Technical Report

Best practices for modern SAN

ONTAP 9

Michael Peppers, NetApp
October 2025 | TR-4080

Abstract

This technical report provides an overview of the block protocols in NetApp® ONTAP® 9 data management software as well as best-practice recommendations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

This technical report provides a comprehensive overview of ONTAP blocks, including ONTAP SAN on unified platforms such as FAS and AFF, as well as the All SAN Array (ASA). For in-depth ASA best practices, refer to TR-5009: Best Practices for NetApp ASA. NetApp ONTAP is a robust data-management platform offering native features like inline deduplication and compression, nondisruptive hardware upgrades, and LUN importation from foreign storage arrays. Clusters can include up to 12 nodes, simultaneously serving SAN data via iSCSI, Fibre Channel (FC), and NVMe protocols. NetApp Snapshot technology enables rapid creation and cloning of backups, supporting extensive disaster protection and recovery capabilities.

Overview

NetApp ONTAP 9.17.1 marks the twenty-first major clustered ONTAP release supporting SAN protocols since their introduction in clustered Data ONTAP 8.1. This report reviews clustered SAN and NetApp ASA implementations from the perspective of SAN-attached hosts, detailing prescribed NetApp AFF SAN configurations for optimal performance and best practices for leveraging ONTAP's high-availability and data mobility features.

Audience

This document targets system and storage architects designing iSCSI, FC, NVMe/FC, and NVMe/TCP solutions with NetApp storage running ONTAP 9.10.x or later. Readers should possess a general understanding of NetApp hardware and software and be familiar with block-access protocols including iSCSI, FC, NVMe/FC, and NVMe/TCP.

Caveats

This report is not intended as a general ONTAP administration introduction. Foundational information is covered in the ONTAP 9 System Administration Reference, ONTAP 9 SAN Administration Guide, ONTAP 9 SAN Configuration Guide, NetApp Host Utilities Kits, and host OS-specific documentation—available on the ONTAP SAN host page.

All ONTAP and SAN limits are available in the NetApp Hardware Universe (HWU). The HWU page is the definitive source for these limits. Always consult the Interoperability Matrix Tool (IMT) on the NetApp Support site to verify that your environment's product and feature versions are supported. The IMT details the product components and versions tested and qualified by NetApp. Specific results rely on each customer’s installation adhering to published specifications. Configurations not listed as IMT-qualified may function but are unsupported and will receive best-effort support from NetApp.

Summary of best practices

Here’s a summary of the ONTAP SAN best practices that will be discussed throughout the rest of this technical report:

- Create FCP, iSCSI, and NVMe-oF services concurrently with SVM (storage virtual machine) creation.
- Increase the number of volumes and LUNs to optimize performance—ideally about 8 volumes and 8-16 LUNs.
- Assign one LIF per Ethernet network or FC fabric on each storage controller serving data via iSCSI, FC, and the NVMe-oF transports – NVMe/FC and NVMe/TCP.

- Enable N_Port ID virtualization (NPIV) on FC fabrics connected to clustered ONTAP storage clusters.
- Configure zoning to use only NPIV virtual WWPNs as targets; avoid using physical WWPNs for target ports.
- Selective LUN mapping: Most LUNs should have four paths—two direct and two indirect, corresponding to the storage controller and its HA partner. Update LUN mappings when moving LUNs to a new HA pair within a cluster.
- Add additional paths as needed for data mobility or increased I/O resources, but do not exceed the host OS's maximum supported path count.
- Follow a standardized host procedure for LUN mapping changes to ensure new paths are discovered and obsolete paths are discarded.
- SVMs serving data via FCP, iSCSI, or NVMe-oF require a management interface.
- When moving deduplicated or compressed LUNs, ensure the destination volume has these policies enabled.
- Utilize LUN move's pause and throttle features for granular control of LUN replication.
- Use LUN move to streamline data mobility and replication workflows.
- Do not exceed cluster size limits for SAN data clusters as specified in NetApp HWU.
- Install the Host Utilities Kit on hosts accessing LUNs.
- Install ONTAP tools for VMware to optimize your VMware/ONTAP SAN experience
- Set In Order Delivery on all FC switches in your fabrics.

ONTAP overview

Storage controllers running ONTAP are known as nodes, which are organized into a clustered system. The nodes within the cluster maintain continuous communication to coordinate activities and transfer data seamlessly between nodes via redundant paths on a dedicated cluster network, consisting of two 100 Gigabit Ethernet (100GbE) switches. Additionally, a single HA pair can form an ONTAP cluster as a two-node-switchless configuration.

While the node is the basic unit of a cluster, nodes are integrated as HA pairs. HA pairs provide high availability by communicating over an HA interconnect (separate from the main cluster network) and maintaining redundant connections to their respective disks. Disks are not shared across different HA pairs.

Clusters are managed at the cluster level rather than per node. Data is delivered from one or more Storage Virtual Machines (SVMs). SVMs are set up to own storage resources such as volumes (and LUNs) provisioned from physical aggregates, with LIFs allocated to either physical Ethernet NICs or FC target HBA ports. LUNs are created inside SVM volumes and mapped to hosts for storage access. SVMs are independent of nodes and operate at the cluster level; they use physical resources including volumes or network ports located anywhere in the cluster. ONTAP functions as a storage hypervisor, enabling abstraction and utilization of resources such as disk, CPU, memory, Ethernet, and FC ports for SVMs.

High availability

- A full explanation of ONTAP high availability features is outside the scope of this document. However, a general understanding of these features is considered when designing storage infrastructure. For information on HA in an ASA configuration, please review TR-5009: Best Practices for NetApp ASA.

HA pairs

The primary structure for high availability is the HA pair, which consists of two storage controllers linked by two redundant cluster interconnect cables. Both nodes can access all disks associated with the HA pair. Each controller within the pair can manage its own data along with a virtual instance of its partner and corresponding data.

NVRAM

Each HA pair includes redundant links for NVRAM data replication. NVRAM does not serve as write cache; instead, the RAM within the controller is used for that purpose. NVRAM's function is to temporarily record data to safeguard against unexpected system failures, similar to how database transaction logs operate. Both NVRAM and database logs allow for rapid storage of data changes, which are later written to persistent drives during a checkpoint process. Normally, neither NVRAM nor database redo logs are accessed during standard operations.

In the event of a controller failure, pending changes may remain in NVRAM before being written to the drives. The partner controller identifies the failure, assumes control over the drives, and, if there are and dirty buffers in NVRAM (any outstanding IOs in NVRAM that have not been written to disk, those will be written to disk. This is the only time that ONTAP writes NVRAM contents to disk.

Figure 1) Two-node switched A1000 cluster

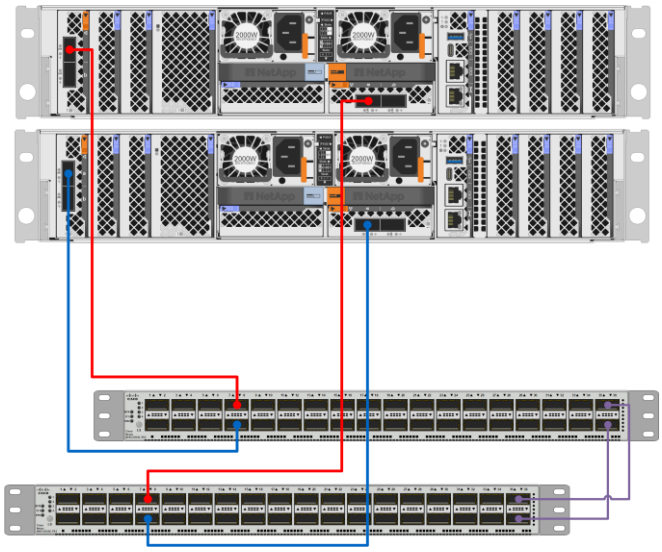
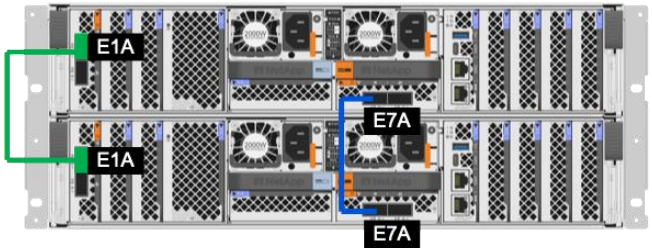


Figure 2) Two-node switchless configuration



Cabling redundancy

Figure 1 illustrates a typical high-availability (HA) pair cabling configuration; however, the specific arrangement may differ depending on the controller and drive shelf type. Regardless of layout, redundant data paths—whether through physical cables or electrical traces on a backplane within a chassis—are always present. Each controller maintains at least two connections to other controllers within the cluster, enabling NVRAM-based replication and facilitating intra-cluster communications for I/O operations and nondisruptive data relocation.

Power redundancy

All controllers, drive shelves, and associated components are equipped with redundant power supplies. Standard deployments position systems within server racks utilizing dual Power Distribution Units (PDUs), each connected to separate UPS-backed circuits within the data center, thereby enhancing power reliability.

Takeover and giveback

Takeover and giveback describe the procedures by which storage resource responsibility is transferred between nodes in an HA pair. This process encompasses two primary aspects: management of network connectivity for host access to the storage system, and management of drives within the storage array. Network interfaces supporting SAN block protocols (such as iSCSI, FC, NVMe/FC, and NVMe/TCP) are not instantaneously moved during takeover or giveback. Should a controller fail, its partner assumes data service provision using its own interfaces. if the node is an ASA, then persistent ports and/or iSCSI LIF failover are enabled and supported, failed interfaces are reactivated on the partner by either migrating the IP address (iSCSI LIF failover) or relocating the HBA WWNN (for FCP and NVMe/FC persistent ports), ensuring hosts do not encounter lost storage paths.

Note: Additional controller paths can be configured to enable data relocation across larger clusters.

The transfer of disk ownership constitutes the second aspect of takeover and giveback. The precise procedure varies based on factors such as the event's cause and command line options provided. The objective is to optimize operational efficiency; while the entire process may span several minutes, the actual handoff of drive ownership typically occurs within seconds.

Takeover triggers

Takeovers may be initiated under various circumstances, including:

- Manually triggering a takeover via the storage failover takeover command - `storage failover takeover`
- Occurrence of software or system failures resulting in a controller panic; following reboot, resources are returned to normal operation.
- Complete system failure of a controller, such as loss of power, preventing reboot.
- Failure of a partner controller to receive a heartbeat signal, possibly due to hardware or software issues that impede normal functioning without causing a panic.

Non-Disruptive Operations (NDO)

Nondisruptive operations encompass both contingency response to sudden controller failures and support for online upgrades and maintenance. After a controller relinquishes data services to its partner, administrators may upgrade the ONTAP OS, replace defective hardware, install new adapters, or even update the controller itself—all while maintaining uninterrupted service.

Takeover time

ONTAP unified (AFF/FAS)

In a takeover, one HA node assumes control of its partners' disks and re-advertises files, directories, and storage units as the new host and owner. ONTAP quickly completes the initial transition in under two seconds. However, IO Resume Time (IORT) better reflects how fast the system returns to normal

operation and varies significantly by OS. Linux and VMware respond quickly, while Windows and legacy UNIX systems show more variability and delays. These differences stem mainly from host settings for path recognition and retry behavior.

Table 1) Unified failover times

Failover Type	IO Resumption Time
Planned takeover	2 - 3 seconds
Unplanned takeover	2 - 3 seconds

NetApp ASA

NetApp ASA offers symmetric active-active paths across controllers, ensuring that hosts experience no delays during path switching—every path remains active regardless of system state or failover. Each controller continuously replicates key LUN metadata to its partner, enabling rapid SAN failover by allowing immediate data access even if a controller fails before the failover process completes.

Features like FC's persistent ports and iSCSI LIF failover also support seamless migration of target endpoints to surviving connections, preventing path disruptions and ensuring uninterrupted I/O operations for hosts.

The measured time reflects full I/O resumption (IORT) at the operating system level. While storage response may resume faster, the most relevant metric is when the host fully resumes I/O.

Table 2) ASA failover times

Failover Type	IO Resumption Time
Planned takeover	2 - 3 seconds
Unplanned takeover	2 - 3 seconds

How ONTAP ensures data integrity

Data integrity

Logical data protection in ONTAP involves three main requirements:

- Protection against data corruption.
- Protection against drive failure.
- Safeguarding changes to data against loss.

The following sections address these three aspects.

Network corruption: checksums

Checksums are basic error-detecting codes stored with data to identify corruption during transmission. For instance, an FC frame uses a cyclic redundancy check (CRC) that is recalculated on receipt; if it doesn’t match, the frame is discarded or rejected. Similarly, iSCSI I/O operations use checksums at multiple protocol layers, with optional CRC protection at the SCSI layer. Any detected errors result in packet retransmission or operation rejection.

Drive corruption: checksums

Checksums verify data integrity on drives by storing a unique number generated from each data block. When reading, the checksum is recalculated and matched against the stored value; any mismatch signals corruption, requiring recovery via the RAID layer.

Data corruption: lost writes

One of the most challenging forms of data corruption to identify is either a lost or misplaced write. When a write operation is confirmed, it must be accurately recorded onto the correct media location. In-place data corruption can typically be detected through a straightforward checksum stored alongside the data. However, if a write is lost, the previous version of the data may persist on the media, and the checksum for those underlying blocks will appear valid. Similarly, writes placed incorrectly at a physical location can result in seemingly legitimate checksums, despite overwriting existing data.

To address these issues:

- Each write operation should include metadata specifying its intended storage location.
- A version identifier should be integrated with every write.

When ONTAP writes a block, it embeds information about the block's expected location. If a subsequent read locates a block with metadata indicating it belongs at location 123 but finds it at location 456, this signals misplacement.

Detecting a lost write is more complex. ONTAP approaches this by storing metadata so that one write operation updates two separate locations on the drives. If a write is lost, a future read of both the data and its associated metadata will reveal disparate version identities, confirming the write was not successfully completed.

While instances of lost or misplaced write corruption are exceptionally rare, the risk grows as storage devices scale to accommodate ever-larger datasets. Implementing lost write detection is essential for any storage system tasked with managing critical data.

Drive failures: RAID, RAID DP, and RAID-TEC

If a block of data on a drive is found to be corrupt or the entire drive fails and becomes unavailable, the data must be recomputed. In ONTAP, parity drives are used for this process. Data is distributed across multiple data drives, with parity information generated and stored separately from the original data.

ONTAP initially implemented RAID 4, which utilizes a single parity drive for each data drive group. While RAID 4 assigns a dedicated parity disk, RAID 5 distributes the parity stripes among all drives in the group through "floating parity." Both RAID 4 and RAID 5 can withstand the loss of one drive without data loss. If the parity drive fails, the data remains intact, and a new parity drive can be reconstructed. When a single data drive is lost, its contents can be regenerated using the remaining drives and the parity drive.

With smaller drive capacities, the likelihood of two drives failing simultaneously was low. As drive sizes increased, the time needed to reconstruct data after a failure grew, extending the period during which a second drive failure could result in data loss. The rebuild process also increases I/O on surviving drives until completed, and older drives may face heightened risks under additional load, potentially causing further failures. Moreover, larger drive capacities mean that the consequences of any data loss become more significant, impacting recovery times and business continuity.

To address this, NetApp introduced NetApp RAID DP® (Dual Parity) technology, based on RAID 6 principles. RAID 6 uses two floating parity calculations across drives, whereas RAID DP uses two dedicated parity drives for horizontal and diagonal parity. These two parity drives allow any two drives in a RAID group to fail without data loss. Continued growth in drive capacities led to the development of NetApp RAID-TEC™ (Triple Erasure Coding), which enables survival of up to three drive failures.

Historical SAN best practices have sometimes favored RAID-10 (striped mirroring) for performance reasons, but RAID-10 offers less protection than RAID DP as it has scenarios where two-disk failures result in data loss, unlike RAID DP. Some legacy documentation recommends RAID-10 over RAID4/5/6

due to concerns about performance penalties related to parity regeneration. Traditional RAID systems require extra disk reads to regenerate parity during write operations, known as a RAID penalty.

In ONTAP, this penalty does not occur; writes are staged in memory, where parity is calculated before being written to disk as a complete RAID stripe, eliminating the need for read operations during writes.

Overall, compared to traditional RAID configurations, RAID DP and RAID-TEC offer higher usable capacity, improved resilience to drive failure, and do not reduce performance.

Hardware failure protection: NVRAM

For storage arrays handling latency-sensitive workloads, it is essential to promptly acknowledge write operations while ensuring data is protected against loss from unexpected events such as power outages. Consequently, every write must be securely stored in at least two distinct locations.

ONTAP systems utilise NVRAM to fulfil these requirements. The write process proceeds as follows:

- Inbound write data is initially stored in RAM.
- Required changes to disk data are then journaled into NVRAM on both the local and partner nodes. Unlike a traditional write cache, NVRAM functions as a journal analogous to a database redo log. It is not accessed during regular operations and is only read during recovery scenarios, such as following a power failure during I/O processing.
- After the write is copied to NVRAM, an acknowledgment is sent to the host.

At this point, from the application's perspective, the write operation is complete and the data is safeguarded due to its presence in two separate locations. The subsequent writing of changes to disk occurs asynchronously and does not impact application latency; this approach is similar to database logging, where changes are swiftly recorded in redo logs before being marked as committed, with updates to datafiles taking place later.

In the event of controller failure, the partner controller assumes control of its counterpart's disks and replays the NVRAM journal to recover any outstanding I/O operations that were in progress at the time of failure.

Redundancy failure: NVFAIL

A write is only acknowledged after being logged in local NVRAM and mirrored to another controller’s NVRAM, preventing data loss during hardware failures or power outages. If local NVRAM or node connectivity fails, mirroring stops. On local NVRAM errors, the node shuts down and failover occurs—either to a partner or remote controller, depending on HA or MetroCluster setup. No data is lost because unacknowledged writes aren’t considered persistent; only acknowledged writes are guaranteed to be stored. Until a write is acknowledged, its state remains uncertain.

ONTAP SAN

Considerations for optimizing SAN performance

ONTAP offers two main SAN platforms: ONTAP unified (FAS/AFF) and All SAN Array (ASA). Specific performance considerations and optimizations apply to one or the other platform and will be noted accordingly.

ONTAP efficiently utilizes processor cores for concurrent workloads, distributing many operations across multiple cores. Some processes, however, are limited to a single core, which can cap performance for those tasks. While non-distributable threads may constrain total output, spreading workloads over several objects ensures full core engagement. Traditionally, increasing volumes and LUNs has boosted performance, but improvements in parallelization have enhanced outcomes even with fewer or single LUNs. Ongoing development continues to improve efficiency. The sections below show possible performance gains for demanding applications, though most customer workflows may not require such configurations.

Volumes

The discussion of volumes in a SAN context applies to unified ONTAP (FAS and AFF), but not to ASA, where there is a 1:1 relationship between volumes and storage units (LUNs or namespaces). In ASA, volumes are abstracted from administration, and their impact on performance is minimal due to architectural changes.

When provisioning volumes for an application, it is important to consider the function of a volume in a SAN environment. Different vendors may define "volume" differently; in ONTAP SAN, volumes serve as management containers for hosted SUs, simplifying the management of multiple SUs within a single volume. Volumes also function as consistency groups, since snapshot copies taken at the volume level include all blocks within the volume, capturing all associated LUNs. Storage efficiency features operate primarily at the volume level, so larger volumes generally increase storage efficiencies due to more shared data blocks.

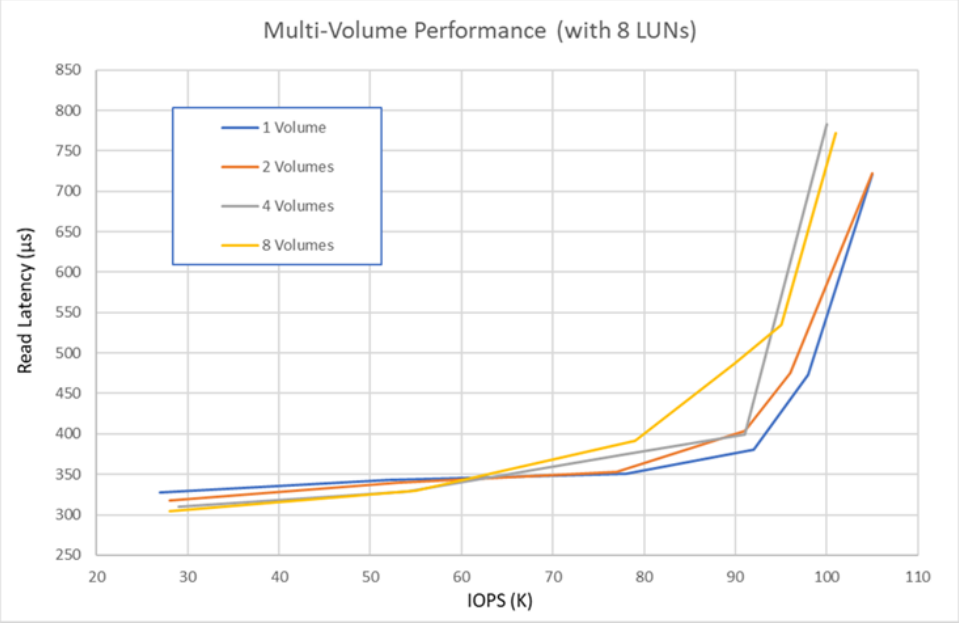
While larger volumes facilitate administration, support consistency groups, and tend to improve efficiency, the number of volumes should also be considered for performance optimisation. Some processing threads cannot be distributed across processor cores; spreading workloads across several volumes enables non-distributable threads to run concurrently on different cores, improving throughput.

Balancing the grouping of multiple LUNs within a single volume against increasing the total number of volumes is necessary. Based on testing and customer experience, NetApp recommends provisioning between 4 and 16 volumes, ideally within the 8–16 range. Increasing volumes beyond this does not typically enhance performance and may add complexity while reducing storage efficiency.

Error! Reference source not found. and Figure 4) Effects of spreading work across multiple LUNs.

show the effects of adding more volumes and LUNs; they are displayed for illustrative purposes only. They are not meant to guarantee or benchmark specific workloads or provide estimates of performance.

Figure 3) The effect of spreading work across more volumes.



Volume configuration

When provisioning volumes in a cluster, considerations such as deduplication, space reservations, and storage efficiency remain consistent. A key distinction is that volumes on ONTAP storage clusters are assigned to SVM containers rather than individual nodes, which allows them to be mapped into an

SVM-wide global namespace for exporting file systems with NFS or CIFS protocols. The inclusion or exclusion of a specific volume in the global namespace does not impact data served by ONTAP SAN.

Storage Units (SUs)

NetApp now uses "storage units" as a generic term for containers of blocks, such as SCSI LUNs and NVMe namespaces. LUN and namespace will refer only to specific container types in the future.

Change to Storage unit/Volume geometry best practices

Recent ONTAP releases (post 9.10.1) have optimized processing so that splitting workloads across multiple LUNs or volumes is no longer needed for peak performance. Improved protocol efficiencies and single LUN/single namespace enhancements mean customers can achieve high performance without following previous recommendations about workload distribution.

Storage units (prior to ONTAP 9.10.1)

LUNs, whether iSCSI or FC, have certain threads that cannot be distributed across multiple processor cores. Therefore, using more smaller LUNs rather than fewer larger ones may improve performance. Spreading workload across several LUNs enables more processor cores to concurrently process I/O.

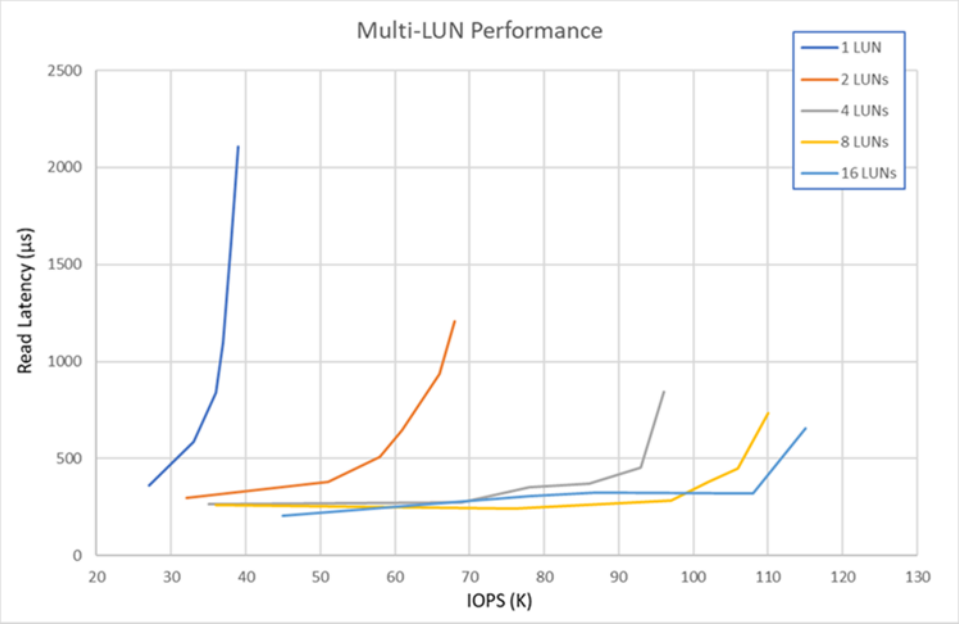
One way to boost application performance is to distribute workloads across additional LUNs. Techniques for increasing the number of LUNs used by an application include:

- **Logical volume managers (LVMs)**, which combine multiple LUNs into one volume for the host OS or application. LVMs are typically used with Linux and UNIX operating systems. Oracle ASM can also aggregate multiple LUNs as a single storage object.
- In some situations, presenting multiple LUNs to an OS or application may be appropriate if the OS or application manages the presented LUNs directly.

Best practice

For optimizing most workloads, NetApp recommends using more smaller LUNs versus fewer bigger ones. Ideally between 8-16 LUNs.

Figure 4) Effects of spreading work across multiple LUNs.



FC in-order delivery

FC switches should be configured to ensure in-order delivery (IOD). While this step isn't necessary for ONTAP operations, it is a best practice because ONTAP will drop the exchange when an out of order or dropped frame is encountered. As a result, ONTAP must rely on the initiator (host) to retransmit the frame when the initiator hits its SCSI timeout threshold. This process might take 60 seconds. ONTAP will survive and recover from this situation but at a cost of the latency caused by the SCSI timeout and retransmit times.

If IOD is configured on all FC switches in the fabric, ONTAP won't receive any out of order frames and therefore, won't endure long host SCSI timeouts while awaiting frame retransmits.

Conclusions about number of volumes and storage units needed to optimize performance

There is essentially little performance benefit from additional volumes. Any variation is just a margin of error. Error! Reference source not found. **and Figure 4) Effects of spreading work across multiple LUNs.**

show a single volume with eight LUNs is delivering an easy 100K random IOPS at good latency. The inference that can be drawn from this is that you can increase application performance by increasing the number of storage units used with a given application. While there are some performance improvements by increasing the number of volumes, those performance improvements are much smaller than those seen from increasing the number of storage units. In both cases, returns from increasing the number of volumes tend to be smaller. Furthermore, increasing the number of volumes can lead to diminishing returns to scale.

Note: On rare occasions, you can spread your workload over multiple volumes, but this mostly applies to cases where a single application is consuming all the capabilities of a controller. For example, if you have a one large database that needs to push 500K IOPS, and you want to minimize every microsecond of latency, then you need more than one volume. If you are in this situation, you should be working with a solution architect to consider all aspects of the configuration, not just the number of volumes in use.

Based on the data presented above, we can draw the following inferences:

- Spreading your work across multiple storage units improves performance significantly.
- A single storage unit can support approximately 35K IOPS. Two storage units almost doubles the limit.
- Benefits start diminishing as you reach eight storage units. This is why TR-3633: Oracle Databases on ONTAP recommends using four to eight storage units for databases. Using four LUNs is acceptable, but eight is slightly better.
- A single volume with eight storage units delivers an easy 100K random IOPS at good latency, which is more I/O than 99% of all databases require.

Error! Reference source not found. **and Figure 4) Effects of spreading work across multiple LUNs.**

Note: show that the numbers in the conclusions previously discussed were from a specific test. The conclusions are valid, and they illustrate the NetApp best practice recommendations. However, the specific numbers listed are included to illustrate the concepts and recommendations. They shouldn't be taken as guarantees or guidelines for what a given volume, storage unit, or application can achieve regarding performance.

Scalable SAN

When an SVM is first created and a block protocol (FC, iSCSI, or NVMe) is enabled, the SVM gets an FC worldwide name (WWNN), an iSCSI qualified name (IQN), or an NVMe qualified name (NQN) respectively. This identifier is used regardless of which physical node is being addressed by a host, with

ONTAP making sure that SCSI and NVMe target ports on all of the cluster nodes work together to present virtual, distributed SCSI and NVMe targets to hosts that are accessing block storage.

In practice, this means that no matter which physical node a host is communicating, it is communicating with the same SCSI or NVMe target. This method of access presents new opportunities for data resiliency and mobility, and it also has implications for best practices when accessing data using block protocols on a cluster.

Best practice

When creating iSCSI, FC or NVMe LIFs for the first time for an existing SVM, make sure that the FC, iSCSI and/or NVMe service for that SVM has been created and is turned on by using the `fcp show`, `iscsi show`, and `vserver nvme show` command or by navigating to the Storage Virtual Machines > Configuration > Protocols pane in System Manager.

Note: This step is not necessary if the SVM was originally set up to serve these protocols by using an automated process such as the System Manager SVM Setup wizard.

Note: This step isn't necessary on ASA because the SVMs are created with block protocols already licensed and enabled.

Host connectivity

Hosts accessing ONTAP storage clusters via block protocols use the Asymmetric Logical Unit Access (ALUA) SCSI extension to distinguish direct (active/optimized) and indirect (active/nonoptimized) paths to LUNs. ALUA information is exchanged over the same iSCSI or FC connection as data. NVMe uses Asymetric Namespace Access (ANA) for similar path management. Hosts discover path status by sending inquiries down each known path; these can be triggered by extra data in SCSI responses indicating a need to update priorities. ALUA and ANA are widely-used standards required for ONTAP block data access and supported by all qualified operating systems.

Path selection

Each SVM's LIF handles both read and write requests for its storage units, but only one cluster node manages the supporting disks at any given time. Paths to storage units are either direct—where the LIF and storage unit are on the same node, avoiding the cluster network—or indirect—where data travels through the cluster network between different nodes. Direct paths are preferred for efficiency, with redundancy provided by multiple connections and networks.

ALUA helps hosts select direct paths first, minimizing reliance on indirect paths in typical operations. ALUA and ANA relay path status to the host server, whose MPIO stack manages path usage, prioritizing active optimized routes; administrators can adjust these settings if needed.

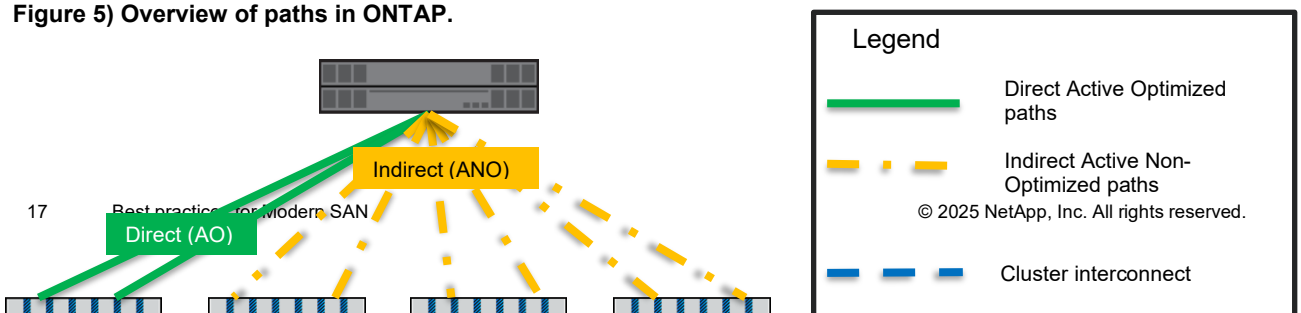
Indirect paths, involving inter-node transfers over the fast, high-availability cluster network, add minimal latency but are less efficient. In SAN environments, hosts mainly use direct paths managed via the host's MPIO software.

Administrators familiar with NAS protocols should note that SAN LIFs differ: NAS LIFs can move across nodes or failover groups, while SAN LIFs establish simultaneous block protocol connections managed by multipathing software. Instead, SAN redundancy is achieved by presenting multiple paths.

Fibre Channel persistent ports and iSCSI LIF failover both reassign target endpoints to functioning ports on HA partners, reducing disruptions during failover events.

Note: Due to these differences, Ethernet LIFs serving iSCSI cannot also serve NAS traffic, though NVMe/TCP and iSCSI may share a LIF. ASA allows iSCSI and NVMe/TCP to operate on shared LIFs.

Figure 5) Overview of paths in ONTAP.



Best practice

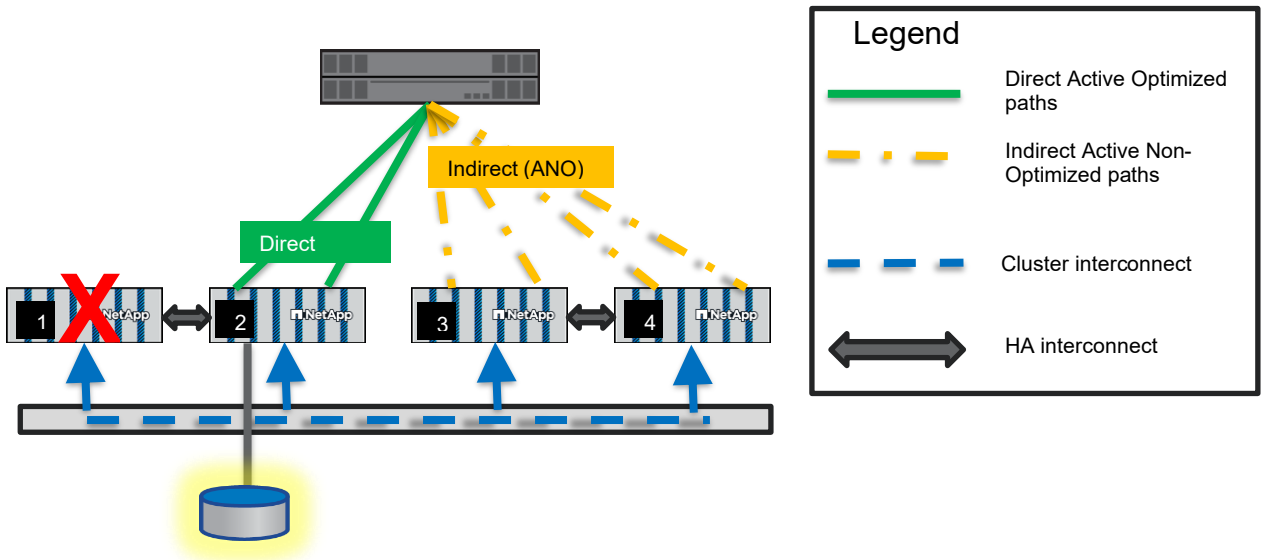
All SVMs should be assigned LIFs on each cluster node and on each FC fabric or Ethernet network. For instance, if a four-node cluster is connected to two independent FC fabrics, A and B, using its 3a and 4a FC target ports, an SVM that serves data by using FC should have eight LIFs, on node1:3a, node1:4a, node2:3a, node2:4a, node3:3a, node3:4a, node4:3a, and node4:4a. Clusters with more than four nodes should limit the number of paths used to access any given LUN for ease of manageability and in deference to operating system path count limitations. For a more in-depth discussion, see the section titled, “Path management and selective LUN mapping.”

Path selection changes

Three events can alter a host's data access path on a cluster:

HA failover. During HA failover, LIFs on the failed node go offline, while those on the HA partner become direct paths. This switch occurs automatically via ALUA path inquiry, requiring no manual intervention. Figure 5 illustrates these paths during HA failover.

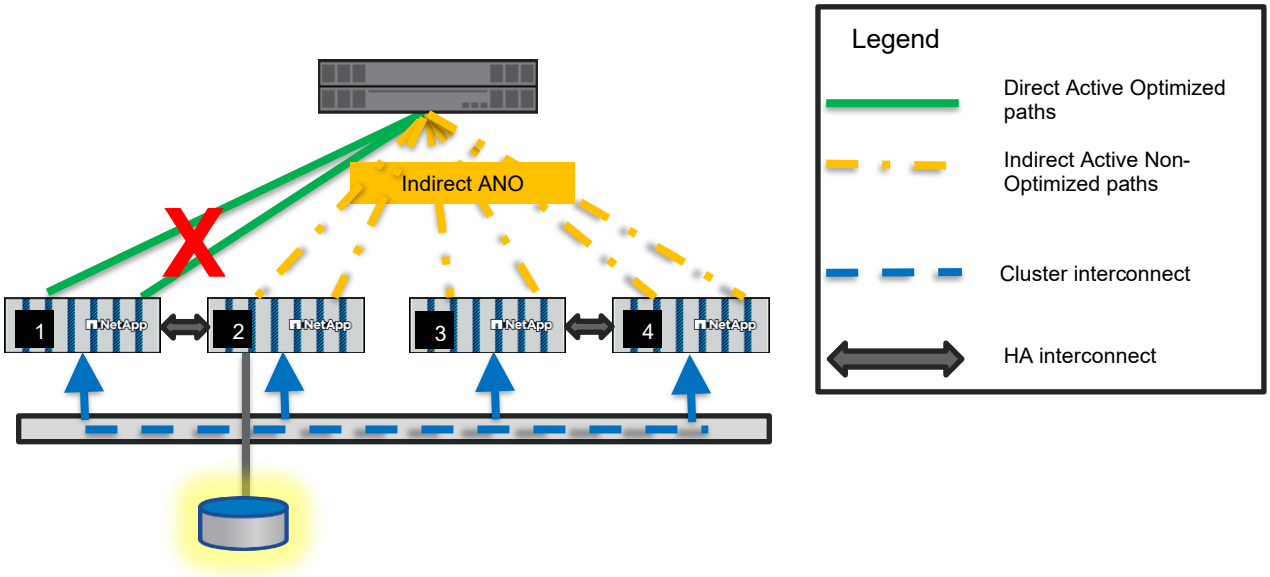
Figure 6) Paths during HA failover.



ASA supports both iSCSI LIF failover and FC persistent ports, enabling fast path redirection. While each protocol handles this differently, both move the target endpoint to a mirror port on the HA partner by virtually changing the IP address or WWPN. This ensures IO is quickly rerouted with no impact on host initiator-target-LUN mappings and therefore connected applications.

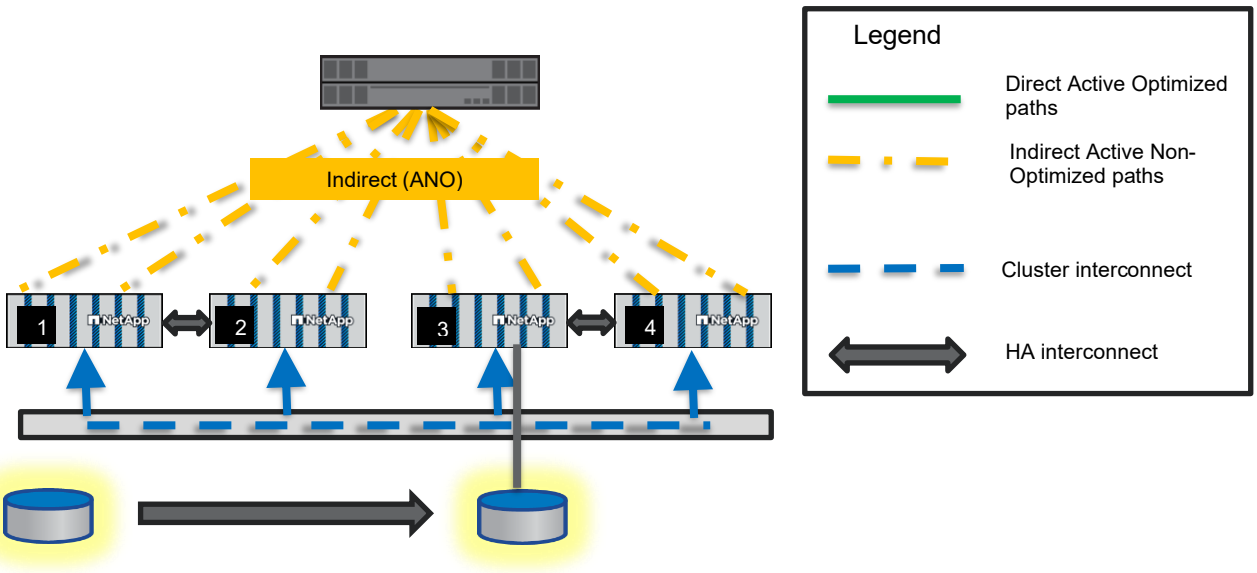
Port or switch failure. When a port or switch fails, direct paths are unavailable. The path priority is unchanged, and MPIO software chooses alternate indirect paths until a direct path recovers. ALUA path states stay the same. Figure 6 illustrates this scenario.

Figure 7) Paths during port or switch failure.



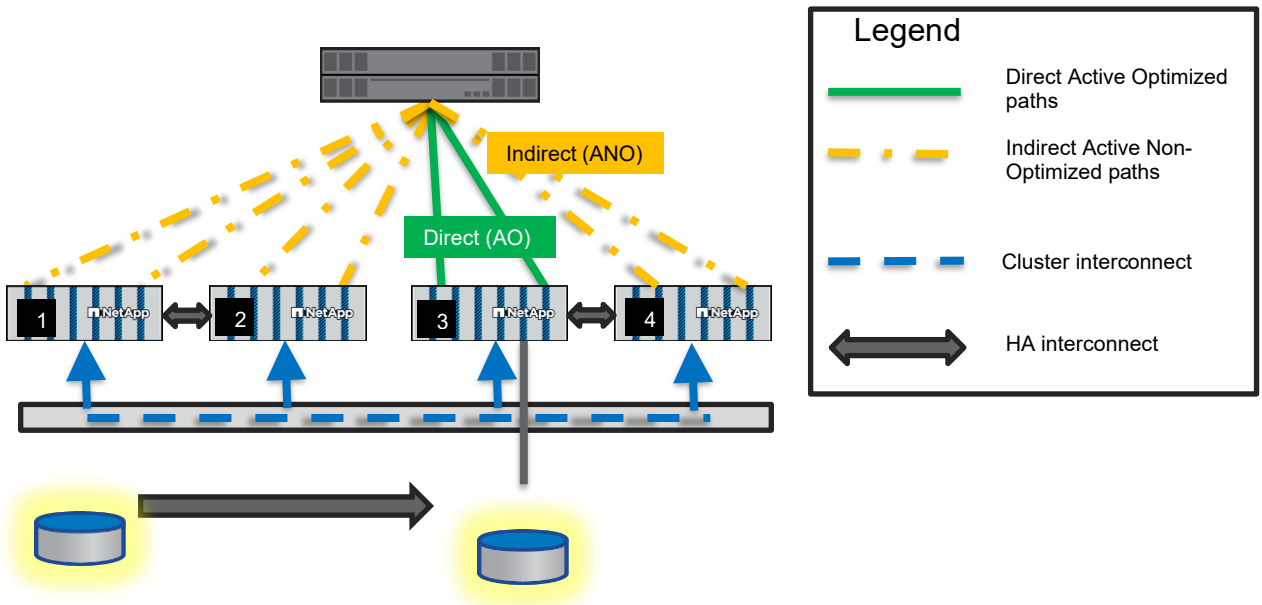
Volume or LUN mobility. Volumes or LUNs can be moved between nodes transparently using volume move or LUN move functions. Figure 7 illustrates the paths during their mobility.

Figure 8) Paths during volume or LUN mobility.



During a volume move, when the cutover occurs, the volume's new node begins to process read and write requests. The path status is updated to reflect that the new node has direct paths while the old node retains indirect paths. All paths remain accessible throughout the process. Figure 8 illustrates the paths following volume or LUN mobility. During a LUN move, cutover happens before all data is transferred; read requests are routed through the cluster network to the source node. For more on LUN move behavior, refer to “NetApp DataMotion for LUNs.”

Figure 9) Paths after volume or LUN mobility



FC and NPIV

ONTAP nodes use N_Port ID Virtualization (NPIV) so each logical interface logs in to an FC fabric with its own WWPN, instead of relying on a single WWNN and adapter-based WWPNs. This allows hosts on

the same FC fabric to reach the same SCSI target regardless of which node owns a LIF. The virtual port handles SCSI target services and data transfer. On Cisco NX-OS, run show npiv status to check NPIV status.

Best practice

NPIV is required for FC LIFs to operate correctly. Before creating FC LIFs, make sure that any fabrics attached to an ONTAP system have NPIV enabled.

```
N5K-A# show npiv status
NPIV is enabled
```

When using Brocade FabOS, the `portcfgshow` command shows NPIV capability and status.

```
BRCD_8K:admin> portcfgshow
Ports of Slot 0    0   1   2   3     4   5   6   7     8   9  10  11    12  13  14  15
----------------+---+---+---+---+-----+---+---+---+-----+---+---+---+-----+---+---+---
Speed             AN  AN  AN  AN    AN  AN  AN  AN    10  10  10  10    10  10  10  10
Fill Word          0   0   0   0     0   0   0   0     -   -   -   -     -   -   -   -
AL_PA Offset 13   ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Trunk Port        ON  ON  ON  ON    ON  ON  ON  ON     -   -   -   -     -   -   -   -
Long Distance     ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
VC Link Init      ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Locked L_Port     ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Locked G_Port     ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Disabled E_Port   ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Locked E_Port     ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
ISL R_RDY Mode    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
RSCN Suppressed   ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Persistent Disable..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
LOS TOV enable    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
NPIV capability   ON  ON  ON  ON    ON  ON  ON  ON    ON  ON  ON  ON    ON  ON  ON  ON
NPIV PP Limit    126 126 126 126   126 126 126 126   126 126 126 126   126 126 126 126
QOS E_Port        AE  AE  AE  AE    AE  AE  AE  AE    ..  ..  ..  ..    ..  ..  ..  ..
EX Port           ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Mirror Port       ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Rate Limit        ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Fport Buffers     ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
Port Auto Disable ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..    ..  ..  ..  ..
```

You can't check NPIV status on a switch from the storage admin console, but reviewing local FC topology reveals if switch ports have NPIV enabled. For example, NPIV is enabled when port 2/1 shows multiple attached WWPNs, including virtual ports.

```
cluster::> node run -node node01 fcp topology show
   Switch Name: N5K-A
 Switch Vendor: Cisco Systems, Inc.
Switch Release: 5.0(2)N2(1a)
 Switch Domain: 200
    Switch WWN: 20:66:00:0d:ec:b4:94:01

Port    Port WWPN                State    Type     Attached WWPN           Port ID
-------------------------------------------------------------------------------
2/1     20:01:00:0d:ec:b4:94:3f  Online   F-Port   50:0a:09:83:8d:4d:bf:f1 0xc80033
                                                   20:1c:00:a0:98:16:54:8c 0xc80052*
                                                   20:0e:00:a0:98:16:54:8c 0xc80034*
                                                   20:10:00:a0:98:16:54:8c 0xc8003f
2/3     20:03:00:0d:ec:b4:94:3f  Online   F-Port   50:0a:09:83:8d:3d:c0:1c 0xc8002c
```

Best practices

Physical WWPNs (beginning with 50:0a:09:8x) do not present a SCSI target service and should not be included in any zone configurations on the FC fabric, though they show as logged in to the fabric. These WWPNs are listed by using the `fcp adapter show -fields fc-wwpn` command or using the FC/FCoE Adapters pane under Network > FC Ports in System Manager, as shown in Figure 10) FC adapters in System Manager

.

Instead, use only virtual WWPNs (WWPNs starting with 20:) visible in the output of the `network interface show` command and in the System Manager Network > Overview pane, as shown in **Error! Reference source not found.**.

Figure 10) FC adapters in System Manager

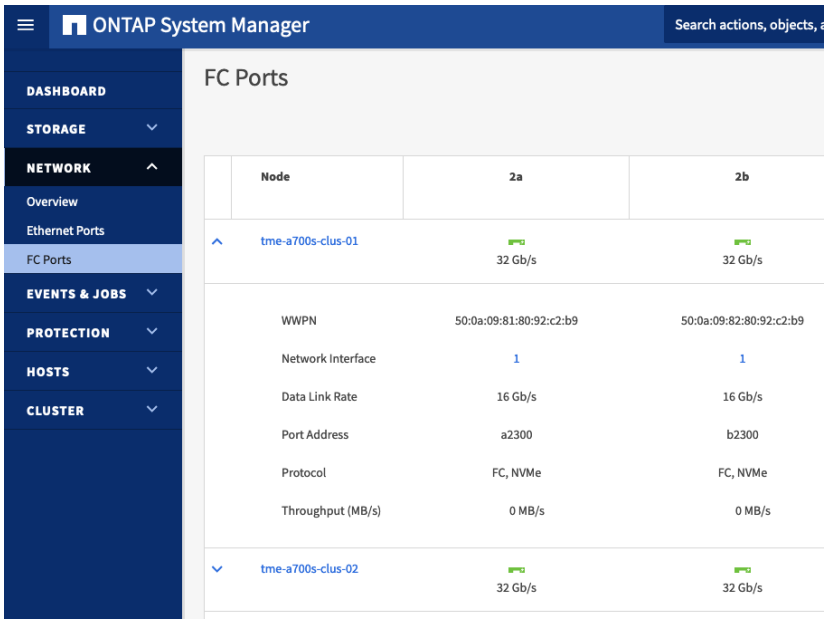
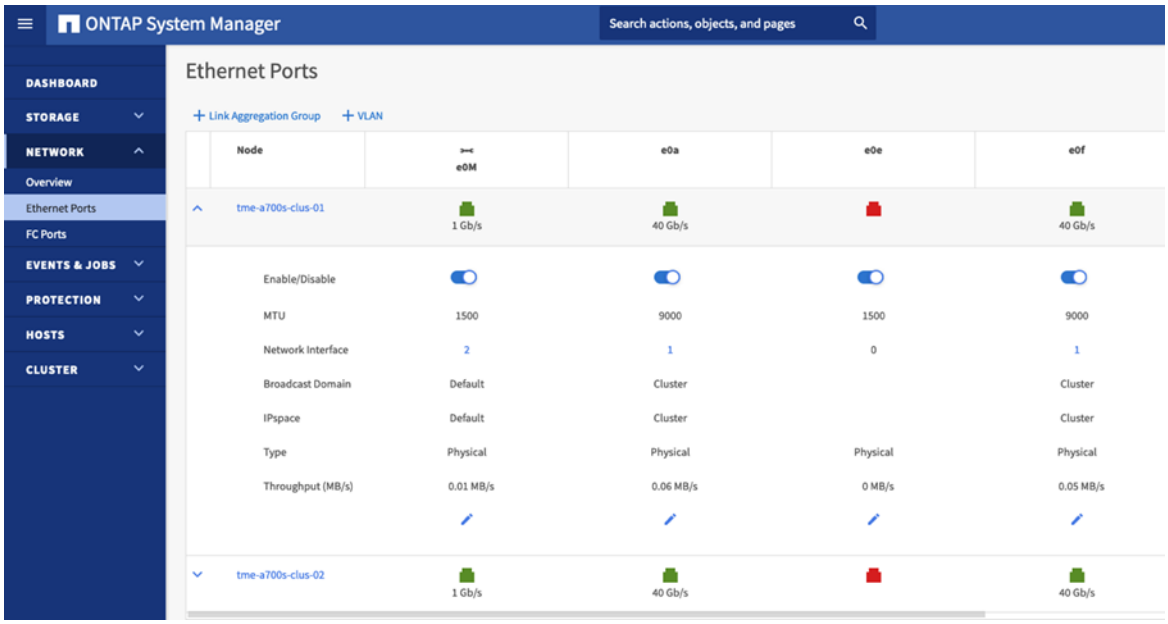


Figure 11) Network interfaces in System Manager



Path management and selective LUN mapping

Clusters configured with more than two nodes are increasingly likely to have a greater number of available paths compared to traditional configurations. Clusters connected to multiple fabrics, or with storage controllers linked more than once per fabric, can rapidly increase the potential path count.

This scenario presents several challenges for storage administrators:

- While a high number of target ports enhances redundancy, it also introduces operational complexity. In Fibre Channel (FC) environments, this necessitates larger and more intricate zones and zonesets, as well as maintaining extensive tables of World Wide Port Names (WWPNs) for cluster SVMs. In iSCSI deployments, there may be numerous sessions required for each host that accesses a LUN.
- Many operating systems impose practical limits on the number of accessible paths. For hosts with many paths and LUNs, exceeding these limits can result in LUN enumeration issues or path status complications.
- Even in the absence of path status problems, managing an excessive quantity of unused paths leads to inefficient utilization of host resources—including processor capacity and memory allocation.
- Certain high-throughput workloads benefit from traffic separation to reduce contention; however, Asymmetric Logical Unit Access (ALUA) path statuses do not provide mechanisms for prioritizing one direct path over another.

The ONTAP storage operating system enforces an upper tested limit on the total number of established paths—referred to as initiator-target nexus (ITN). For additional information regarding these limits for any NetApp storage controller, consult the Hardware Universe site.

It is advisable to limit the total number of presented paths. Nevertheless, to maintain both direct data access and redundancy for high-availability failovers or path failures, at minimum, both the node containing the relevant volume and its HA partner should present paths.

There are two primary methods for limiting LUN-presented paths through storage OS features (rather than solely via FC zoning or iSCSI session management): selective LUN mapping (SLM), which is enabled by default in ONTAP 8.3 and later, and port sets.

Selective LUN mapping

SLM expands the existing LUN mapping table in a Data ONTAP cluster, detailing the relationships between LUN paths, igroups, and LUN IDs. Since LUNs can map to multiple igroups and vice versa, this table provides comprehensive mapping details. In ONTAP 8.3 and later, each mapping also lists reporting nodes, indicating which controllers present that LUN to each igroup. By default, new LUN mappings use selective LUN mapping, showing the LUN only from its home node and HA partner, while older mappings use a wildcard to present the LUN from all nodes. Administrators can specify any combination of nodes—grouped in HA pairs—or use blank/wildcard entries for cluster-wide presentation, enabling granular control over path visibility. Below is the lun mapping show command displaying the field “reporting nodes”. This query shows the nodes that are advertising paths to a given LUN.

```
san-cluster::> lun mapping show -fields reporting-nodes
vserver          path                igroup lun-id reporting-nodes
---------------  ------------------- ------ ------ ----------------
SAN_Default_SVM  /vol/host1/lun0     linux1      0 node-01,node-02
SAN_Default_SVM  /vol/host2/lun0     linux2      0 node-01,node-02
SAN_Default_SVM  /vol/host2/lun1     linux2      1 node-03,node-04
```

Port sets

Port sets let administrators restrict access to LUNs on specific target ports, rather than exposing them to all available ports.

A LIF in a port set can't be modified until removed, but it can be re-added after changes, ensuring enough LIFs remain to meet host path requirements.

To maintain direct access and redundancy during failover or non-disruptive events, paths to the node with the data volume and its HA partner are sufficient.

Management interfaces

LIFs connected to SVMs that provide data using block protocols cannot simultaneously be used for administrative tasks. Since the SVM serves as the main management unit within an ONTAP storage cluster, each SVM requires a distinct management interface in addition to its data interfaces for block protocols.

During SVM creation, the vsadmin account becomes available but needs a password to be set using the security login password command, followed by unlocking it with the security login unlock command. For more details, refer to Delegating Administration to SVM Administrators in the ONTAP System Administration Guide.

When managing a cluster through System Manager, an SVM management LIF may be established during SVM creation or assigned as a management LIF during standard LIF creation. Refer to Figure 11, Figure 12, and Figure 13.

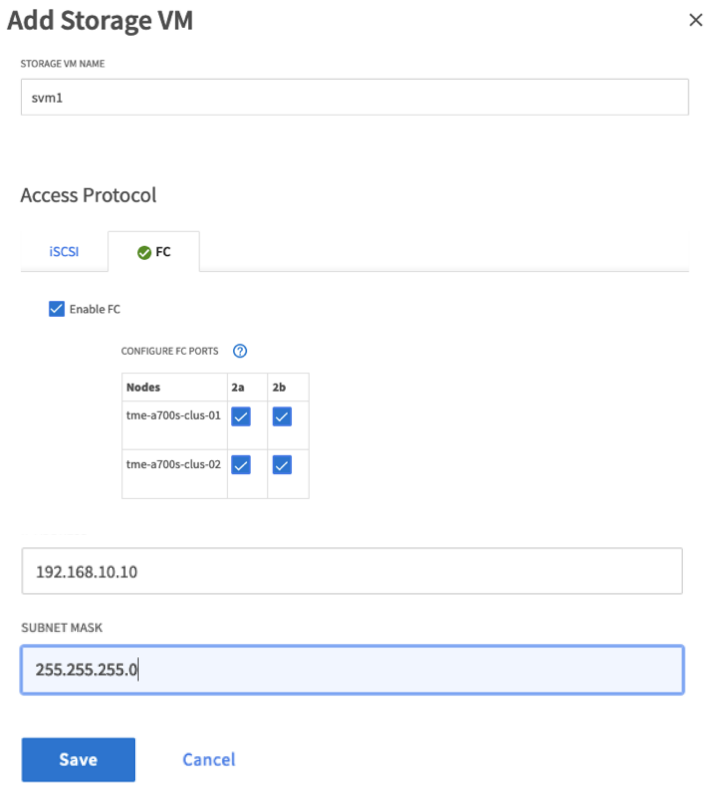
Best practices

A management interface on an SVM serving block data should have the following properties:

- A LIF type of `data`
- No data protocols assigned (`-data-protocols none`)
- A firewall policy that permits management access (`-firewall-policy mgmt`)
- A failover group and policy that keep the LIF accessible to hosts that might need to contact it for data management purposes, such as creating or managing Snapshot® copies (For more

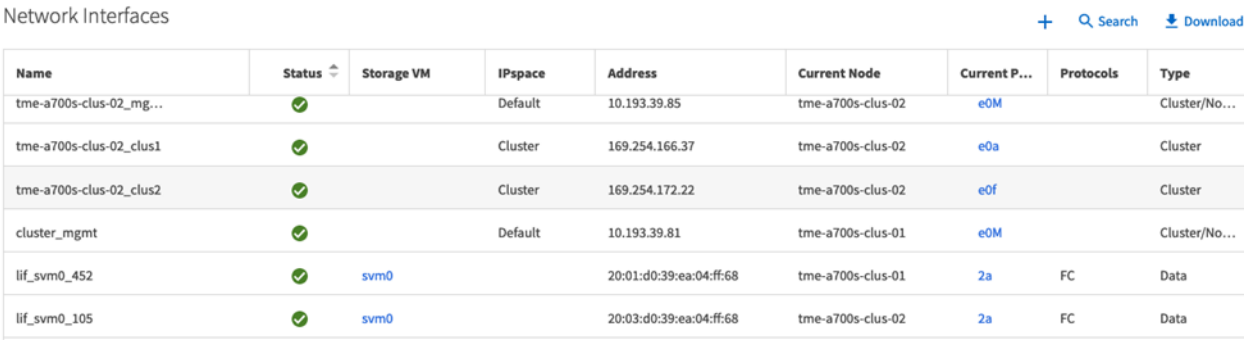
information about failover groups, see “Configuring Failover Groups and Policies for LIFs” in the ONTAP Network Management Guide.)

Figure 12) Creating a management LIF during SVM creation



The management LIF details should look like the details in **Error! Reference source not found.**13.

Figure 13) Management LIF details



NetApp DataMotion for LUNs

You can move or copy LUNs individually between volumes, aggregates, controllers, and HA pairs using commands or API calls. LUNs become available almost immediately after creation, with metadata and attributes in place. Data is copied in the background; read requests for missing data fetch it from the source, while writes go straight to the destination.

LUN move and LUN copy comparison

NetApp DataMotion for LUNs allows LUNs to be copied between volumes in the same or different SVMs, while LUN moves are limited to within the same SVM due to differing configurations across SVMs. LUN copies can use snapshots as their source, but moves require the source to be in the active file system; data cannot be moved from snapshot copies, as they are immutable. By default, LUN moves are promoted early (allowing I/O without new snapshots), and LUN copies are promoted late (permitting both I/O and snapshots).

Storage efficiency considerations

LUNs moved or copied via NetApp DataMotion for LUNs arrive at their destination without compression or deduplication. Deduplication or cloning can only share data within a volume; copied LUN data is simply a duplicate from its source. Any data in a locked Snapshot copy remains on disk until the Snapshot expires or is deleted, even if the LUN has been moved. Volumes with inline compression enabled do not compress LUNs received through DataMotion operations.

Best practice

If a destination volume has not previously contained deduplicated or compressed data, turning on deduplication or compression adds the arriving LUN's blocks to the list of blocks to be processed during the next storage efficiency run, and they do not need to be discovered through a block scan.

Data protection considerations

Data protection considerations primarily apply to LUNs that are moved, as copied LUNs preserve the original source data. When a LUN is moved, any Snapshot copies associated with the original volume are not transferred. To include data within Snapshots, employ LUN copy to duplicate the Snapshot LUNs. Deduplication enables the new LUN to reference existing duplicate blocks efficiently. Upon moving a LUN, previous data protection relationships may be discontinued, necessitating the establishment of a new SnapMirror relationship if required. If the destination is already engaged in such a relationship, additional steps may be necessary to address increased space consumption resulting from data replication.

Best practice

When using LUN move in conjunction with software external to the ONTAP storage cluster to manage Snapshot copies containing LUNs, make sure that the software is aware of the capabilities of NetApp DataMotion for LUNs and can (for example) restore a LUN from Snapshot copies in volumes in which it might no longer exist. If this is not possible, LUN move might have an effect on data protection workflows.

Scalability and throughput considerations

You can throttle LUN move or copy operations individually using the -max-throughput argument, either at the start or later via the lun copy modify or lun move modify commands. Up to 50 move or copy operations can run simultaneously; additional ones are queued. This limit applies to destination operations.

Best practice

A LUN copy or move operation can be paused and resumed at any time after data begins copying in the background. Pausing the move or copy only prevents data from being moved in the background, but does not prevent requests for data that hasn't yet arrived from being forwarded to the source LUN for fulfillment.

Data management and workflow considerations

Consider these ONTAP interactions when using NetApp DataMotion for LUNs:

- LUNs that are being moved or copied cannot be removed during the operation.
- SnapRestore® cannot replace LUNs involved in move or copy actions until those operations finish.
- If a LUN is in a volume being moved, any ongoing LUN move or copy pauses during the volume's cutover.

Best practices

Some existing workflows can take advantage of NetApp DataMotion for LUNs to shorten the number of required steps:

- Previously, to duplicate a volume containing LUNs, the entire volume needed to be cloned. Now any empty or already-occupied volume can be filled with LUN copies from another volume's Snapshot copies, even if that volume is in a separate SVM. Effectively, the subvolume LUN cloning capability previously available within a volume can now be extended to other volumes.
- Previously, to change the existing layout and ratio of LUNs and volumes, it was necessary to clone volumes and then remove unnecessary LUNs or to use a host-side copy using volume management to fill a new LUN with an old LUN's data. Now, if storage efficiency can be better served by consolidating LUNs in fewer volumes, or if a single LUN in a volume containing others needs to relocate to satisfy performance or storage tiering needs, LUNs can be moved nondisruptively between volumes on the fly.

NetApp DataMotion and selective LUN mapping: Discovering and discarding paths

When altering the LUN mapping on the storage cluster to create new paths or remove existing ones, the hosts attached to that LUN must perform a SCSI bus rescan. Therefore, when moving LUNs between HA pairs, the procedure should be as follows:

1. Change the LUN mapping to add the new reporting nodes using the `lun mapping add-reporting-nodes` command.
2. Perform a SCSI bus rescan on the hosts accessing the LUN, discovering the new paths.
3. Move the LUN nondisruptively, ALUA signals a path status change to the host, and the host begins moving I/O down the new direct paths.
4. Change the LUN mapping to remove the old reporting nodes using the `lun mapping remove-reporting-nodes` command.
5. Perform a SCSI bus rescan on the hosts accessing the LUN, discarding the old paths.

More than one LUN can have new paths discovered or old ones removed during a rescan.

For step-by-step instructions on how to perform a host SCSI bus rescan for all supporting operating systems, see the KB article on the NetApp Support site describing the procedure.

Caution

Do NOT remove reporting nodes until the LUN move is complete and any host remediation steps, for example, SCSI bus rescans, are completed. If reporting nodes are removed prior to adding new reporting nodes, completing the LUN move, and all host remediation steps are completed, you could lose access to the LUN that was moved.

Path management best practices

You should use ONTAP features to limit the number of available paths at the storage management level.

Best practices

- For storage controllers that have a single target LIF on each connected FC fabric or Ethernet network, the default number of paths presented by a LUN mapping is two direct paths from the storage controller that contains the volume and LUN being accessed and two indirect paths from its HA partner, for a total of four paths.
- Selective LUN mapping by default limits a LUN's paths to the storage controller that owns it and its HA partner, but extra nodes might be part of a mapping on either a temporary or permanent basis.
- In clusters that have more than one target LIF per connected FC fabric or Ethernet network, you can use the extra paths to provide more bandwidth or queue depth on a per-LUN basis, or port sets can be used to channel traffic on a per-igroup basis to specific LIFs.
- For LUNs that require more paths than a default LUN mapping provides, eight paths are almost always sufficient and is a path count supported by all host SAN implementations. For LUNs that require even more paths, the SAN Configuration Guide lists the tested maximum number of paths for each supported host OS.
- LUN mobility events such as `vol move` or `lun move` that involve moving a LUN from one HA pair in the cluster to another should include a step to confirm that the LUN is being presented using the destination storage controllers before the mobility event is initiated. The `lun mapping add-reporting-nodes` command can be used to add the new paths. After the move is complete, use the `lun mapping remove-reporting-nodes` command to remove the original, no longer direct path.
- Changing the paths presented for a LUN also means that a host SCSI bus rescan should be performed in order to discover new paths and discard stale ones. For best practices from a host perspective on path changes and for the procedure to be used when a LUN mapping must change to accommodate its moving to an HA pair that currently does not present paths, see the section titled, "NetApp NetApp DataMotion and selective LUN mapping: Discovering and discarding paths."
- Because a change on the host accessing the LUN is necessary for a LUN mapping change, consider expanding the list of nodes in LUN mapping situations where administrative steps taken on the host are undesirable or when LUN mobility between HA pairs is frequent.

Storage unit settings

Recommended volume and file or SU configuration combinations overview

Table 3) Settings for thin provisioned Storage units

Guarantee	None
Snapshot reserve	0
Snapshot autodelete	Any
Autogrow	Any
File or SU setting	

Space reservation	Disabled

NetApp All-SAN Array

NetApp All-SAN Array (ASA) systems, based on NetApp AFF platforms running ONTAP, offer a robust, enterprise-grade SAN solution designed for customers seeking consolidation and sharing of storage resources across multiple workloads.

AFF SAN systems provide:

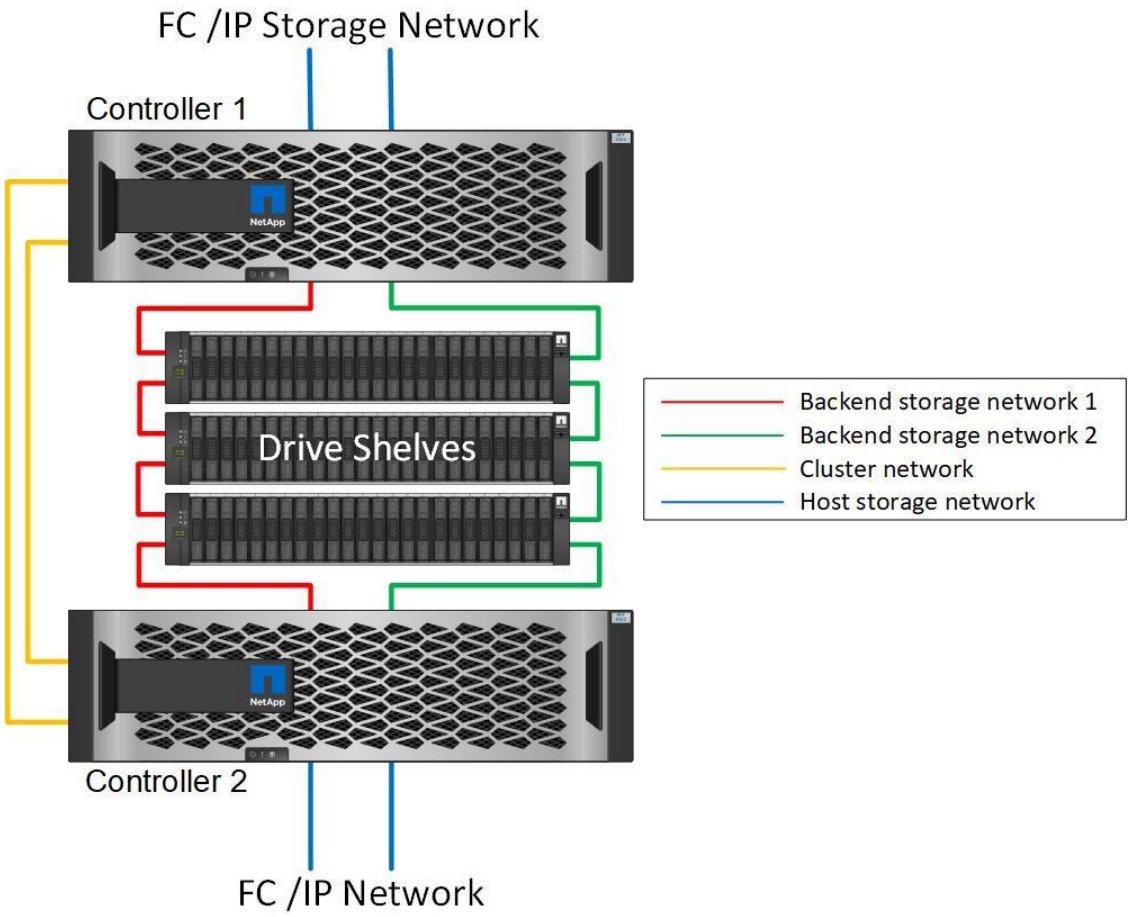
- Industry-leading availability exceeding 99.9999%
- Scalable clusters that expand both vertically and horizontally
- Exceptional enterprise performance verified by audited SPC-1 results
- Superior storage efficiency
- Comprehensive cloud-enabled connectivity options
- Cost-effective and seamless data protection

NetApp ASA systems enhance the AFF platform to ensure continuous SAN availability, delivering uninterrupted data access during both planned and unplanned storage failovers. These systems facilitate streamlined implementation, configuration, and management through a solution dedicated exclusively to SAN workloads. NetApp recommends ASA configurations for customers with requirements such as:

- Support for mission-critical workloads, including databases that necessitate symmetric active-active paths between hosts and storage
- A preference for dedicated systems that isolate SAN workloads

AFF systems remain the optimal choice for customers who:

- Require scaling out SAN clusters up to 12 nodes
- Do not need active-active SAN path management
- Prefer cluster environments supporting unified protocols for mixed NAS and SAN workloads



Provisioning Storage units (SUs)

What are thick and thin provisioning SUs

Thick, thin, and semi-thick provisioning define how storage units (SUs) reserve space. Thick provisioning reserves the SU's full size upfront; thin provisioning only uses space as needed; semi-thick is when a thick-provisioned SU resides in a volume that isn't fully reserved, which can lead to space shortages.

By default, ONTAP creates SUs and namespaces with thin provisioning. To use thick provisioning, you must enable it during SU creation. Namespaces are always thin provisioned unless unmap/space allocation is enabled.

Note: The ASA provisions thick Storage Units (SUs) within abstracted volumes that are thin provisioned and are provisioned to their a maximum size of 600TB. As a result, these volumes only utilize the space required by their actual constituents, rather than occupying the total maximum volume capacity.

Review the SAN administration guide for a full discussion of ONTAP's space reservation options, the settings required for each option, and the pros and cons of each approach. Here are links to the settings needed for thick, thin and semi-thick provisioning:

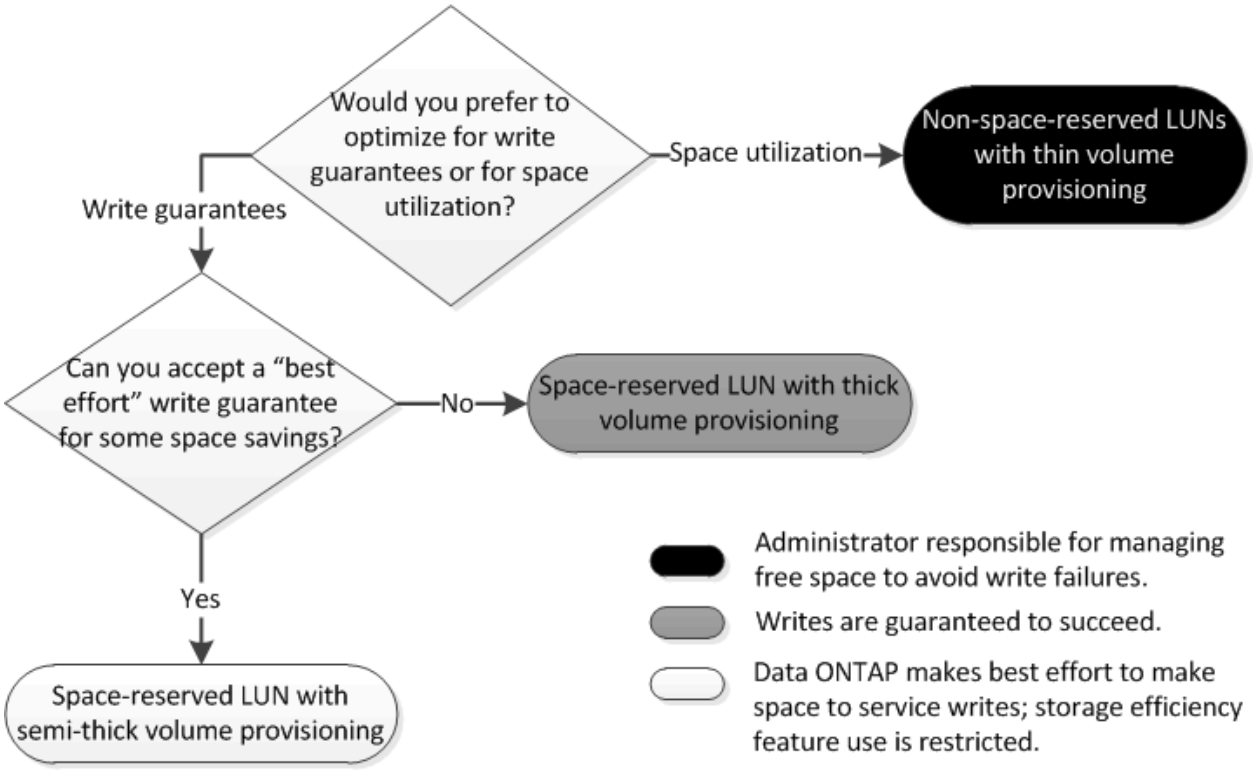
1. Thick provisioning a LUN
2. Thin provision an LUN
3. Semi-thick provision a LUN

Best practices for provisioning SUs

Figure 12, below, is a flowchart that can be used to determine which space reservation strategy you should use.

1. Review the SAN administration guide for details about your space configuration options.
2. Use the default Thin provisioning when setting up new SUs and namespaces assuming that you have the capability and commitment to actively and systemically monitor SU space usage so that you can react in a timely manner to low space conditions.
3. Systemically monitor space usage and setup timely alerting to provide enough time to react and remediate any low space conditions that might occur. Consider using NetApp Console and Cloud Insights (CI), and Harvest to monitor and alert when low space thresholds are reached
4. If you are unable or uninterested in provisioning thin provisioned SUs because you don't have sufficient systemic monitoring or don't want the higher monitoring requirements of thin provisioning, you should thick provision your SUs and namespaces

Figure 14) LUN provisioning flowchart



ONTAP SAN key value propositions and features

This section outlines several of NetApp's primary design objectives. These objectives include creating a unified architecture at scale that supports nondisruptive operations for data mobility, performance optimization, capacity planning, and system-level hardware replacement. While not a comprehensive list of available features, this demonstrates how scalable SAN capabilities and ONTAP differ from other solutions in the storage industry.

SVM as unified target and unit of management

Storage controllers with ONTAP operating as part of an HA configuration or larger cluster, present a single WWNN per SVM to the connected FC fabric. The storage cluster replicates this single WWNN on an SVM basis across all cluster members, enabling each node to display the same target and allowing multiple targets to exist on the same hardware.

This approach also applies to storage management. All data is provided from volumes linked to an SVM and via iSCSI or FC targets configured within an SVM. As a result, cluster administration is performed per SVM instead of managing each storage node individually.

Managing at the SVM level supports the implementation of a multitenancy model for storage management.

Scalability at the node and cluster levels

ONTAP provides scalable options at the node and cluster levels, with enhanced scalability since Data ONTAP 8.1 introduced SAN protocols. For complete SAN configuration limits, consult the NetApp Hardware Universe (HWU). Table 2 offers a summary.

Table 4) Scalability in ONTAP

Version of ONTAP	9	9.1	9.8
Nodes per cluster	8	12	12
LUNs per node	12,288	8,192	8,192
LUNs per cluster	98,304	98,304	98,304
iSCSI sessions/node	8,192	8,192	8,192
FC I_T_Ns/node	8,192	8,192	8,192

Cluster-wide consistency groups

Snapshot consistency groups, introduced in ONTAP 9.10.1, enable simultaneous Snapshot copies across multiple storage controllers. This allows hosts to synchronize Snapshot creation on volumes from different cluster nodes within an SVM, ensuring consistent copies. With a single command, a host can take a Snapshot across several nodes and volumes at once. Consistency groups operate per SVM, so any volume owned by the target SVM can be included.

Intracluster SU and LIF mobility

Earlier ONTAP versions allowed nondisruptive volume migration within a cluster. Starting with ONTAP 8.3, administrators can now copy and migrate individual Storage Units (SUs) between volumes and controllers, enabling quicker cloning as SUs are instantly accessible. Normally, moving Logical Interfaces (LIFs) or volumes between nodes is unnecessary, but some cases may require nondisruptive migration. The destination node must have a direct host path as detailed in "Path selection." LIF migration overhead is minimized by modifying existing LIFs, keeping IP addresses or WWPNs unchanged, so fabric zoning or host updates aren't needed. SAN LIFs must be offline to modify them, using `network interface modify –status-admin down`.

Best practice

Do not exceed the cluster size limit when making changes to cluster membership. For information about the cluster size limit when using block protocols, see the ONTAP 9 SAN Configuration Guide.

Foreign LUN Import (FLI)

Starting with Data ONTAP 8.3, you can import LUNs from third-party arrays and E-Series/EF-Series controllers using FC, without extra licenses or equipment—just set some storage controller FC or UTA2 ports to initiator mode (UTA2 ports must use FCP personalities). ONTAP creates an identical LUN in its own volume and copies data block-by-block, allowing protocol-agnostic use (imported LUNs via FC may be presented to hosts as iSCSI).

The import can be done online or offline. Online imports (from 8.3.1 onward) briefly take the LUN offline to establish the relationship; I/O then resumes as ONTAP mirrors data until import completion. Offline imports keep both LUNs inaccessible until finished. Supported third-party arrays are listed in the FLI Interoperability Matrix Tool (IMT). Additionally there is a SAN LUN Migrate tool in the NetApp support site tools menu that allow you to field qualify a source array to work with FLI. For more information please review appendix

For more details about the SAN LUN migrate tool please review Appendix B: Field Qualifying Foreign LUN Import with a new source array. For more information about FLI, refer to the Foreign LUN Import overview.

Using FabricPools with ONTAP SAN

If a FabricPool (FP) S3 tiering connection fails, ONTAP issues retriable errors for up to 120 seconds, then non-retriable errors affecting only cold tiered data; hot local storage stays available. This can lead to SAN Host filesystem inconsistencies requiring recovery.

NetApp advises tiering only SAN volume snapshots with FabricPool to prevent data loss if lower-tier data becomes unavailable. FP behavior is uniform across ONTAP SAN protocols—FCP, iSCSI, NVMe/FC, and NVMe/TCP—and applies to both RW and RO volumes, as well as all DR and DP features.

Customers using *All* or *Auto* tiering policies should recognize risks during S3 tiering outages. Avoid the *All* policy on LUN-hosting volumes, and do not use *Auto* for critical LUNs such as SANboot or cluster quorum.

Host integration

NetApp Host Utilities Kit

Installing the Host Utilities Kit applies recommended OS-specific timeout settings and provides tools to examine LUNs from NetApp storage, including both clustered and 7-Mode systems. For complete details on supported configurations, refer to the NetApp Interoperability Matrix Tool.

UNIX or Linux Host Utilities Kit

The NetApp Host Utilities Kit offers tools for viewing SU configuration at the SVM level, giving detailed information about an attached SU's volume and path within its SVM.

```
# sanlun lun show all
controller(7mode)/              device       host                  lun
vserver(Cmode) lun-pathname     filename     adapter    protocol   size    mode
--------------------------------------------------------------------------------
vs             /vol/vol1/linux1 /dev/sdcx    host1      FCP        25g     C
vs             /vol/vol2/linux2 /dev/sdcw    host1      FCP        25g     C
vs             /vol/vol3/linux3 /dev/sdck    host1      FCP        25g     C
```

In addition, the Host Utilities Kit can indicate which of an SVM's logical interfaces are serving as direct and indirect paths for a specific SU, with primary designating direct paths and secondary indicating indirect paths.

```
# sanlun lun show -p
                    ONTAP Path: vs:/vol/vol1/linux1
                           LUN: 0
                      LUN Size: 25g
                          Mode: C
                   Host Device: 3600a09803246664c422b2d51674f7470
              Multipath Policy: round-robin 0
            Multipath Provider: Native
--------- ---------- ------- ------------ ----------------------------------------
host      vserver
path      path       /dev/   host         vserver
state     type       node    adapter      LIF
--------- ---------- ------- ------------ ----------------------------------------
up        primary    sdfo    host0        fcoe_lif_1
up        primary    sdfk    host1        fcoe_lif_2
up        secondary  sdga    host0        fcoe_lif_3
up        secondary  sdge    host1        fcoe_lif_4
up        secondary  sdgm    host1        fcoe_lif_5
up        secondary  sdgj    host0        fcoe_lif_6
up        secondary  sdfw    host0        fcoe_lif_7
up        secondary  sdgq    host1        fcoe_lif_8
```

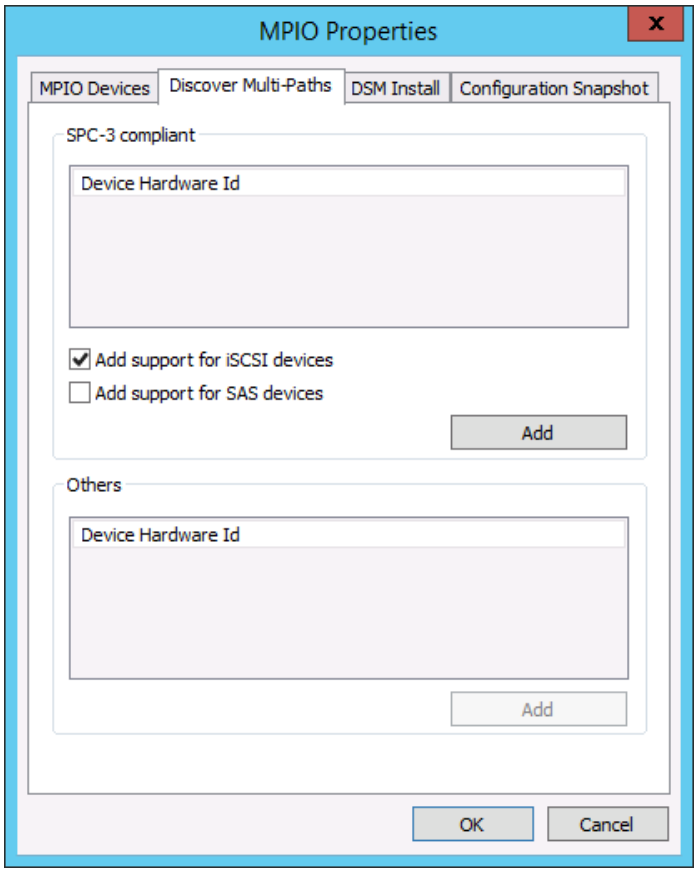
Microsoft Windows and native MPIO

Accessing ONTAP storage clusters requires hosts to use MPIO and ALUA. For Microsoft Windows 2008 and 2012, these features are supported when multipath I/O is installed. To enable multipathing for iSCSI devices, open the MPIO properties application, go to the Discover Multi-Paths tab, select Add Support for iSCSI Devices, and click Add.

Windows Host Utilities Kit

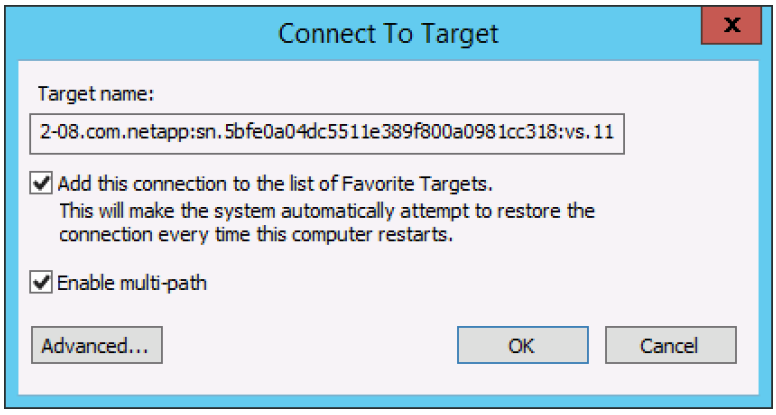
The NetApp Host Utilities Kit can be installed to adjust Windows registry settings for performance and failover. If Data ONTAP DSM is already present, the Host Utilities Kit does not modify the registry and relies on Data ONTAP DSM to set correct values.

Figure 15) MPIO properties in Windows 2012.



It's also necessary to create multiple sessions from the host initiators to the target iSCSI LIFs on the storage cluster. This can be accomplished using the native iSCSI initiator. Select the Enable Multi-path option (as shown in Figure 15) when logging on to a target.

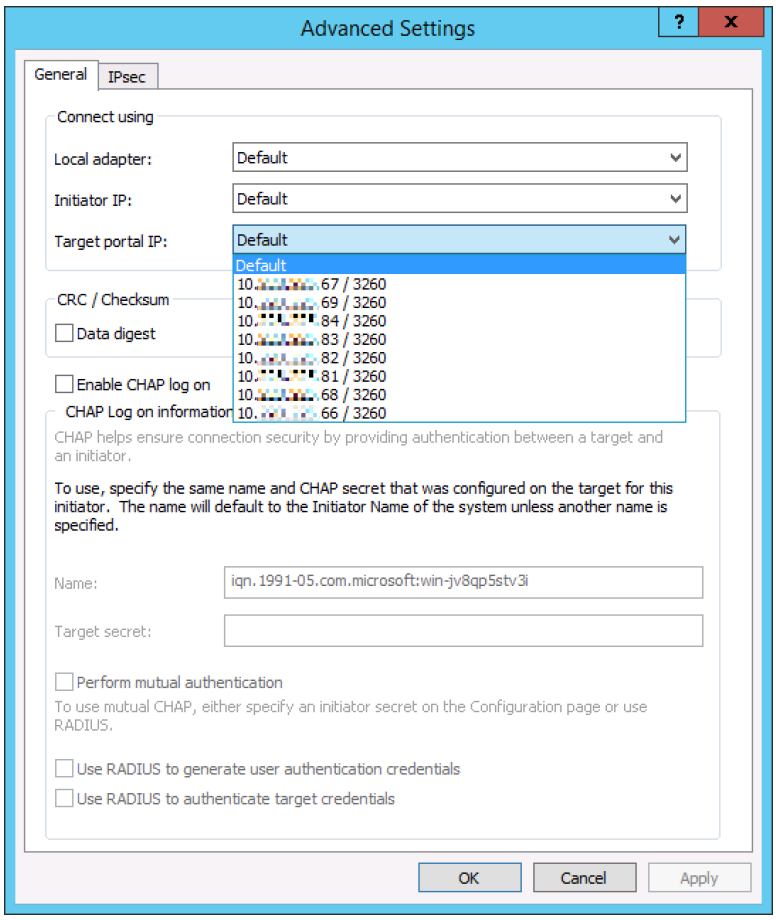
Figure 16) Connecting with multipath in Windows iSCSI initiator.



To manually create additional sessions, highlight the corresponding target in the Targets pane of the iSCSI initiator and click Log on. Make sure that the session is automatically restored after the next reboot and that the new session is identified as a multipath session by selecting both checkboxes.

Click Advanced. From the Target Portal IP drop-down menu (as illustrated inFigure 16) select the IP address of the logical interface that is the target of the new iSCSI session.

Figure 17) Multiple target ports in Windows iSCSI initiator.



Sessions can also be managed by using the NetApp SnapDrive® iSCSI management pane. This is the preferred method, because SnapDrive remembers which target logical interfaces already have an established session and preselects an unused target portal.

SnapCenter

SnapCenter is a scalable platform for application-consistent data protection, offering centralized management while allowing users to handle backup, restore, and cloning tasks for specific applications. It enables both database and storage administrators to use one tool for these operations across various endpoints within NetApp's data fabric.

ONTAP tools for VMware vSphere

ONTAP tools for VMware vSphere provide integrated storage management within vSphere environments. The vCenter plug-in (formerly Virtual Storage Console) streamlines storage tasks, improves efficiency, supports both SAN and NAS, and optimizes ESXi host settings for NFS and block storage. NetApp recommends using these tools as a best practice with ONTAP systems. Features include a server appliance, vCenter UI extensions, VASA Provider, and Storage Replication Adapter, with extensive automation available via REST APIs.

IBM AIX and ONTAP

IBM AIX systems can access data on a cluster via Fibre Channel (FC) starting from ONTAP 8.2 storage clusters. For detailed information on supported AIX technology levels and service packs, please refer to the NetApp Interoperability Matrix.

Cross-platform utilities

RBAC User Creator

When setting up a Data ONTAP storage cluster at the SVM level, the RBAC User Creator tool in NetApp Utility Toolchest enables precise assignment of roles to users for specific tasks. It also simplifies providing access for external orchestration, monitoring, or backup and recovery tools, such as SnapDrive for Windows and Virtual Storage Console for VMware vSphere.

Data protection

While the previous section covers data availability and integrity for storage hardware, it is also essential to plan for recovery from user or application errors. Enterprises seeking 99.999% uptime should implement backup and recovery strategies that enable quick and reliable restoration of growing datasets.

Data protection with NetApp Snapshot copies

NetApp ONTAP data protection relies on Snapshot technology, offering:

Simplicity. Snapshot copies capture data containers at a specific time as read-only versions.

Efficiency. No additional space is used when snapshots are created; only changes consume space.

Manageability. Snapshots are built into the storage OS, making backup strategies straightforward and easy to manage.

Scalability. Up to 1023 snapshots per SU can be stored, and multiple data containers can be protected consistently.

Snapshot performance remains consistent regardless of volume count. Backups require no data movement, allowing flexible strategies based on business needs instead of network or hardware limitations.

Data restoration with ONTAP SnapRestore

NetApp SnapRestore® technology enables rapid recovery of data in ONTAP from Snapshot copies. Files or LUNs, whether 4KB or 128TB, and entire FlexVol® volumes up to 600TB can be restored within seconds. Unlike tape or slow disk-based backups, SnapRestore provides near-instant restoration—even for large databases—minimizing downtime and streamlining critical operations.

SnapMirror

SnapMirror is a scalable, efficient replication technology that manages both data and snapshot backups. You can store backups remotely or in the cloud, ensuring availability while minimizing network and storage requirements.

SnapMirror Replication Offerings

Asynchronous Replication Offerings

Asynchronous replication is particularly well-suited for disaster recovery and data archiving over long distances. Unlike synchronous replication, it does not require continuous synchronization between the source and destination systems. Instead, data is replicated at scheduled intervals, allowing for efficient use of bandwidth and resources.

SnapMirror DR (Data Protection)

SnapMirror DR serves as the core SnapMirror solution for asynchronous, volume-level disaster recovery. It creates a complete mirror of the data, which can be activated if a failure occurs at the primary site, ensuring business continuity and minimizing downtime.

SnapMirror Archive (formerly SnapVault)

SnapMirror Archive is designed for backup and long-term data retention needs. It allows organizations to create multiple point-in-time Snapshot copies on the destination volume. These copies are invaluable for restoring previous versions of data and meeting compliance or regulatory requirements.

Unified Replication

The Unified Replication policy brings together the features of SnapMirror DR and SnapMirror Archive within a single relationship. This approach provides both disaster recovery capabilities and a comprehensive backup history of Snapshot copies, supporting both failover and restore operations.

SnapMirror Cloud

Introduced with ONTAP 9.8, SnapMirror Cloud enables asynchronous replication from an ONTAP system to object storage endpoints. This feature supports long-term data archival to both private and public cloud object storage solutions, such as Amazon S3 or Azure Blob Storage, offering flexibility and scalability for cloud-based storage strategies.

Amazon FSx for ONTAP

NetApp SnapMirror can replicate data between FSx for ONTAP file systems, either within the same AWS region or across different regions. It also supports replication from on-premises ONTAP systems to FSx for ONTAP, facilitating seamless hybrid cloud data protection and disaster recovery solutions.

Synchronous Replication Offerings

Synchronous replication is designed for situations where zero data loss is acceptable. It ensures a zero Recovery Point Objective (RPO) by writing data to both the source and destination volumes at the same time, keeping them perfectly synchronized.

SnapMirror Synchronous (SM-S)

SnapMirror Synchronous offers volume-level replication for mission-critical workloads that cannot tolerate any data loss. The destination storage remains in a continuously synchronized state with the source, ensuring data integrity and immediate availability in the event of a failure.

SnapMirror Active Sync

NetApp's solution for SAN environments requiring the highest levels of availability. It provides automatic and transparent application failover, achieving both zero Recovery Time Objective (RTO) and zero RPO, so applications remain operational without any downtime or data loss.

MetroCluster technology

NetApp MetroCluster offers high availability and zero data loss for critical workloads. It simplifies complex enterprise applications and virtualization by consolidating backup, recovery, disaster recovery, and high availability into a single clustered storage system, eliminating the need for multiple external data protection tools.

HA with MetroCluster

MetroCluster uses NetApp SyncMirror® technology to efficiently switch between synchronous and nonreplicated modes, meeting customer needs for both synchronous replication and high availability. If a remote site loses connectivity, the system continues locally rather than stopping service, unlike other solutions limited to synchronous-only (domino mode), which halt data access if unsynchronized. SyncMirror® allows rapid resynchronization to RPO=0 when connectivity returns and preserves usable remote data during this process, ensuring both local and remote copies are maintained.

MetroCluster and SyncMirror

Synchronous replication in ONTAP is implemented through SyncMirror, which establishes two complete sets of RAID-protected data at separate locations. These locations can be within the same data center or many kilometers apart. SyncMirror is integrated with ONTAP and functions above the RAID level, ensuring compatibility with ONTAP features such as Snapshot copies, SnapRestore, and NetApp FlexClone®. The system operates as ONTAP with an additional layer for synchronous data mirroring.

A group of ONTAP controllers that manage SyncMirror data is referred to as NetApp MetroCluster. Various configurations are available, with MetroCluster designed to provide high-availability access to synchronously mirrored data for routine operations and disaster recovery scenarios.

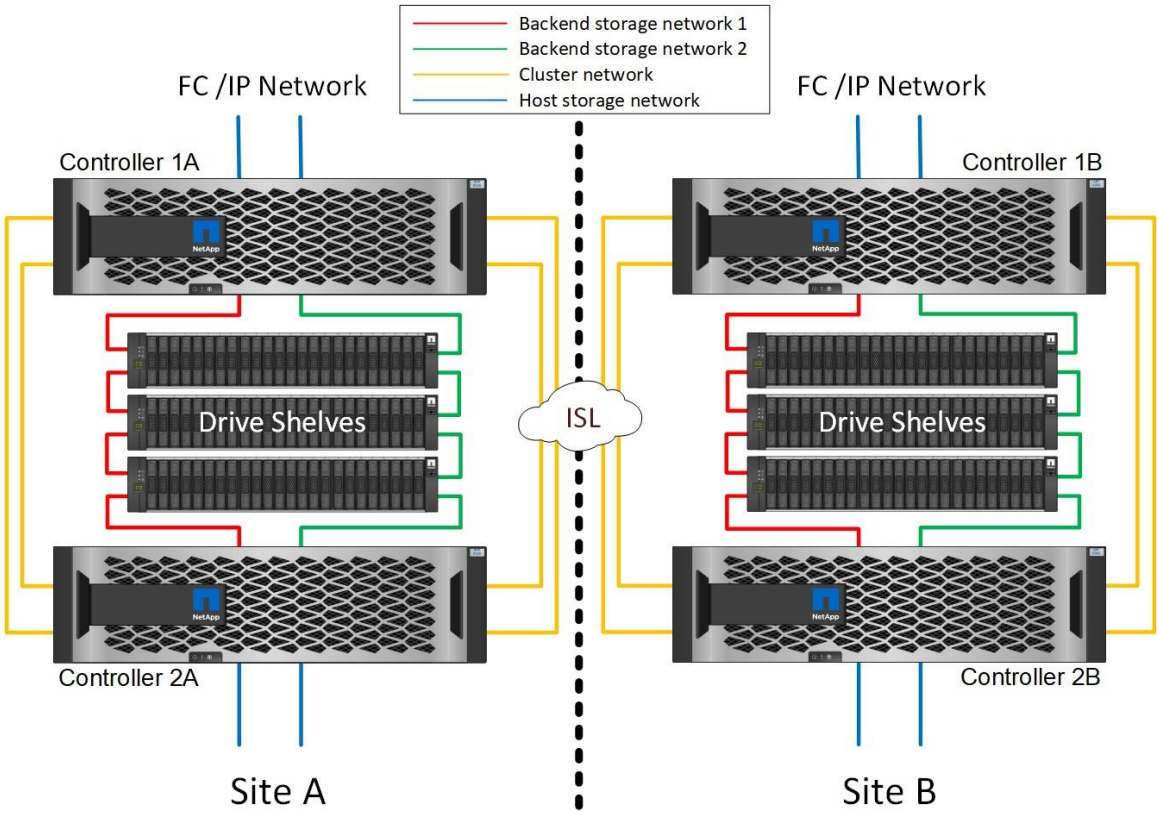
MetroCluster and SyncMirror provide data protection by guaranteeing synchronous mirroring across locations during normal operations. Write operations are acknowledged only when the data is stored on nonvolatile media at both sites. If connectivity between sites is disrupted, SyncMirror switches to asynchronous mode, allowing the primary site to continue serving data until connectivity is restored. Upon restoration, resynchronization is achieved by updating only the changes from the primary site, without requiring full reinitialization.

SnapMirror remains compatible with systems utilizing SyncMirror. For instance, a primary database operating on a MetroCluster distributed across two geographic sites can also replicate backups to a third location for archival purposes or to create clones for use in a DevOps environment.

MetroCluster architecture

While a full overview of MetroCluster is outside this document’s scope, it’s important to understand its key availability features. These examples use IP-based MetroCluster, as most customers prefer IP for its simpler infrastructure. Previously, cross-site connections relied on dark fibre and FC switches, but now high-speed, low-latency IP circuits are commonly available. For details, refer to ONTAP documentation and MetroCluster IP Solution Architecture and Design. MetroCluster systems with IP connectivity have HA pairs at each site.

Figure 18) MetroCluster IP basic architecture



MetroCluster resiliency features

MetroCluster has no single points of failure: each controller has two independent connections to both local and remote drive shelves, to controllers at the other site, and to its local partner in HA-pair setups. Any component can be removed without disrupting data availability. The main difference is that after a site failure, only the HA-pair configuration maintains overall high availability.

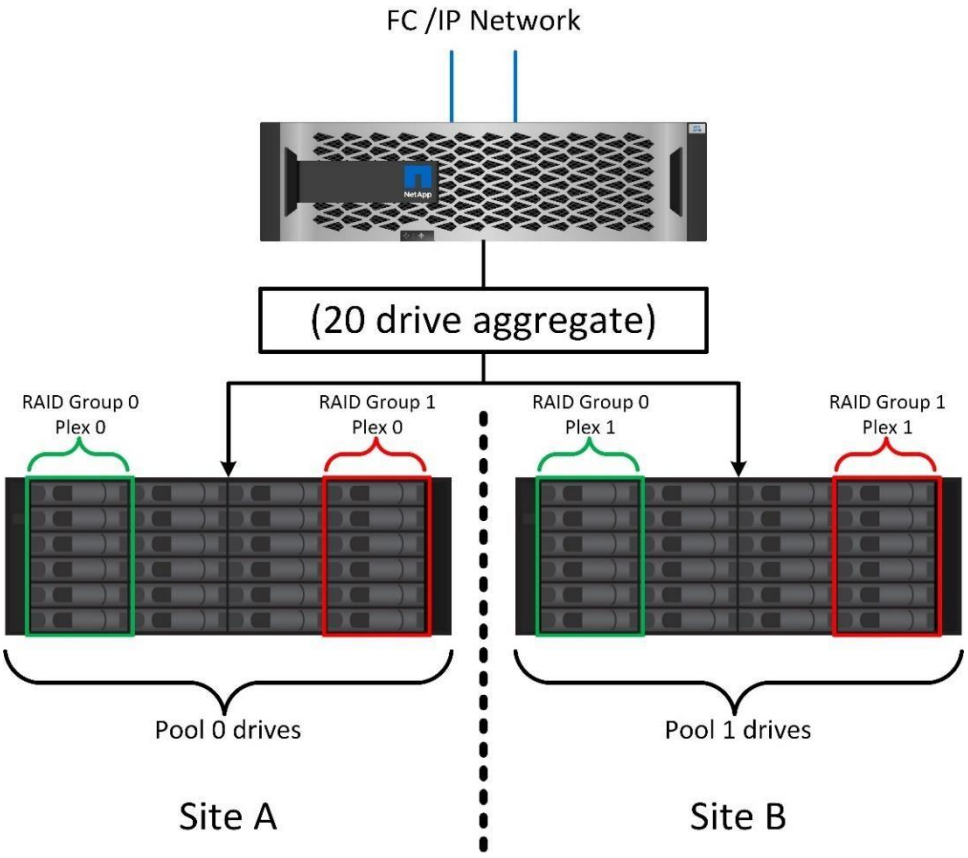
Site failure protection: NVRAM and MetroCluster

MetroCluster safeguards NVRAM data by replicating it to both local and remote partners, only acknowledging writes after all partners receive the data. This protects in-flight I/O from site failures but does not involve drive-level replication. The controller manages aggregate data replication by writing to each plex and relies on replicated NVRAM if a partner controller must take over after a failure.

Site and shelf failure protection: SyncMirror and Plexes

SyncMirror mirrors data between two RAID groups, enhancing the protection of RAID DP or RAID-TEC. Drives are divided equally between site A and site B, forming mirrored aggregates—e.g., 20 drives split for redundancy. Each pool is an independent RAID group, ensuring data integrity if one site fails. For example, a 24-drive aggregate uses 12 drives from each site, creating two mirrored RAID groups. SyncMirror is mainly used in MetroCluster systems for remote mirroring, with data copies at both sites, but sometimes adds extra redundancy within a single system—for instance, in automotive testing platforms where racks have separate power and UPS supplies.

Figure 19) SyncMirror



Hardware-assisted takeover

The service processor is an out-of-band management device integrated into AFF and FAS systems. It has its own IP address and provides direct console access and additional management functions regardless of the controller's operational status.

ONTAP can initiate a takeover of a failed node once it no longer receives a heartbeat from the partner node, subject to certain timeouts. Hardware-assisted takeover utilizes the service processor to expedite this process by detecting failures more rapidly and initiating the takeover without delay. This approach begins the takeover procedure without waiting for ONTAP to identify that the partner's heartbeat has stopped.

Switchover and switchback

Switchover and switchback describe moving volumes between remote controllers in a MetroCluster setup, applying only to remote nodes. In a four-volume MetroCluster configuration, local node failover uses the standard takeover and giveback methods.

Planned switchover and switchback

A planned switchover or switchback is a controlled process involving several steps to transfer storage and network resources between nodes. While the full command can take minutes, the actual control handoff happens quickly. The main distinction from takeover-giveback lies in SAN connectivity: during local takeover-giveback, all SAN paths to the affected node are lost and MPIO handles path changes

without relocating ports. In switchover-switchback, virtual FC target ports move between sites, briefly disappearing from the SAN before reappearing on another controller.

ONTAP Mediator with MetroCluster IP

The ONTAP Mediator, used with MetroCluster IP and some ONTAP solutions, acts as a tiebreaker to ensure NVRAM and SyncMirror sync. It determines if switchover is safe by verifying data consistency between sites. Failover or switchover is blocked when data is out of sync unless forced, accepting possible data loss.

SnapMirror Active Sync (SMas)

MetroCluster offers zero data loss in the event of a disaster with minimal service disruption for an entire environment. Not all customers require RPO=0 data protection for every dataset; some may only need synchronous data protection for specific datasets. SMAS (SnapMirror Active Sync), introduced in ONTAP 9.8, provides this functionality. SMAS and SM-S (SnapMirror Synchronous) use the same replication engine, but SMAS includes additional capabilities such as transparent application failover and failback.

Modes

SMas has two modes: Synchronous and StrictSync. In Synchronous mode (similar to MetroCluster), RPO=0 is maintained and writes go to both local and remote systems; if replication fails, the process continues, risking data loss if a site fails. Most customers use this mode. For workloads needing absolute consistency, StrictSync mode stops processing and returns an error if replication fails, usually causing application shutdown.

Path access

SMAS enables storage devices to appear on host operating systems from both primary and remote arrays. Local controller paths (ASA) are Active/Optimized, while remote controller paths are Active/Nonoptimized. Normally, I/O is handled by local controllers, but in a site failure or storage failover, remote paths switch to optimized status.

Failover

SMAS offers planned and unplanned storage failover. Planned failover is done manually for upgrades, patching, disaster recovery tests, or to rotate sites for business continuity. Unplanned failover happens automatically by the mediator for emergencies.

Storage hardware

SMAS provides platform flexibility by allowing different hardware at each site, unlike other disaster-recovery storage solutions. You can choose identical systems for full workloads or smaller, cost-effective hardware at remote sites if less I/O is needed after a disaster.

ONTAP Mediator

ONTAP Mediator is NetApp software that automates failover between primary and remote storage clusters. It runs on a lightweight VM on-premises or in the cloud, acting as a third site to monitor and manage failover events. In split-brain cases, it resumes I/O on the master node and automatically resyncs when sites reconnect.

SAN configuration best practices

Adhering to the following best practices is essential for optimizing SAN availability. Most recommendations pertain to host and Fibre Channel network configuration, reflecting the nuances and constraints of SAN technologies, operating systems, and multipathing software. Although there may be

situations where diverging from these guidelines is appropriate, administrators are advised to thoroughly evaluate potential implications and risks before proceeding.

Independent FC fabrics

FC SAN hosts should have redundant network connections, each on separate FC fabrics, to minimize downtime from port failures. Full-mesh configurations provide many host paths but also raise the risk that user mistakes could impact the entire SAN.

Independent IP subnets

For iSCSI and NVMe/TCP hosts requiring high availability, it is recommended to deploy a minimum of two network adapters (NICs), each allocated to a different subnet. Utilizing a single subnet for all TCP/IP operations increases vulnerability to network-wide disruptions and could result in service outages. Additionally, most operating systems employ internal routing tables that route network traffic through only one NIC per subnet, leaving any additional NICs on the same subnet inactive.

To improve redundancy and optimize throughput, host bonding solutions such as LACP trunking should be utilized on every subnet. A standard high-availability configuration for iSCSI or NVMe/TCP environments generally includes:

- NIC #1 assigned the IP address 192.168.1.10/24 on the host
- NIC #2 assigned the IP address 192.168.2.10/24 on the host
- ONTAP controller #1 featuring a two-port LACP trunk at 192.168.1.1/24
- ONTAP controller #2 featuring a two-port LACP trunk at 192.168.2.1/24

This architecture delivers load-balanced SAN access via the trunked interfaces on ONTAP controllers and upholds robust redundancy for hosts. The use of segregated subnets additionally ensures that a failure in any individual network segment does not affect overall SAN connectivity.

Administrators may implement separate subnets and VLANs for logical traffic isolation. While redundant subnets strengthen resilience, distinct VLANs bolster security by maintaining discrete workflows.

SU path limits

SAN hosts should be limited to four paths to a SU, with a maximum of eight. Too many paths can delay OS booting, cause path failover issues, reveal bugs in path management, and increase the likelihood of user errors when administering SAN devices.

Storage unit sizing

ONTAP controllers reach peak performance with just eight Storage Units (SUs), and adding more usually brings little benefit. More SUs may only be necessary if one application uses all available capacity. Fewer, larger SUs are preferred to avoid path issues—for example, using four 2TB or eight 1TB SUs for an 8TB database. For a single dataset, the recommended maximum is 64 SUs or 16 namespaces per controller; multiple datasets can each have their own SUs, such as ten databases with eight SUs apiece.

Single-initiator zoning

Use single-initiator zoning to prevent issues from initiator crosstalk, which can cause unexpected problems with certain operating systems or HBA/firmware setups. Multi-target zoning is allowed; avoid multi-initiator zoning.

Verify HBA/firmware/OS against IMT

Always verify your SAN setup with the NetApp Interoperability Matrix (IMT), particularly when upgrading your host OS or ONTAP. NetApp thoroughly tests operating systems, HBAs, firmware, SAN drivers, file systems, and features such as SMAS, MetroCluster, and cloning to ensure your SAN stays stable and continues to support ONTAP features—even through frequent SU changes.

Recently, NetApp streamlined the IMT by removing categories and variables where years of interoperability testing showed minimal issues. This effort reduced IMT components by 46%, making the IMT leaner and more user-friendly. For example, Ethernet switches are now considered commodity hardware and generally assumed to be supported. However, if you need to confirm compatibility for specific switch firmware versions or other variables, NetApp recommends also consulting vendor websites for their qualification results. There, you can find specifics about interoperability with particular firmware or OS versions.

SAN configuration against the SAN Host Utilities documentation

Most operating systems function as installed, but some setups need extra configuration. For details, see the ONTAP San Host Utilities documentation.

Use of sanlun utilities to verify path health

NetApp Host Utilities should be deployed on all supported operating systems. The primary utility is the sanlun command, which enables users to execute `sanlun lun show -p` to assess path integrity. This verification is crucial prior to ONTAP or SAN infrastructure upgrades. Numerous support cases involving outages have been attributed to missing paths, often due to initial zoning of only one controller or modifications to the SAN configuration over time.

Ensuring the appropriate number of paths are present, and that both controllers in a high-availability pair are included, mitigates potential configuration oversights. This process also identifies possible operating system misconfigurations or faults before changes are made to the SAN, thereby reducing the risk of outages.

If use of the sanlun command is not feasible, suitable multipath management tools available within the operating system should be utilized as an alternative.

Note on Linux LVM

A design limitation in Linux LVM can result in I/O errors and application failures during path transitions. At boot time, the multipath and LVM drivers initiate nearly simultaneously, which may cause a race condition. Typically, the multipath driver completes device creation before LVM begins; however, this sequence is not always assured.

Consequently, LVM may establish physical volume (PV) devices using single-path devices if the multipath device is unavailable when LVM scans for devices. If a logical volume (LV) relying on such a PV is mounted and a failover occurs, the PV may become inaccessible due to the loss of its sole path. While only specific configurations with sufficient storage units are susceptible to this issue, it has been reported by NetApp customers.

An indicator of a potentially unsafe environment can be found in the output of the `pvs` command, which may flag certain PVs that are not associated with a multipath device.

```
  WARNING: Not using device /dev/mapper/3600a0980383038616e3f4a53716a4c7a for PV 4ZZweF-
tjt9wLxC-CdPU-oQmT-78Wy-My6st2.
  WARNING: Not using device /dev/mapper/3600a0980383038616e3f4a53716a4d32 for PV O3IihV-zEaH-
J82B-fF8B-NGvz-dlPe-uUgb1r.
  WARNING: Not using device /dev/mapper/3600a0980383038616e3f4a53716a4d31 for PV XvjZty-T1qx-
7aHc-nrtI-yh3N-CWAv-U5gwrX.
  WARNING: Not using device /dev/mapper/3600a0980383038616e3f4a53716a4d30 for PV tl9BmZ-
3dCYLfvs-s7xR-3jfN-NLLT-dFGLc0.
```

This issue can be addressed by changes to `/etc/lvm/lvm.conf`. The default setting is as follows, and results in lvmd scanning all devices for physical volumes.

```
filter = [ "a|.*/|" ]
```

In general, the following setting works, but it must be tested carefully.

```
filter = [ "a|^/dev/sda[1-9]$|", "a|^/dev/mapper/*|", "r|^/dev/*|" ]
global_filter = [ "a|^/dev/sda[1-9]$|", "a|^/dev/mapper/*|", "r|^/dev/*|" ]
```

This filter results in lvmd scanning for physical volumes on `/dev/sda*` and `/dev/mapper/*` only. If your boot device is not a `/dev/sda` partion, this setting can interfere with rebooting. If for example, your server's local boot device might appear as a `/dev/xda` device. Consult the official LVM documentation for more details.

Caution: If you change this file, reboot the server to make sure that a reboot is successful. Also, be prepared to log on at a console to fix errors.

Note on /etc/sysconfig/oracleasm errors

With Oracle database and ASMlib, ensure /etc/sysconfig/oracleasm ignores single path devices. Linux supports both single- and multipath devices, but ASMlib should only detect multipath devices.

Note on host_config script with Solaris

Always review the ONTAP San Host Utilities documentation. Solaris needs specific configuration steps for proper ONTAP multipath device recognition. Not following host configuration instructions can compromise resilience and cause serious ZFS performance issues.

NVFAIL

For any SAN volume on ONTAP storage holding critical data, enable the nvfail parameter. SAN workloads face risk of corruption during forced failover or switchover because disk caches may lose previously acknowledged changes, causing mismatched database states. The nvfail setting also safeguards against NVRAM journaling failures that threaten data integrity. When enabled, if ONTAP detects NVRAM errors at startup, affected volumes enter innvfailed-state, blocking access and triggering a controlled shutdown to protect the database.

Appendix A: Improving performance with QoS, LUN Striping and SU/volume/aggr layouts

Note: This appendix applies to Unified ONTAP on AFF/FAS, These recommendations aren't meant for ASA

Table 5) QoS, LUN stripping, and SU/Vol/Aggr layouts to optimize performance

Best Practice Summary

Container	Best practice recommendations
Aggregate	Use multiple aggregates per node/controller; avoid very large aggregates.
Volume	Distribute volumes across aggregates; avoid overloading single volumes.
Storage Unit (LUN/namespace)	Stripe heavy workloads across multiple LUNs/volumes/aggregates.

1. **Quality of Service (QoS)**

Definition:
QoS in ONTAP lets administrators set limits on throughput or IOPS for storage objects to manage workload performance.

Purpose:

- Prevents any single workload from monopolizing resources and affecting others.
- Useful in multi-tenant or mixed environments like VMware, Hyper-V, SAP HANA, and databases.
- Helps control the effects of bursty, write-heavy workloads that cause long Consistency Points and higher latency.

Implementation Notes:

- QoS policies may be applied at the SU, volume, or SVM level.
- In this instance, the customer was hesitant due to administrative demands and potential effects on workloads.
- NetApp suggests using QoS to limit the maximum throughput for specific workloads.

2. **LUN Striping**

Definition:

LUN striping involves distributing an application's data across multiple Storage Units (SUs), and ideally across various volumes and aggregates, instead of directing all input/output operations to a single SU.

Purpose in this Case:

- Helps manage performance limits related to single SUs, particularly for workloads requiring high throughput or experiencing bursts in write activity.
- ONTAP processes one Consistency Point (CP) per aggregate at a time; substantial write activity on a single SU may saturate a CP and increase latency.
- Striping data across several SUs distributes the I/O load, enables parallel processing, and reduces the likelihood of a CP bottleneck.

Implementation Notes:

- For workloads such as SAP HANA, Oracle, or large databases, distributing data and logs over multiple SUs is considered standard practice.
- SU striping may be performed at the host level using tools like LVM, ASM, or other volume management solutions.
- In this situation, it was advised to separate high-throughput workloads onto multiple SUs, volumes, or aggregates; however, the customer initially had reservations due to potential complexity.

3. **SU, Volume, and Aggregate Layout**

Definition:

This describes how storage resources are mapped and distributed across the physical and logical layers in ONTAP (Storage Units mapped to volumes, which are contained within aggregates).

Case Observations:

a. Aggregates

- Each aggregate processes one consistency point (CP) at a time.
- Aggregates with larger amounts of data and more volumes or Storage Units may result in longer CP durations, particularly when using MetroCluster aggregate-level snapshots.
- Provisioning additional aggregates per controller allows for increased parallelism by enabling multiple CPs to run concurrently, which can decrease the workload per CP and lower the risk of aggregate-level bottlenecks.

b. Volumes

- Volumes distributed across different aggregates can help prevent concentrated activity or "hot spots."
- Volumes that are appropriately sized and not overloaded with excessive Storage Units or I/O maintain consistent performance.

c. Storage Units (SUs)

- Placing all high-I/O or critical workloads on a single Storage Unit can create performance issues.

- Workloads spread across multiple Storage Units, and ideally across various volumes and aggregates, support balanced system operations.

Operating system specific recommendations

This paragraph shows to how use SU striping for some operating systems (OS).

LINUX

Use LVM2 to create striped logical volumes. The example below is using 4 LUNs which have been already discovered by the OS:

1. Initialize all LUNs as a physical volume.
   pvcreate /dev/mapper/lun1

   pvcreate /dev/mapper/lun2

   pvcreate /dev/mapper/lun3

   pvcreate /dev/mapper/lun4

2. Create the volume group for each data and log partition.
   vgcreate <volumegroup> /dev/mapper/lun1 /dev/mapper/lun2 dev/mapper/lun3 /dev/mapper/lun4

3. Create a logical volume. Use a stripe number that is equal to the number of LUNs used per volume group (in this example, it is four) and a stripe size recommended for your application.
   lvcreate --extents 100%FREE -i 4 -I 64k --name vol <volumegroup>

4. In case of clustered environments scan the physical volumes, volume groups, and vol groups at all other hosts.
   modprobe dm_mod

   pvscan

   vgscan

   lvscan

Note: If the runtime of a file system creation seems to hang or takes very long, specify the option *-K* during XFS file system creation (*mkfs.xfs*) or option *-E nodiscard* during EXT4 file system creation (*mkfs.ext4*). This is LVM specific and independent of the used LUNs.

Windows

Configure LUN striping by using the Windows disk manager option “New striped volume” and choose the desired SUs.

QoS

ONTAP QoS offers the option to restrict workloads, so these don’t contend with other workloads. Best practice is to apply a non-shared throughput ceiling QoS group-policy to each LUN within each Storage Virtual Machine (SVM) to restrict the max throughput of each individual storage object to the given value. This reduces the possibility that a single workload can negatively influence other workloads.

To do so, a group-policy needs to be created using the CLI of the ONTAP cluster for each SVM:

qos policy-group create -policy-group <policy-name> -vserver <vserver name> -max-throughput 1000Mb/S -is-shared false

and applied to each SU within the SVM. Below is an example to apply the policy group to all existing SUs within an SVM:

*lun modify -vserver <vserver name> -path * -qos-policy-group <policy-name>*

This needs to be done for every SVM. The name of the QoS police group for each SVM needs to be different.

For new SUs to be created, the policy can be applied directly:

lun create -vserver <vserver_name> -path /vol/<volume_name>/<lun_name   -size <size> -ostype <e.g. linux> -qos-policy-group <policy-name>

In addition, the QoS burst percentage should be changed from 50% to 5% to limit the possibility of microbursts influencing the overall performance.

[cluster-name]::> set diag

[cluster-name]::*> options qos.rbr.burst_percent –vserver * -option-value 5
[cluster-name]::*> options qos.rbr.grace_period -vserver * -option-value 5

[cluster-name]::> set admin*

Documentation about QoS is available here: Set a throughput ceiling with QoS

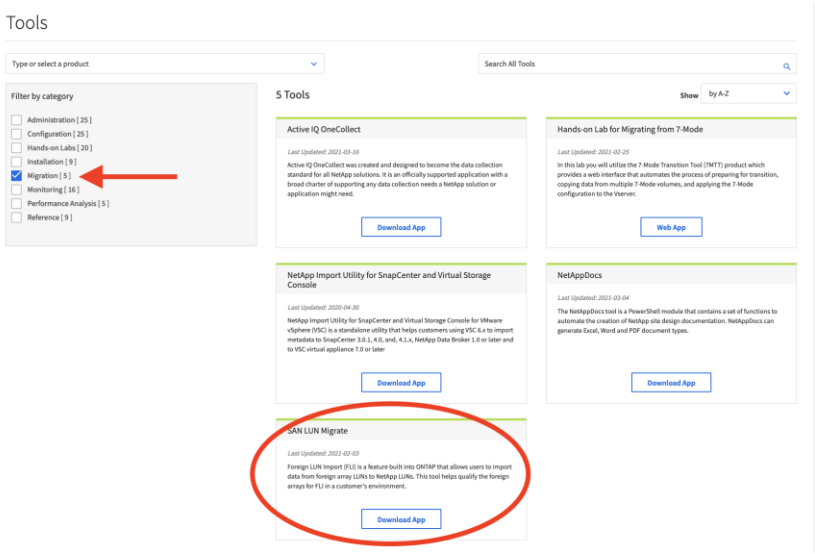
Appendix B: Field Qualifying Foreign LUN Import with a new source array

The FLI IMT inherited the qualification results from NetApp FlexArray. Once FlexArray End of Availability (EOA) was announced further qualifications and therefore IMT entries dwindled. We therefore decided to offer a tool that could be used to “field qualify” arrays not listed on the FLI IMT. The SAN LUN Migration tool can be found here:

https://mysupport.netapp.com/site/tools/tool-eula/san-lun-migrate

An explanation of what the tool is, how it works, how to run it, and how to interpret results are all covered in the accompanying SAN LUN Migrate.pdf. ONTAP Foreign LUN Import documentation covers foreign LUN Import (FLI) migrations.

Figure 20) Downloading the FLI field qualification tool FLI IMT and SAN LUN Migrate tool



Appendix C: Can ONTAP support AS/400

Can IBM operating systems (IBM i) be connected to NetApp storage?

Support for AS/400 is not available in most current configurations, except when their VIOS mode is used to emulate NFS.

The IBM iSeries, also referred to as IBM System i, AS/400, iSeries, OS/400, i5/OS, or System/36, has undergone multiple re-brandings by IBM. Users often refer to the system with different names based on when they began working with the product.

History

IBM has a Unix-like OS commonly known as AS/400, despite several name changes. Its use of 520-byte block LUNs instead of standard 512-byte blocks was initially manageable when IBM dominated the market, but later posed difficulties as their monopoly faded and other vendors were reluctant to support this niche requirement. Maintaining dedicated hardware and storage for a shrinking customer base became inefficient for IBM.

iSeries -> IBM Power

IBM resolved hardware issues by merging AS/400 and AIX server lines, making their differences minimal. For NetApp support, customers must use a standard IBM Power platform, not older System i or iSeries hardware. Most clients today run IBM i on Power platforms, which also allow simultaneous partitions for IBM i, AIX, and Linux.

520-bytes -> 512-bytes

The customer must use IBM PowerVM virtualization with vSCSI storage. To address the 520-byte block requirement, IBM implemented a feature where the PowerVM VIOS combines nine ONTAP LUNs—eight for data and one for the extra eight bytes per LUN—so 512-byte LUNs become 520-byte LUNs. Many customers now use this setup, which competitors can replicate if they support AIX.

Supportability

In this configuration, ownership of the LUNs resides with the hypervisor or VIOS. When provisioning AS/400 LUNs, they are mapped to the HBA on the VIOS partition, which functions as a streamlined version of AIX. VIOS is included as an operating system on our IMT, offering broad compatibility.

The VIOS manages the LUNs as standard AIX multipath LUNs and re-virtualizes each collection of nine LUNs into a single AS/400 520-byte LUN for the AS/400 partition. The AS/400 partition remains unaware of the LUNs' original source, recognizing only a single IBM LUN shared via VIOS.

Additionally, certain newer iterations of IBM i have the capability to utilize 512-byte LUNs directly. While customer adoption has not yet been observed, support for this configuration is available; however, it would still require VIOS. In such cases, the VIOS partition would own the ONTAP LUN and share it directly with the IBM i partition, where it appears as an IBM LUN. VIOS maintains responsibility for connectivity and multipathing.

Provided that the VIOS partition is configured in a supportable manner and the LUNs are accessed as vSCSI devices, NetApp does not need to distinguish the underlying operating system. This arrangement remains transparent to our systems.

Conversely, operating systems employing NPIV devices instead of vSCSI devices access the LUNs directly, necessitating full NetApp support due to additional requirements for multipathing, failover, and error handling. While NPIV is supported for various clients, it has not been certified for IBM i partitions and is unlikely to function correctly in that scenario. Virtualization through vSCSI is straightforward to support, as ONTAP treats it similarly to a standard AIX OS.

Where to find additional resources

To learn more about the information that is described in this document, review the following documents and/or websites:

Quick answers from ASA and SAN SMEs

ng-ASA_Answers

Labs on demand

- **Hands-On Lab: Getting started with NetApp All SAN Array systems**
  https://labondemand.netapp.com/lab/gsnewasa
- **LOD: ASA Business Continuity for Virtualized SAN Workloads**
  https://labondemand.netapp.com/node/1350

ONTAP documentation

- **Express Guides**

  http://docs.netapp.com/ontap-9/index.jsp?topic=%2Fcom.netapp.doc.dot-cm-sanconf%2Fhome.html
- **All SAN Array Documentation Center**
  http://docs.netapp.com/allsan/index.jsp
- **Using the ONTAP command-line interface**
  https://docs.netapp.com/us-en/ontap/system-admin/command-line-interface-concept.html
- **ONTAP CLI support for ASA r2 storage systems**
  https://docs.netapp.com/us-en/asa-r2/learn-more/cli-support.html
- **SnapMIrror Active Sync (SMas)**
  https://docs.netapp.com/us-en/ontap/snapmirror-active-sync/
- **SnapCenter and SCV 6.1 release TOI links**
  https://netapp.hosted.panopto.com/Panopto/Pages/Sessions/List.aspx?folderID=8b924fa4-4291-4297-b81a-b26d00f5bc14
- **Data Infrastructure Insights**
  https://docs.netapp.com/us-en/data-infrastructure-insights/index.html
- **NetApp Harvest**
  https://netapp.github.io/harvest/latest/

Automation and REST API Documentation

- **ONTAP Automation**
  https://docs.netapp.com/us-en/ontap-automation/index.html
- **Summary of REST resources**
  https://docs.netapp.com/us-en/ontap-automation/resources/overview_categories.html
- **Getting started with REST API**
  https://docs.netapp.com/us-en/ontap-restapi/ontap/getting_started_with_the_ontap_rest_api.html
- **REST API implementation details**
  https://docs.netapp.com/us-en/ontap-automation/rest/rest_web_services_foundation.html

NetApp validated architectures

Reference application architectures designed to modernize your enterprise SAN

- **NVA-1126-Design: Oracle and SUSE NetApp Verified Architecture Design Edition**
  http://www.netapp.com/us/media/nva-1126-DESIGN.pdf
- **NVA-1127-Design: MongoDB and SUSE NetApp Verified Architecture Design Edition**
  http://www.netapp.com/us/media/nva-1127-DESIGN.pdf
- **NVA-1145-Design: NetApp, VMware, and Broadcom Verified Architecture Design Edition: With MS Windows Server 2019 and MS SQL Server 2017 Workloads**
  http://www.netapp.com/us/media/nva-1145-DESIGN.pdf
- **NVA-1147-DESIGN: SAP HANA on NetApp All SAN Array**
  http://www.netapp.com/us/media/nva-1145-DESIGN.pdf
- **NVA-1159-Design: Epic, Cache and Clarity a NetApp Verified Architecture Design Edition**
  https://www.netapp.com/pdf.html?item=/media/27905-nva-1159-design.pdf

Version history

Version	Date	Document version history
Version 1.0	June 2012	Covers scalable SAN in clustered Data ONTAP 8.1 storage operating system. Gives an overview of the technology and provides a comparison between the 7-Mode and clustered ONTAP storage OSs. Covers the multipathing model used by the clustered Data ONTAP storage OS.

Version	Date	Document version history
Version 2.0	June 2013	Covers scalable SAN in the clustered Data ONTAP 8.2 storage OS. Includes sections about path management for larger clusters and port sets.
Version 3.0	May 2015	Covers scalable SAN for ONTAP 8.3 storage clusters. Introduces and covers NetApp DataMotion™ for LUNs and selective LUN mapping.
Version 3.1	August 2015	Covers scalable SAN for ONTAP 8.3.1 storage clusters. Introduces online FLI capability.
Version 4.0	August 2016	Updated to cover new features in ONTAP 9: prescribed AFF SAN configurations, fast failover, consistent performance, and simplified provisioning.
Version 5.0	June 2017	Updated for ONTAP 9.2.
Version 6.0	November 2017	Updated for ONTAP 9.3.
Version 7.0	October 2017	Updated for ONTAP 9.4.
Version 8.0	December 2018	Updated for ONTAP 9.5.
Version 9.0	April 2019	Updated for ONTAP 9.6.
Version 10.0	November 2019	Updated for ONTAP 9.7.
Version 11.0	December 2020	Updated for ONTAP 9.8.
Version 12.0	June 2021	Updated for ONTAP 9.9.1.
Version 13.0	April 2023	Updated for ONTAP 9.10.1 – 9.12.1 P2
Version 14.0	April 2024	Updated for ONTAP 9.15.1
Version 15.0	October 2025	Updated for ONTAP 9.17.1 and ASA New

Contact us

Let us know how we can improve this technical report.

Contact us at docfeedback@netapp.com.

Include TECHNICAL REPORT 4080 in the subject line.

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-4080-0423