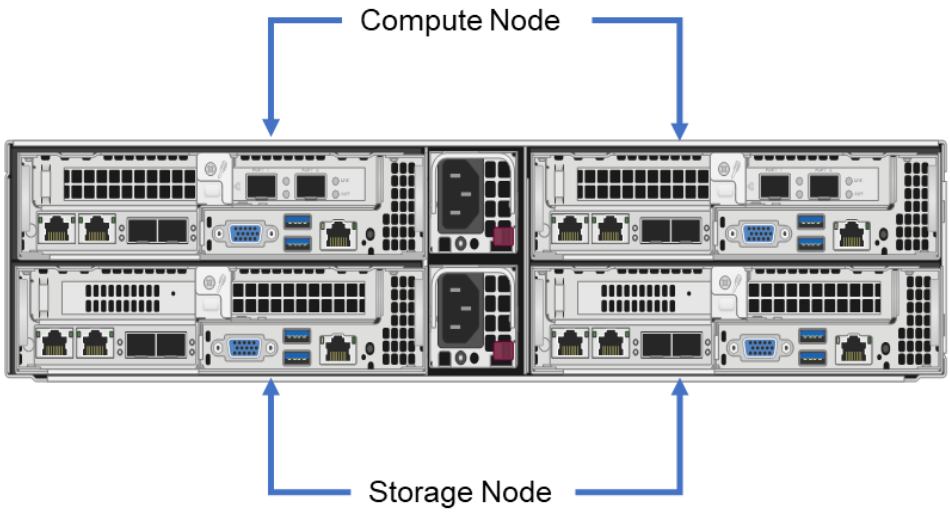
4 Installation

4.1 Hardware

The 2-node cluster for the initial release is for the 4U chassis and the 1U half-width compute and storage nodes. Table 1 lists the compatible models.

The node placement is the same as for any NetApp HCI installation. Figure 7 shows the recommended order for the compute nodes and storage nodes for a 2-node cluster.

Figure 7) Hardware chassis placement.



4.2 NetApp Deployment Engine

NDE supports installation with 2, 3, 4, or 4+ supported storage nodes.

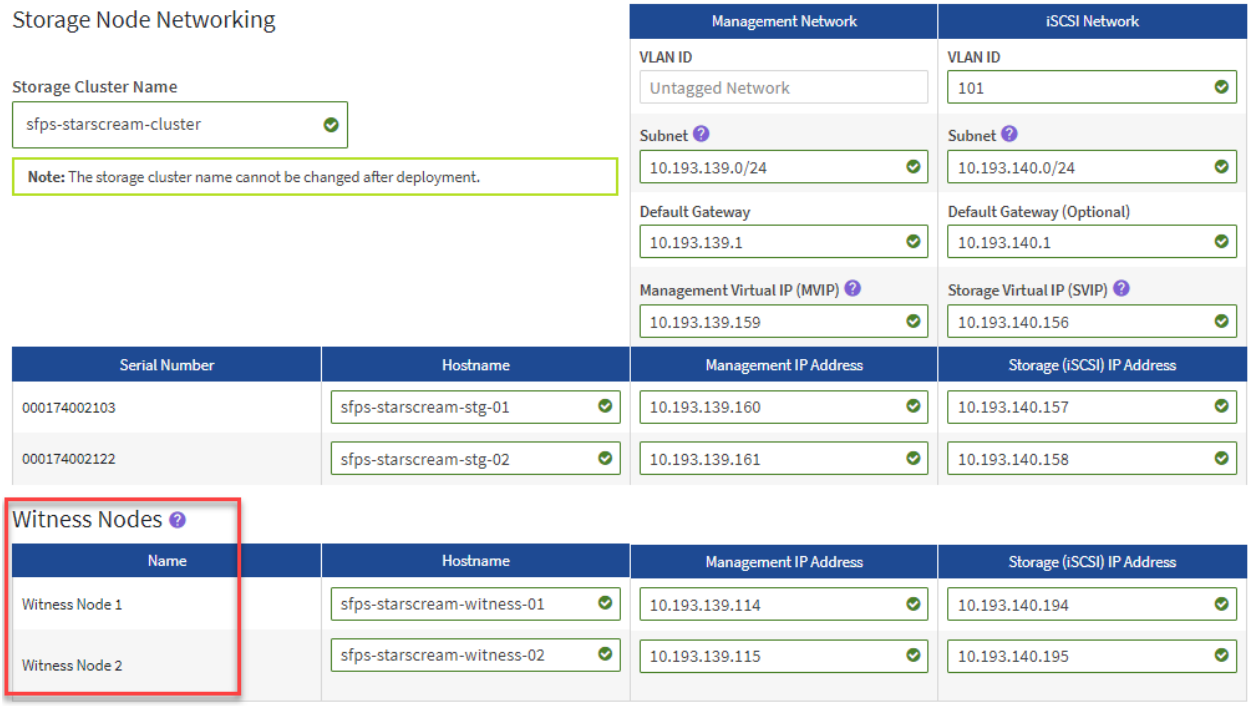
For a 2-node cluster, the minimum number of compute nodes required for an NDE installation stays the same, at 2 nodes, to provide HA. The witness node can be deployed on all supported versions of vSphere that NetApp HCI supports for the release that contains the 2-node cluster feature.

NDE deployment of a 2-node cluster installation follows these basic steps and requirements:

- NDE collects all necessary information from the NetApp HCI administrator.
- Witness nodes will be deployed with static management and storage IP addresses. This implies that the interfaces are associated to the correct management and storage PortGroups.
- Management IP is assigned to net0, storage IP is assigned to net1.
- NDE creates a datastore on local storage of the compute nodes for witness node deployment.
- NDE creates the witness node vitrual machines on the compute node local datastore.
- NDE creates an internal virtual network for witness node storage management.
- NDE creates an internal virtual network for witness node data.

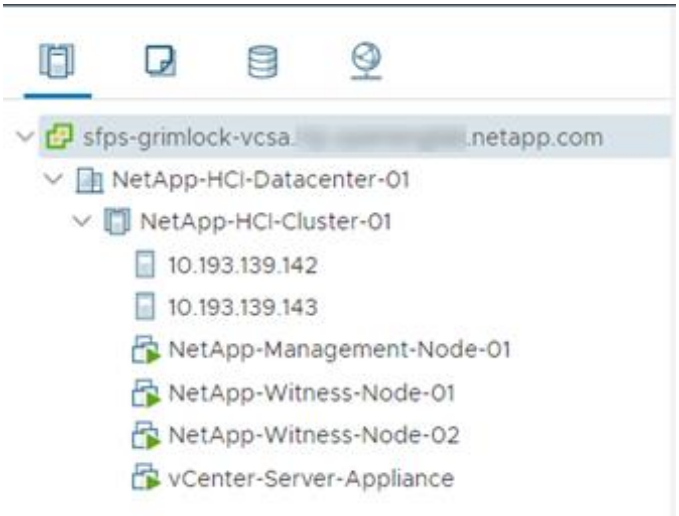
NDE creates the 2-node storage cluster. Figure 8 shows an example of the NDE installation information input for a 2-node storage cluster using existing IP addresses and FQDNs.

Figure 8) NDE 2-node cluster setup.



When installation is complete and VMware vSphere is ready to be accessed, Figure 9 shows an example of an installed 2-node deployment. For more information about installation, setup, management, and expansion, see the [NetApp HCI Documentation Center.]

Figure 9) Post-NDE installation in VMware.



5 Integration with VMware

5.1 Witness Node Virtual Machines

Witness nodes are deployed on all supported compute node models on the local disk by creating a local datastore during NDE installation. A minimum of two compute nodes with witness node VMs is required for 2-node clusters.

Table 2 lists the resource requirements and properties of the witness node VM when created in VMware.

Table 2) Witness node VM requirements.

Virtual Machine	vCPU	Memory	Disk Size
Witness node	6	12GB	67GB

Witness Node Properties

Witness node properties include:

- Witness nodes are never installed on top of vsphere datastores backed by Element storage.
- Witness nodes are deployed to all supported compute node models on the local M2 disk (and the datastore created on it).
- Witness nodes cannot be VMware HA, because it is backed by local storage of a compute node. When one witness node fails, the second witness node on another compute node takes over as a cluster ensemble node.
- Witness nodes are deployed as a thick provisioned VM.
- Witness node size is 67GB, using 7GB for installation and 60GB for log scratch.
- The compute ISO image contains the witness node VM image file.
- Witness nodes are deployed to all supported versions of vSphere at the time of deployment.
- Auto-deploy of witness node by NDE is supported for 6.5 and 6.7.
- Manual deployment of a witness node can only be done with the assistance of NetApp Support.
- VMware vSphere 6.0 is no longer a supported version in NDE for the 2-node node cluster release.

To enable the witness nodes to be deployed in VMware, an OVA from the compute node ISO is installed on a compute node local disk during NDE intallation. Additional networks are required for the witness node installation and for normal postinstallation operations. The networks are created using net0 and net1 and are added to the NDE-installed VDS for NetApp HCI.

The following networks are created in vSphere in the VDS created during NDE installation:

- Net0: Internal Storage Management Network
- Net1: Internal Storage Data Network

Figure 10 shows the additional VDS networks installed, highlighting the networks that are specific for witness nodes.

Figure 10) VMware witness node VDS networks.

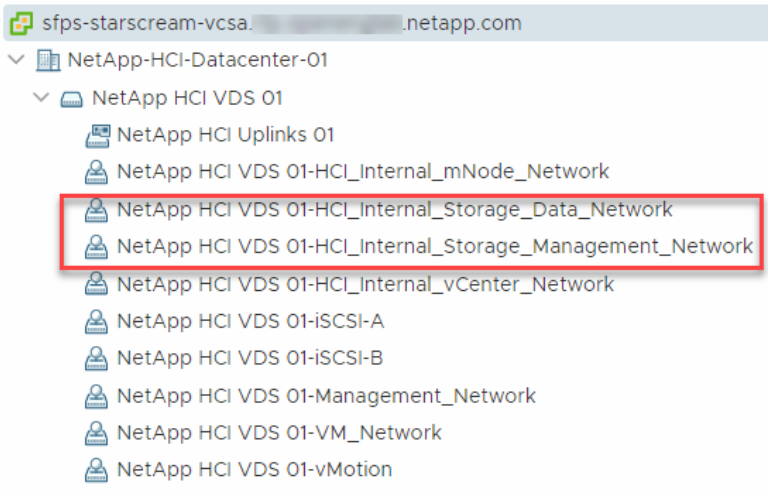


Figure 11 and Figure 12 are examples of VMware vSphere screens for the installed datastore and witness node VM.

Figure 11) Witness node datastore on compute node device storage.

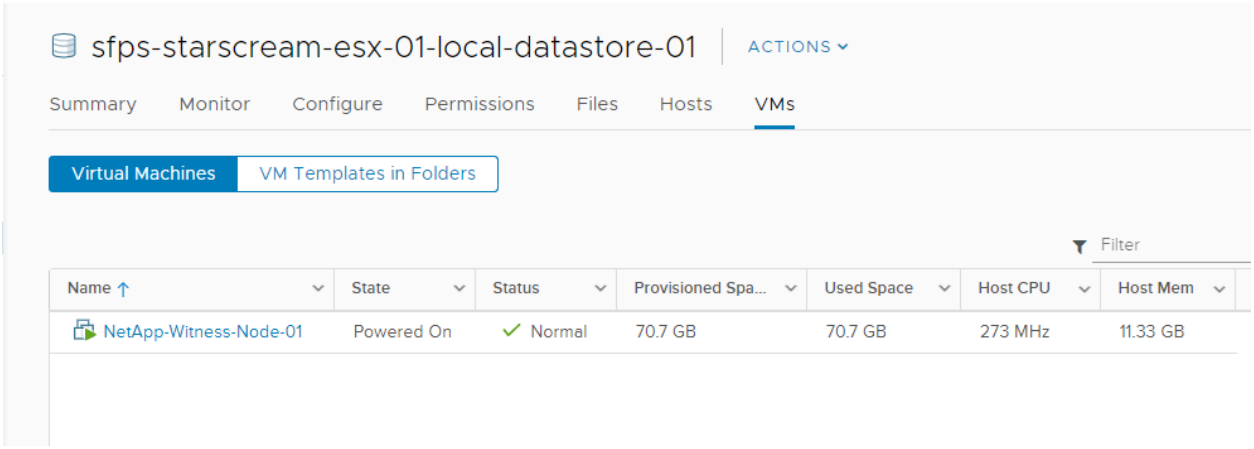
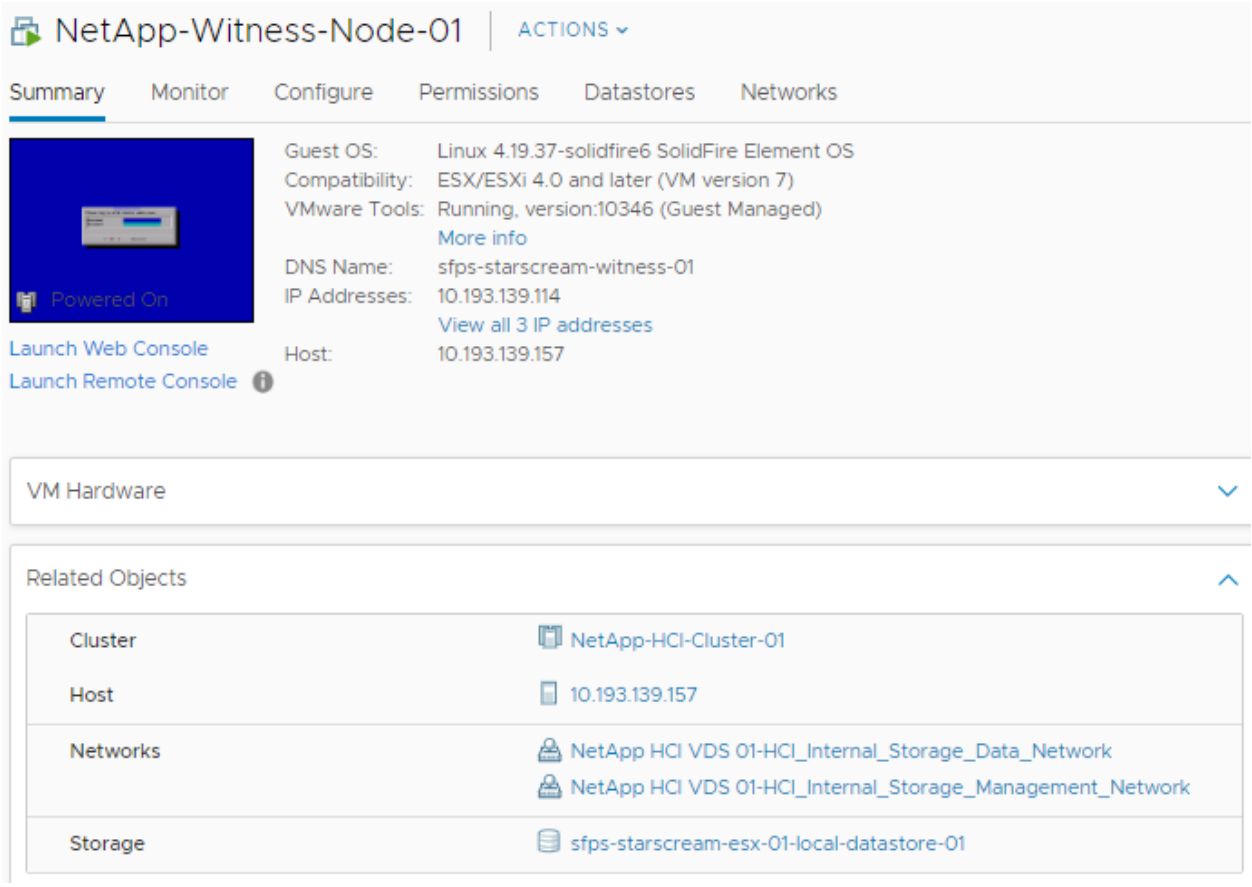
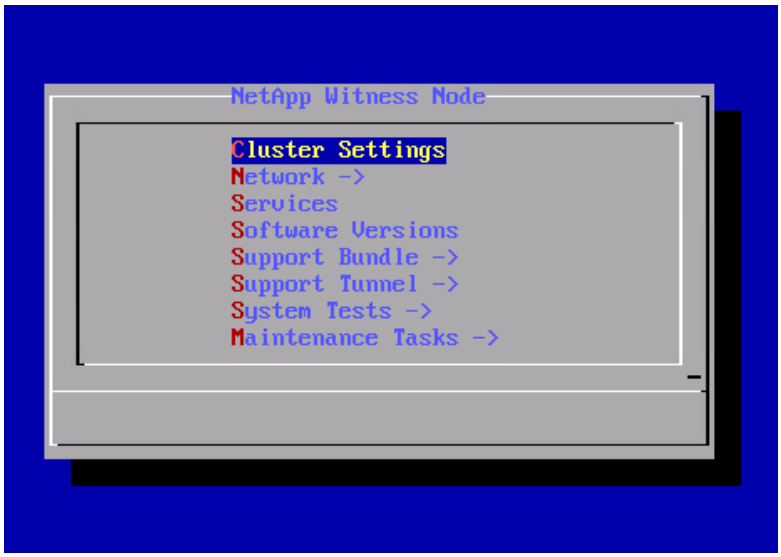


Figure 12) Witness node VM summary.



The witness node VM is installed as a Linux type. When connecting by the VMware remote console, the interface is not typical and is presented as a terminal user interface (TUI), which is the same as a physical storage node. To connect, the user enters the HCI configuration administrator username and password into NDE during installation. Figure 13 is an example of the menu options after logging in to the remote console.

Figure 13) Witness node remote console menu.



The witness node VMware networks net0 and net1 can be viewed or edited. Other common tasks are to view cluster information and to create and upload support bundles. For more information, see the NetApp HCI Documentation Center.

5.2 VMware Updates

VMware Update Manager (VUM) updates for ESX hosts must have the witness node disabled during the upgrade. VMware vMotion cannot move the witness node to another ESX host during an upgrade. This is because the witness node resides on the local disk of the compute node and not on shared storage. A disabled witness node is not a functioning Element ensemble member during the ESX host updates.

Mitigation

Only one compute node can be updated at a time. The VUM update must be done to a single ESX host. Before updating ESX, the end-user must power off the witness. The ESX host on the compute node can then be placed in maintenance mode, while the ESX host on the second compute node retains the witness node as the functioning ensemble member. After updates are completed to the initial ESX host and removed from maintenance mode, the witness node can be powered on to rejoin the ensemble. The ESX host on the second compute node can then safely repeat the process for updates.

6 Conclusion

The NetApp HCI 2-node cluster is an entry-level solution for customers that takes advantage of NetApp HCI performance, reliability, and multitenancy QoS control just like larger systems and clusters. The 2-node cluster creates a NetApp HCI data center footprint with growth flexiblity. When customer workloads need additional performance or capacity to meet business requirements, the 2-node cluster is easy to scale to larger configurations to meet demands.

Where to Find Additional Information

To learn more about the information described in this document, refer to the following documents and websites:

- NetApp HCI datasheet
  https://www.netapp.com/us/media/ds-3881.pdf

- NetApp HCI product page
  https://www.netapp.com/us/products/converged-systems/hyper-converged-infrastructure.aspx
- NetApp HCI documentation resources
  https://www.netapp.com/us/documentation/hci.aspx
- SolidFire cloud storage product page
  https://www.netapp.com/us/products/storage-systems/all-flash-array/solidfire-scale-out.aspx
- SolidFire all-flash storage documentation resources
  https://www.netapp.com/us/documentation/solidfire.aspx

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

Copyright Information

Trademark Information

TR-4823-0420