Lenovo

MetroCluster IP Installation and Configuration Guide

ONTAP® 9

Fourth edition (April 2023)

Contents

Chapter 1. Deciding whether to use the MetroCluster IP Installation and Configuration Guide

This guide describes how to install and configure the MetroCluster IP hardware and software components.

You should use this guide for planning, installing, and configuring a MetroCluster IP configuration under the following circumstances:

- You want to understand the architecture of a MetroCluster IP configuration.
- You want to understand the requirements and best practices for configuring a MetroCluster IP configuration.
- You want to use the command-line interface (CLI), not an automated scripting tool.

General information about ONTAP and MetroCluster configurations is also available.

ONTAP 9 Documentation Center

Chapter 2. Preparing for the MetroCluster installation

As you prepare for the MetroCluster installation, you should understand the MetroCluster hardware architecture and required components.

Differences between the ONTAP MetroCluster configurations

The various MetroCluster configurations have key differences in the required components.

In all configurations, each of the two MetroCluster sites is configured as an ONTAP cluster. In a two-node MetroCluster configuration, each node is configured as a single-node cluster.

Feature	IP configurations	Fabric-attached configurations
		Four- or eight-node
Number of controllers	Four	Four or eight
Uses an FC switch storage fabric	No	Yes
Uses an IP switch storage fabric	Yes	No
Uses FC-to-SAS bridges	No	Yes
Uses direct-attached SAS storage	Yes (local attached only)	No
Supports ADP	Yes (starting in ONTAP 9.4)	No
Supports local HA	Yes	Yes
Supports automatic switchover	No	Yes
Supports unmirrored aggregates	No	Yes

Access to remote storage in MetroCluster IP configurations

In MetroCluster IP configurations, the only way the local controllers can reach the remote storage pools is via the remote controllers. The IP switches are connected to the Ethernet ports on the controllers; they do not have direct connections to the disk shelves. If the remote controller is down, the local controllers cannot reach their remote storage pools.

This is different than MetroCluster FC configurations, in which the remote storage pools are connected to the local controllers via the FC fabric or the SAS connections. The local controllers still have access to the remote storage even if the remote controllers are down.

Considerations for using ONTAP Mediator or MetroCluster Tiebreaker

Starting with ONTAP 9.7, you can use either the ONTAP Mediator-assisted automatic unplanned switchover (MAUSO) in the MetroCluster IP configuration or you can use the MetroCluster Tiebreaker software. Only one of the two services can be used with the MetroCluster IP configuration.

The different MetroCluster configurations perform automatic switchover under different circumstances:

MetroCluster FC configurations using the AUSO capability (not present in MetroCluster IP configurations)

In these configurations, AUSO is initiated if controllers fail but the storage (and bridges, if present) remain operational.

MetroCluster IP configurations using the ONTAP Mediator service (ONTAP 9.7 and later)

In these configurations, MAUSO is initiated in the same circumstances as AUSO, as described above, and also after a complete site failure (controllers, storage, and switches).

Note: MAUSO is initiated only if nonvolatile cache mirroring (*DR mirroring*) and SyncMirror plex mirroring is in sync at the time of the failure.

MetroCluster IP or FC configurations using the Tiebreaker software in active mode

In these configurations, the Tiebreaker initiates unplanned switchover after a complete site failure.

Before using the Tiebreaker software, review the *Tiebreaker Software Installation and Configuration Guide*.

MetroCluster Tiebreaker Software Installation and Configuration Guide

Interoperability of ONTAP Mediator with other applications and appliances

You cannot use any third-party applications or appliances that can trigger a switchover in combination with ONTAP Mediator. In addition, monitoring a MetroCluster configuration with MetroCluster Tiebreaker software is not supported when using ONTAP Mediator.

How the ONTAP Mediator supports automatic unplanned switchover

The ONTAP Mediator stores state information about the MetroCluster nodes in mailboxes located on the Mediator host. The MetroCluster nodes can use this information to monitor the state of their DR partners and implement a Mediator-assisted automatic unplanned switchover (MAUSO) in the case of a disaster.

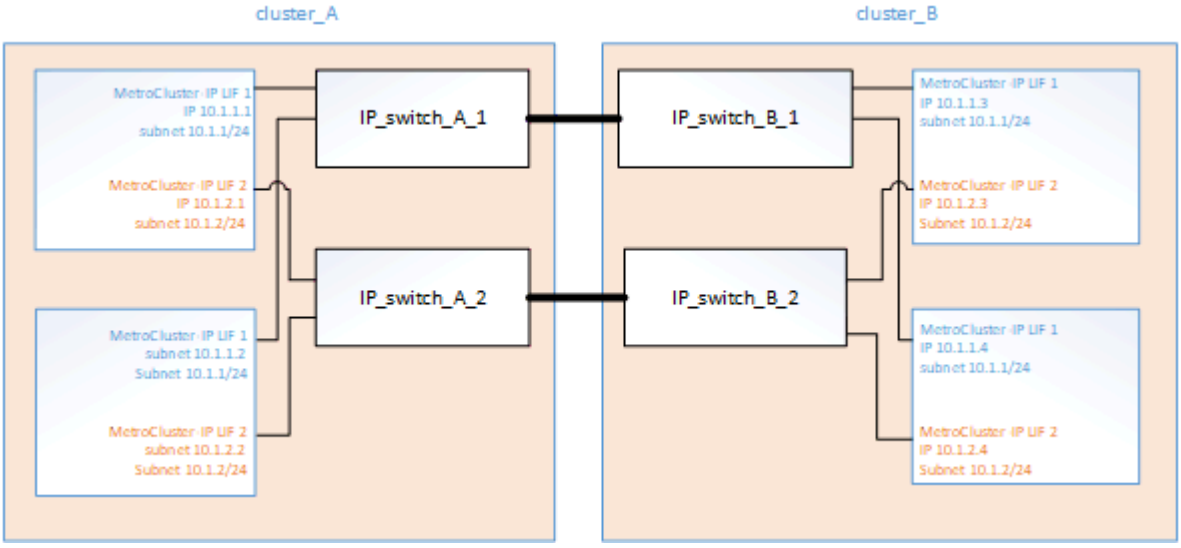
When a node detects a site failure requiring a switchover, it takes steps to confirm that the switchover is appropriate and, if so, performs the switchover.

MAUSO is only initiated if both SyncMirror mirroring and DR mirroring of each node's nonvolatile cache is operating and the caches and mirrors are synchronized at the time of the failure.

Considerations for MetroCluster IP configuration

You should be aware of how the MetroCluster IP addresses and interfaces are implemented in a MetroCluster IP configuration, as well as the associated requirements.

In a MetroCluster IP configuration, replication of storage and nonvolatile cache between the HA pairs and the DR partners is performed over high-bandwidth dedicated links in the MetroCluster IP fabric. iSCSI connections are used for storage replication. The IP switches are also used for all intra-cluster traffic within the local clusters. The MetroCluster traffic is kept separate from the intra-cluster traffic by using separate IP subnets and VLANs. The MetroCluster IP fabric is distinct and different from the cluster peering network.



The MetroCluster IP configuration requires two IP addresses on each node that are reserved for the back-end MetroCluster IP fabric. The reserved IP addresses are assigned to MetroCluster IP logical interfaces (LIFs) during initial configuration, and have the following requirements:

Note: You must choose the MetroCluster IP addresses carefully because you cannot change them after initial configuration.

- They must fall in a unique IP range.

  They must not overlap with any IP space in the environment.
- They must reside in one of two IP subnets that separate them from all other traffic.

For example, the nodes might be configured with the following IP addresses:

Node	Interface	IP address	Subnet
node_A_1	MetroCluster IP interface 1	10.1.1.1	10.1.1/24
	MetroCluster IP interface 2	10.1.2.1	10.1.2/24
node_A_2	MetroCluster IP interface 1	10.1.1.2	10.1.1/24
	MetroCluster IP interface 2	10.1.2.2	10.1.2/24
node_B_1	MetroCluster IP interface 1	10.1.1.3	10.1.1/24
	MetroCluster IP interface 2	10.1.2.3	10.1.2/24
node_B_2	MetroCluster IP interface 1	10.1.1.4	10.1.1/24
	MetroCluster IP interface 2	10.1.2.4	10.1.2/24

Characteristics of MetroCluster IP interfaces

The MetroCluster IP interfaces are specific to MetroCluster IP configurations. They have different characteristics from other ONTAP interface types:

- They are created by the `metrocluster configuration-settings interface create` command as part the initial MetroCluster configuration.

  They are not created or modified by the `network interface` commands.
- They do not appear in the output of the `network interface show` command.
- They do not fail over, but remain associated with the port on which they were created.

- MetroCluster IP configurations use specific Ethernet ports (depending on the platform) for the MetroCluster IP interfaces.

Considerations for automatic drive assignment and ADP systems

MetroCluster IP configurations support new installations with AFA systems using ADP (Advanced Drive Partitioning). In most configurations, partitioning and disk assignment is performed automatically during the initial configuration of the MetroCluster sites.

ONTAP 9.4 and later releases include the following changes for ADP support:

- Pool 0 disk assignments are done at the factory.
- The unmirrored root is created at the factory.
- Data partition assignment is done at the customer site during the setup procedure.
- In most cases, drive assignment and partitioning is done automatically during the setup procedures.

Note: When upgrading from ONTAP 9.4 to 9.5, the system recognizes the existing disk assignments.

Automatic partitioning

ADP is performed automatically during initial configuration of the platform.

Note: Starting with ONTAP 9.5, disk autoassignment must be enabled for automatic partitioning for ADP to occur.

How shelf-by-shelf automatic assignment works

If there are four external shelves per site, each shelf is assigned to a different node and different pool, as shown in the following example:

- All of the disks on site_A-shelf_1 are automatically assigned to pool 0 of node_A_1
- All of the disks on site_A-shelf_3 are automatically assigned to pool 0 of node_A_2
- All of the disks on site_B-shelf_1 are automatically assigned to pool 0 of node_B_1
- All of the disks on site_B-shelf_3 are automatically assigned to pool 0 of node_B_2
- All of the disks on site_B-shelf_2 are automatically assigned to pool 1 of node_A_1
- All of the disks on site_B-shelf_4 are automatically assigned to pool 1 of node_A_2
- All of the disks on site_A-shelf_2 are automatically assigned to pool 1 of node_B_1
- All of the disks on site_A-shelf_4 are automatically assigned to pool 1 of node_B_2

Manual drive assignment (ONTAP 9.5)

In ONTAP 9.5, manual drive assignment is required on systems with the following shelf configurations:

- Three external shelves per site.

  Two shelves are assigned automatically using a half-shelf assignment policy, but the third shelf must be assigned manually.
- More than four shelves per site and the total number of external shelves is not a multiple of four.

  Extra shelves above the nearest multiple of four are left unassigned and the drives must be assigned manually. For example, if there are five external shelves at the site, shelf five must be assigned manually.

You only need to manually assign a single drive on each unassigned shelf. The rest of the drives on the shelf are then automatically assigned.

Manual drive assignment (ONTAP 9.4)

In ONTAP 9.4, manual drive assignment is required on systems with the following shelf configurations:

- Fewer than four external shelves per site.

  The drives must be assigned manually to ensure symmetrical assignment of the drives, with each pool having an equal number of drives.
- More than four external shelves per site and the total number of external shelves is not a multiple of four.

  Extra shelves above the nearest multiple of four are left unassigned and the drives must be assigned manually.

When manually assigning drives, you should assign disks symmetrically, with an equal number of drives assigned to each pool. For example, if the configuration has two storage shelves at each site, you would one shelf to the local HA pair and one shelf to the remote HA pair:

- Assign half of the disks on site_A-shelf_1 to pool 0 of node_A_1.
- Assign half of the disks on site_A-shelf_1 to pool 0 of node_A_2.
- Assign half of the disks on site_A-shelf_2 to pool 1 of node_B_1.
- Assign half of the disks on site_A-shelf_2 to pool 1 of node_B_2.
- Assign half of the disks on site_B-shelf_1 to pool 0 of node_B_1.
- Assign half of the disks on site_B-shelf_1 to pool 0 of node_B_2.
- Assign half of the disks on site_B-shelf_2 to pool 1 of node_A_1.
- Assign half of the disks on site_B-shelf_2 to pool 1 of node_A_2.

Adding shelves to an existing configuration.

Automatic drive assignment supports the symmetrical addition of shelves to an existing configuration.

When new shelves are added, the system applies the same assignment policy to newly added shelves. For example, with a single shelf per site, if an additional shelf is added, the systems applies the quarter-shelf assignment rules to the new shelf.

ADP and disk assignment differences by system in MetroCluster IP configurations

The operation of Advanced Drive Partitioning (ADP) and automatic disk assignment in MetroCluster IP configurations varies depending on the system model.

Note: In systems using ADP, aggregates are created using partitions in which each drive is partitioned in to P1, P2 and P3 partitions. The root aggregate is created using P3 partitions.

You must meet the MetroCluster limits for the maximum number of supported drives and other guidelines.

ADP and disk assignment on AFA DM5000F systems

Guideline	Shelves per site	Drive assignment rules	ADP layout for root partition
Minimum recommended shelves (per site)	Internal drives only	The internal drives are divided into four equal groups. Each group is automatically assigned to a separate pool and each pool is assigned to a separate controller in the configuration. **Note:** Half of the internal drives remain unassigned before MetroCluster is configured.	Two quarters are used by the local HA pair. The other two quarters are used by the remote HA pair. The root aggregate includes the following partitions in each plex: • Three partitions for data • Two parity partitions • One spare partition
Minimum supported shelves (per site)	16 internal drives	The drives are divided into four equal groups. Each quarter-shelf is automatically assigned to a separate pool. Two quarters on a shelf can have the same pool. The pool is chosen based on the node that owns the quarter: • If owned by the local node, pool0 is used. • If owned by the remote node, pool1 is used. For example: a shelf with quarters Q1 through Q4 can have following assignments: • Q1: node_A_1 pool0 • Q2: node_A_2 pool0 • Q3: node_B_1 pool1 • Q4:node_B_2 pool1 **Note:** Half of the internal drives remain unassigned before MetroCluster is configured.	Each of the two plexes in the root aggregate includes the following partitions: • One partition for data • Two parity partitions • One spare partition

ADP and disk assignment on AFA DM7000F systems

Guideline	Shelves per site	Drive assignment rules	ADP layout for root partition
Minimum recommended shelves (per site)	Two shelves	The drives on each external shelf are divided into two equal groups (halves). Each half-shelf is automatically assigned to a separate pool.	One shelf is used by the local HA pair. The second shelf is used by the remote HA pair. Partitions on each shelf are used to create the root aggregate. The root aggregate includes the following partitions in each plex: • Eight partitions for data • Two parity partitions • Two spare partitions
Minimum supported shelves (per site)	One shelf	The drives are divided into four equal groups. Each quarter-shelf is automatically assigned to a separate pool.	Each of the two plexes in the root aggregate includes the following partitions: • Three partitions for data • Two parity partitions • One spare partition

Disk assignment on DM5000H systems

Guideline	Shelves per site	Drive assignment rules	ADP layout for root partition
Minimum recommended shelves (per site)	One internal and one external shelf	The internal and external shelves are divided into two equal halves. Each half is automatically assigned to different pool	Not applicable.
Minimum supported shelves (per site) (active/passive HA configuration)	Internal drives only	Manual assignment required.	

Disk assignment on DM7000H systems

Guideline	Shelves per site	Drive assignment rules	ADP layout for root partition
Minimum supported shelves (per site)	Two shelves	The drives on the external shelves are divided into two equal groups (halves). Each half-shelf is automatically assigned to a separate pool.	Not applicable.
Minimum supported shelves (per site) (active/passive HA configuration)	One shelf	Manual assignment required.	

Considerations for configuring cluster peering

Each MetroCluster site is configured as a peer to its partner site. You should be familiar with the prerequisites and guidelines for configuring the peering relationships and when deciding whether to use shared or dedicated ports for those relationships.

Prerequisites for cluster peering

Before you set up cluster peering, you should confirm that the connectivity, port, IP address, subnet, firewall, and cluster-naming requirements are met.

Connectivity requirements

Every intercluster LIF on the local cluster must be able to communicate with every intercluster LIF on the remote cluster.

Although it is not required, it is typically simpler to configure the IP addresses used for intercluster LIFs in the same subnet. The IP addresses can reside in the same subnet as data LIFs, or in a different subnet. The subnet used in each cluster must meet the following requirements:

- The subnet must have enough IP addresses available to allocate to one intercluster LIF per node.

  For example, in a six-node cluster, the subnet used for intercluster communication must have six available IP addresses.

Each node must have an intercluster LIF with an IP address on the intercluster network.

Intercluster LIFs can have an IPv4 address or an IPv6 address.

Note: ONTAP 9 enables you to migrate your peering networks from IPv4 to IPv6 by optionally allowing both protocols to be present simultaneously on the intercluster LIFs. In earlier releases, all intercluster relationships for an entire cluster were either IPv4 or IPv6. This meant that changing protocols was a potentially disruptive event.

Port requirements

You can use dedicated ports for intercluster communication, or share ports used by the data network. Ports must meet the following requirements:

- All ports that are used to communicate with a given remote cluster must be in the same IPspace.

  You can use multiple IPspaces to peer with multiple clusters. Pair-wise full-mesh connectivity is required only within an IPspace.
- The broadcast domain that is used for intercluster communication must include at least two ports per node so that intercluster communication can fail over from one port to another port.

  Ports added to a broadcast domain can be physical network ports, VLANs, or interface groups (ifgrps).
- All ports must be cabled.
- All ports must be in a healthy state.
- The MTU settings of the ports must be consistent.

Firewall requirements

Firewalls and the intercluster firewall policy must allow the following protocols:

- ICMP service
- TCP to the IP addresses of all the intercluster LIFs over the ports 10000, 11104, and 11105
- Bidirectional HTTPS between the intercluster LIFs

The default intercluster firewall policy allows access through the HTTPS protocol and from all IP addresses (0.0.0.0/0). You can modify or replace the policy if necessary.

Considerations when using dedicated ports

When determining whether using a dedicated port for intercluster replication is the correct intercluster network solution, you should consider configurations and requirements such as LAN type, available WAN bandwidth, replication interval, change rate, and number of ports.

Consider the following aspects of your network to determine whether using a dedicated port is the best intercluster network solution:

- If the amount of available WAN bandwidth is similar to that of the LAN ports and the replication interval is such that replication occurs while regular client activity exists, then you should dedicate Ethernet ports for intercluster replication to avoid contention between replication and the data protocols.
- If the network utilization generated by the data protocols (CIFS, NFS, and iSCSI) is such that the network utilization is above 50 percent, then you should dedicate ports for replication to allow for nondegraded performance if a node failover occurs.
- When physical 10 GbE or faster ports are used for data and replication, you can create VLAN ports for replication and dedicate the logical ports for intercluster replication.

  The bandwidth of the port is shared between all VLANs and the base port.
- Consider the data change rate and replication interval and whether the amount of data that must be replicated on each interval requires enough bandwidth that it might cause contention with data protocols if sharing data ports.

Considerations when sharing data ports

When determining whether sharing a data port for intercluster replication is the correct intercluster network solution, you should consider configurations and requirements such as LAN type, available WAN bandwidth, replication interval, change rate, and number of ports.

Consider the following aspects of your network to determine whether sharing data ports is the best intercluster connectivity solution:

- For a high-speed network, such as a 40-Gigabit Ethernet (40-GbE) network, a sufficient amount of local LAN bandwidth might be available to perform replication on the same 40-GbE ports that are used for data access.

  In many cases, the available WAN bandwidth is far less than 10 GbE LAN bandwidth.
- All nodes in the cluster might have to replicate data and share the available WAN bandwidth, making data port sharing more acceptable.
- Sharing ports for data and replication eliminates the extra port counts required to dedicate ports for replication.
- The maximum transmission unit (MTU) size of the replication network will be the same size as that used on the data network.
- Consider the data change rate and replication interval and whether the amount of data that must be replicated on each interval requires enough bandwidth that it might cause contention with data protocols if sharing data ports.
- When data ports for intercluster replication are shared, the intercluster LIFs can be migrated to any other intercluster-capable port on the same node to control the specific data port that is used for replication.

Considerations for sharing private layer 2 networks

Starting with ONTAP 9.6, MetroCluster IP configurations with supported Cisco switches can share existing networks for ISLs, rather than using dedicated MetroCluster ISLs. Earlier ONTAP versions require dedicated ISLs.

MetroCluster IP switches are dedicated to the MetroCluster configuration and cannot be shared. Therefore, a set of MetroCluster IP switches can only connect one MetroCluster configuration. Only the MetroCluster ISL ports on the MetroCluster IP switches can connect to the shared switches.

CAUTION:
If using a shared network, the customer is responsible for meeting the MetroCluster network requirements in the shared network.

MetroCluster ISL requirements in shared networks

When sharing ISL traffic in a shared network, you must ensure that you have adequate capacity and size the ISLs appropriately. Low latency is critical for replication of data between the MetroCluster sites. Latency issues on these connections can impact client I/O.

You should review these sections to correctly calculate the required end-to-end capacity of the ISLs. Continuous nonvolatile cache and storage replication with low latency is critical for MetroCluster configurations. The latency in the back-end network impacts the latency and throughput seen by client IO.

Latency and packet loss limits in the ISLs

The following requirements must be met for round-trip traffic between the MetroCluster IP switches at site_A and site_B, with the MetroCluster configuration in steady state operation:

- Round trip latency must be less than or equal to 7 ms.

  The maximum distance is 700 km, so the distance between the sites is limited by the latency or the maximum distance, whichever is reached first.

  As the distance between two MetroCluster sites increases, latency increases, usually in the range of 1 ms round-trip delay time per 100 km (62 miles). This latency also depends on the network service level agreement (SLA) in terms of the bandwidth of the ISL links, packet drop rate, and jitter on the network. Low bandwidth, high jitter, and random packet drops lead to different recovery mechanisms by the switches or the TCP engine on the controller modules for successful packet delivery. These recovery mechanisms can increase overall latency.

  Any device that contributes to latency must be accounted for.
- Packet loss must be less than or equal to 0.01%.

  Packet loss includes physical loss or loss due to congestion or over-subscription.

  Packet drops can cause retransmissions and a reduced congestion window.
- The supported jitter value is 3 ms for round trip (or 1.5 ms for one way).
- The network should allocate and maintain the SLA for the bandwidth required for MetroCluster traffic, accounting for microbursts and spikes in the traffic.

  Low bandwidth can cause queuing delays and tail drops on switches. If you are using ONTAP 9.7 or later, the network intermediate between the two sites must provide a minimum bandwidth of 4.5 Gbps for the MetroCluster configuration.
- MetroCluster traffic should not consume the complete bandwidth and have negative impact on non-MetroCluster traffic.
- The shared network should have network monitoring configured to monitor the ISLs for utilization, errors (drops, link flaps, corruption, etc.) and failures.

Connection limits and trunking in the customer switches

The intermediate customer-provided switches must meet the following requirements:

- The number of intermediate switches is not limited, and more than two switches between the MetroCluster IP switches is supported.

The MetroCluster IP switches should be located as close as possible to the intermediate switches providing the long-haul link. All of the ISL connections along the route must meet all of the requirements for MetroCluster ISL.

- The ISLs in the customer network (the ISLs between the customer switches) must be configured in such way that sufficient bandwidth is provided and order of delivery is preserved.

  This can be done with trunking a sufficient number of links and enforcing load balancing policies to preserve order.

Other network requirements

The intermediate, customer-provided switches must meet the following requirements:

- The MetroCluster traffic uses fixed VLAN IDs that are set in the provided RCF files.

  Layer 2 VLANs with IDs that match the MetroCluster VLAN IDs must span the shared network. AFA DM5000F and DM5000H systems require VLAN 10 and VLAN 20 unless they are changed during interface creation. Other systems are not restricted to specific VLAN IDs.
- The MTU size must be set to 9216 on all devices in the end-to-end network.
- No other traffic can be configured with a higher priority than class of service (COS) five.
- ECN (explicit congestion notification) must be configured on all end-to-end paths.

ISL cabling requirements

When using shared ISLs in a MetroCluster IP configuration, you must be aware of the requirements for the end-to-end MetroCluster ISL running from controller ports on site A to controller ports on site B.

Basic MetroCluster ISL requirements

The following requirements must be met:

- A native-speed ISL switch port must connect to a native-speed ISL switch port.

  For example, a 40 Gbps port connects to a 40 Gbps port.
- A 10 Gbps port that is in native mode (i.e., not using a breakout cable) can connect to a 10 Gbps port that is in native mode.
- The ISLs between the MetroCluster IP switches and the customer network, as well as the ISLs between the intermediate switches, follow the same rules in terms of speed.
- The number of ISLs that are between the MetroCluster switches and the customer network switches, and the number of ISLs that are between the customer network switches, do not need to match.

  For example, the MetroCluster switches can connect using two ISLs to the intermediate switches, and the intermediate switches can connect to each other using 10 ISLs.
- The speed of ISLs that are between the MetroCluster switches and the customer network switches, and the speed of ISLs that are between the customer network switches, do not need to match.

  For example, the MetroCluster switches can connect using a 40-Gbps ISL to the intermediate switches, and the intermediate switches can connect to each other using 100-Gbps ISLs.
- The number of and speed of ISLs connecting each MetroCluster switch to the intermediate switch must be the same on both MetroCluster sites.

Number of ISLs and breakout cables in the shared network

The number of ISLs connecting the MetroCluster IP switches to the shared network varies depending on the switch model and port type.

MetroCluster IP switch model	Port type	Number of ISLs
Broadcom-supported BES-53248 switches	Native ports	4 ISLs using 10 or 25-Gbps ports
Cisco 3132Q-V	Native ports	6 ISLs using 40 Gbps ports
	Breakout cables	16 x 10 Gbps ISLs
Cisco 3232C	Native ports	6 ISLs using 40 or 100 Gbps ports
	Breakout cables	16 x 10 Gbps ISLs

- The use of breakout cables (one physical port is used as 4 x 10 Gbps ports) is supported on Cisco switches.
- The RCF files for the IP switches have ports in native and breakout mode configured.

  A mix of ISL ports in native port speed mode and breakout mode is not supported. All ISLs from the MetroCluster IP switches to the intermediate switches in one network must be of same speed and length.
- The use of external encryption devices (for example, external link encryption or encryption provided via WDM devices) are supported as long as the round-trip latency remains within the above requirements.

For optimum performance, you should use at least a 1 x 40 Gbps or multiple 10 Gbps ISLs per network.

The maximum theoretical throughput of shared ISLs (for example, 240 Gbps with six 40 Gbps ISLs) is a best-case scenario. When using multiple ISLs, statistical load balancing can impact the maximum throughput. Uneven balancing can occur and reduce throughput to that of a single ISL.

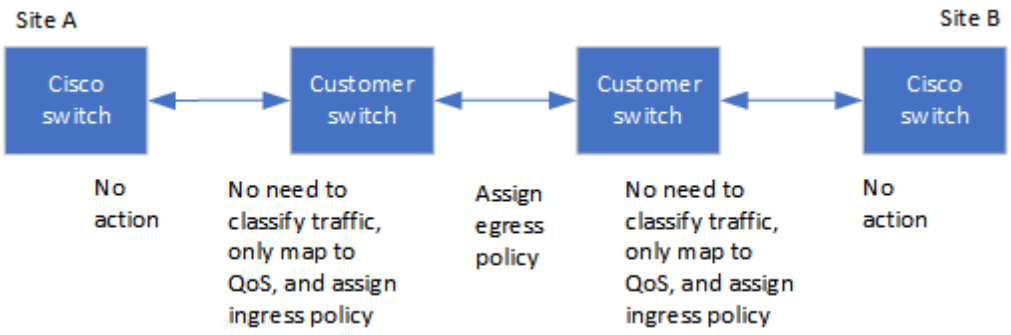
Required settings on intermediate switches

When sharing ISL traffic in a shared network, the configuration of the intermediate switches provided by the customer must ensure that the MetroCluster traffic (RDMA and storage) meets the required service levels across the entire path between the MetroCluster sites.

The following examples are for Cisco Nexus 3000 switches and IP Broadcom switches. Depending on your switch vendor and models, you must ensure that your intermediate switches have an equivalent configuration.

Cisco Nexus switches

The following diagram gives an overview of the required settings for a shared network when the external switches are Cisco switches.



In this example, the following policies and maps are created for MetroCluster traffic:

- A MetroClusterIP_Ingress policy is applied to ports on the intermediate switch that connect to the MetroCluster IP switches.

The MetroClusterIP_Ingress policy maps the incoming tagged traffic to the appropriate queue on the intermediate switch. Tagging happens on the node-port, not on the ISL. Non-MetroCluster traffic that is using the same ports on the ISL remains in the default queue.

- A MetroClusterIP_Egress policy is applied to ports on the intermediate switch that connect to ISLs between intermediate switches

You must configure the intermediate switches with matching QoS access-maps, class-maps, and policy-maps along the path between the MetroCluster IP switches. The intermediate switches map RDMA traffic to COS5 and storage traffic to COS4.

The following example shows the configuration for a customer-provided Cisco Nexus 3000 switch. If you have Cisco switches, you can use the example to configure the switch along the path without much difficulty. If you do not have Cisco switches, you must determine and apply the equivalent configuration to your intermediate switches.

The following example shows the class map definitions:

Note: This example is for configurations using Cisco MetroCluster IP switches. You can follow this example regardless of the switch types of the switches carrying MetroCluster traffic that do not connect to a MetroCluster IP switch.

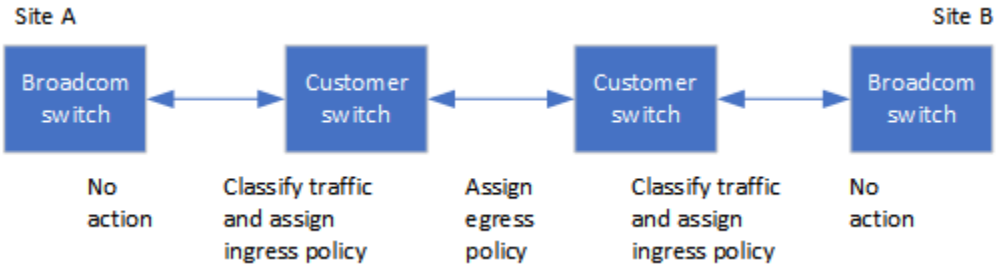
```
class-map type qos match-all rdma
          match cos 5
        class-map type qos match-all storage
          match cos 4
```

The following example shows the policy map definitions:

```
policy-map type qos MetroClusterIP_Ingress
           class rdma
             set dscp 40
             set cos 5
             set qos-group 5
           class storage
             set dscp 32
             set cos 4
             set qos-group 4
         policy-map type queuing MetroClusterIP_Egress
           class type queuing c-out-8q-q7
             priority level 1
           class type queuing c-out-8q-q6
             priority level 2
           class type queuing c-out-8q-q5
             priority level 3
             random-detect threshold burst-optimized ecn
           class type queuing c-out-8q-q4
             priority level 4
             random-detect threshold burst-optimized ecn
           class type queuing c-out-8q-q3
             priority level 5
           class type queuing c-out-8q-q2
             priority level 6
           class type queuing c-out-8q-q1
             priority level 7
           class type queuing c-out-8q-q-default
             bandwidth remaining percent 100
             random-detect threshold burst-optimized ecn
```

MetroCluster IP Broadcom switches

The following diagram gives an overview of the required settings for a shared network when the external switches are IP Broadcom switches.



Configurations using MetroCluster IP Broadcom switches require additional configuration:

- For exterior switches you must configure the access and class maps to classify the traffic on ingress to the customer network.

Note: This is not required on configurations using MetroCluster IP switches.

The following example shows how to configure the access and class maps on the first and last customer switches connecting the ISLs between the MetroCluster IP Broadcom switches.

```
ip access-list storage
  10 permit tcp any eq 65200 any
  20 permit tcp any any eq 65200
ip access-list rdma
  10 permit tcp any eq 10006 any
  20 permit tcp any any eq 10006

class-map type qos match-all storage
  match access-group name storage
class-map type qos match-all rdma
  match access-group name rdma
```

- You need to assign the ingress policy to the ISL switch port on the first customer switch.

The following example shows the class map definitions:

Note: This example is for configurations using Cisco MetroCluster IP switches. You can follow this example regardless of the switch types of the switches carrying MetroCluster traffic that do not connect to a MetroCluster IP switch.

```
class-map type qos match-all rdma
   match cos 5
class-map type qos match-all storage
   match cos 4
```

The following example shows the policy map definitions:

```
policy-map type qos MetroClusterIP_Ingress
   class rdma
      set dscp 40
      set cos 5
      set qos-group 5
   class storage
      set dscp 32
      set cos 4
      set qos-group 4
policy-map type queuing MetroClusterIP_Egress
   class type queuing c-out-8q-q7
      priority level 1
   class type queuing c-out-8q-q6
      priority level 2
   class type queuing c-out-8q-q5
      priority level 3
      random-detect threshold burst-optimized ecn
```

```
class type queuing c-out-8q-q4
   priority level 4
   random-detect threshold burst-optimized ecn
class type queuing c-out-8q-q3
   priority level 5
class type queuing c-out-8q-q2
   priority level 6
class type queuing c-out-8q-q1
   priority level 7
class type queuing c-out-8q-q-default
   bandwidth remaining percent 100
   random-detect threshold burst-optimized ecn
```

Intermediate customer switches

- For intermediate customer switches, you must assign the egress policy to the ISL switch ports.
- For all other interior switches along the path that carry MetroCluster traffic, follow the class map and policy map examples in the section *Cisco Nexus 3000 switches*.

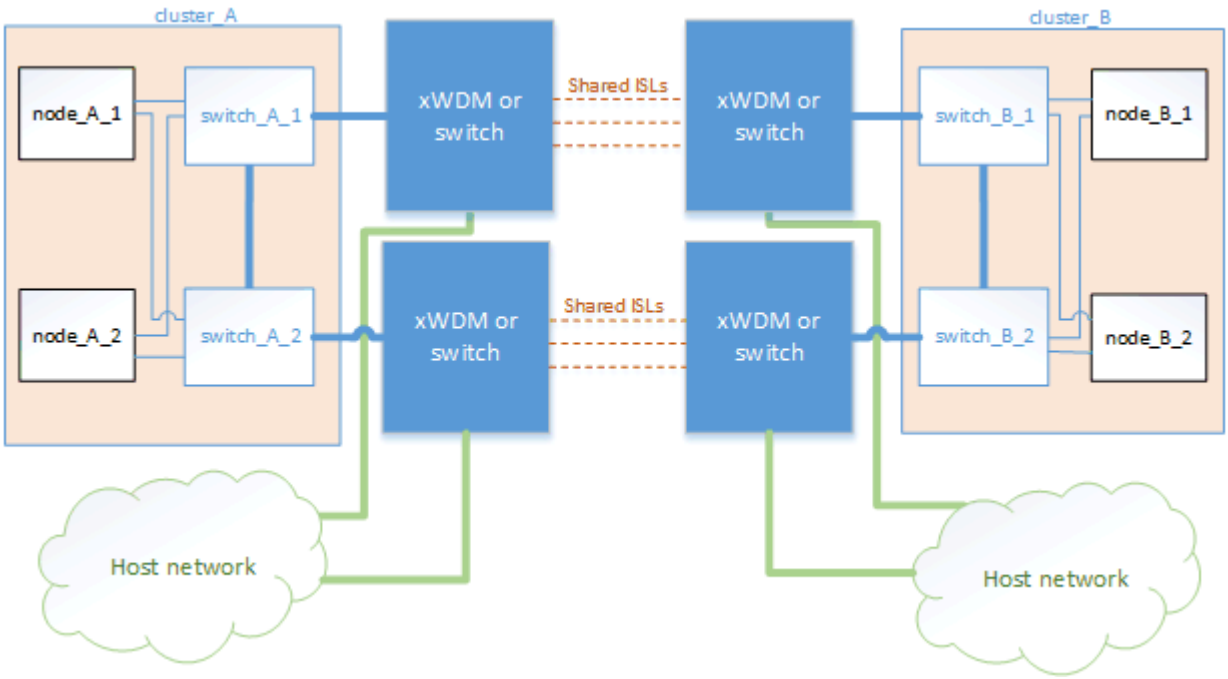
Examples of MetroCluster network topologies

Starting with ONTAP 9.6, some shared ISL network configurations are supported for MetroCluster IP configurations.

Shared network configuration with direct links

In this topology, two distinct sites are connected by direct links. These links can be between Wavelength Division Multiplexing equipment (xWDM) or switches. The capacity of the ISLs is not dedicated to the MetroCluster traffic but is shared with other traffic.

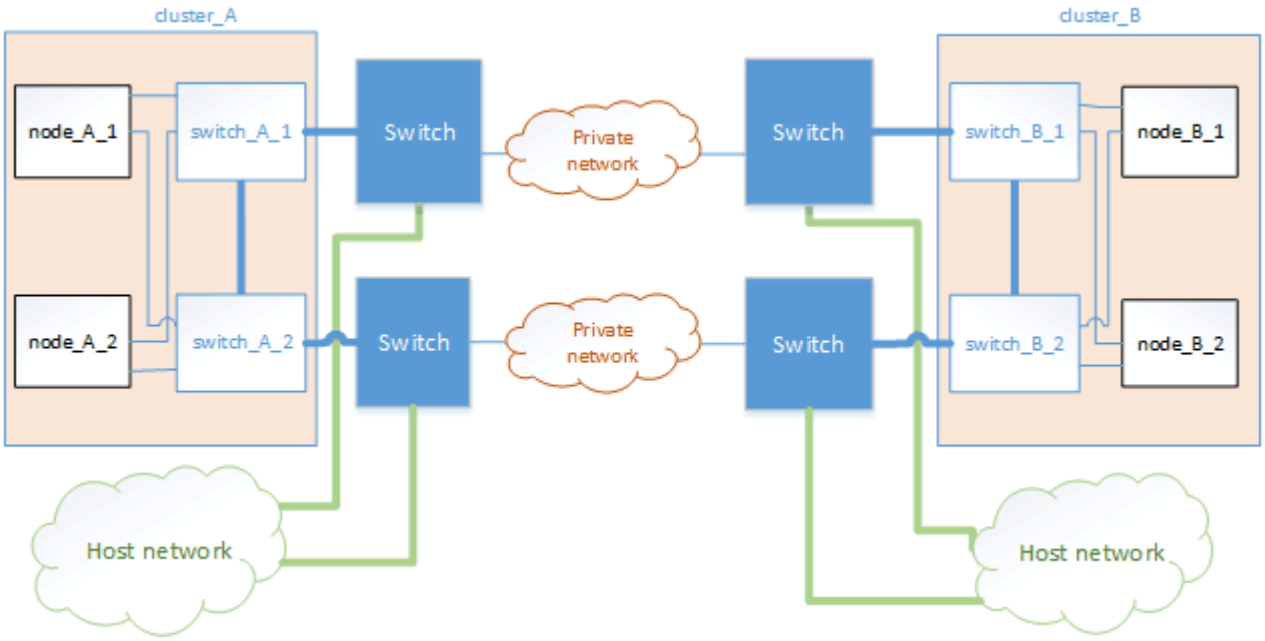
The ISL capacity must meet the minimum requirements. Depending on whether you use xWDM devices or switches a different combination of network configurations might apply.



Shared infrastructure with intermediate networks

In this topology, the MetroCluster IP core switch traffic and the host traffic travel through a network that is not provided by Lenovo. The network infrastructure and the links (including leased direct links) are outside of the MetroCluster configuration. The network can consist of a series of xWDM and switches but unlike the shared configuration with direct ISLs, the links are not direct between the sites. Depending on the infrastructure between the sites, any combination of network configurations is possible. The intermediate infrastructure is represented as a “cloud” (multiple devices can exist between the sites), but it is still under the control of the customer. Capacity through this intermediate infrastructure is not dedicated to the MetroCluster traffic but is shared with other traffic.

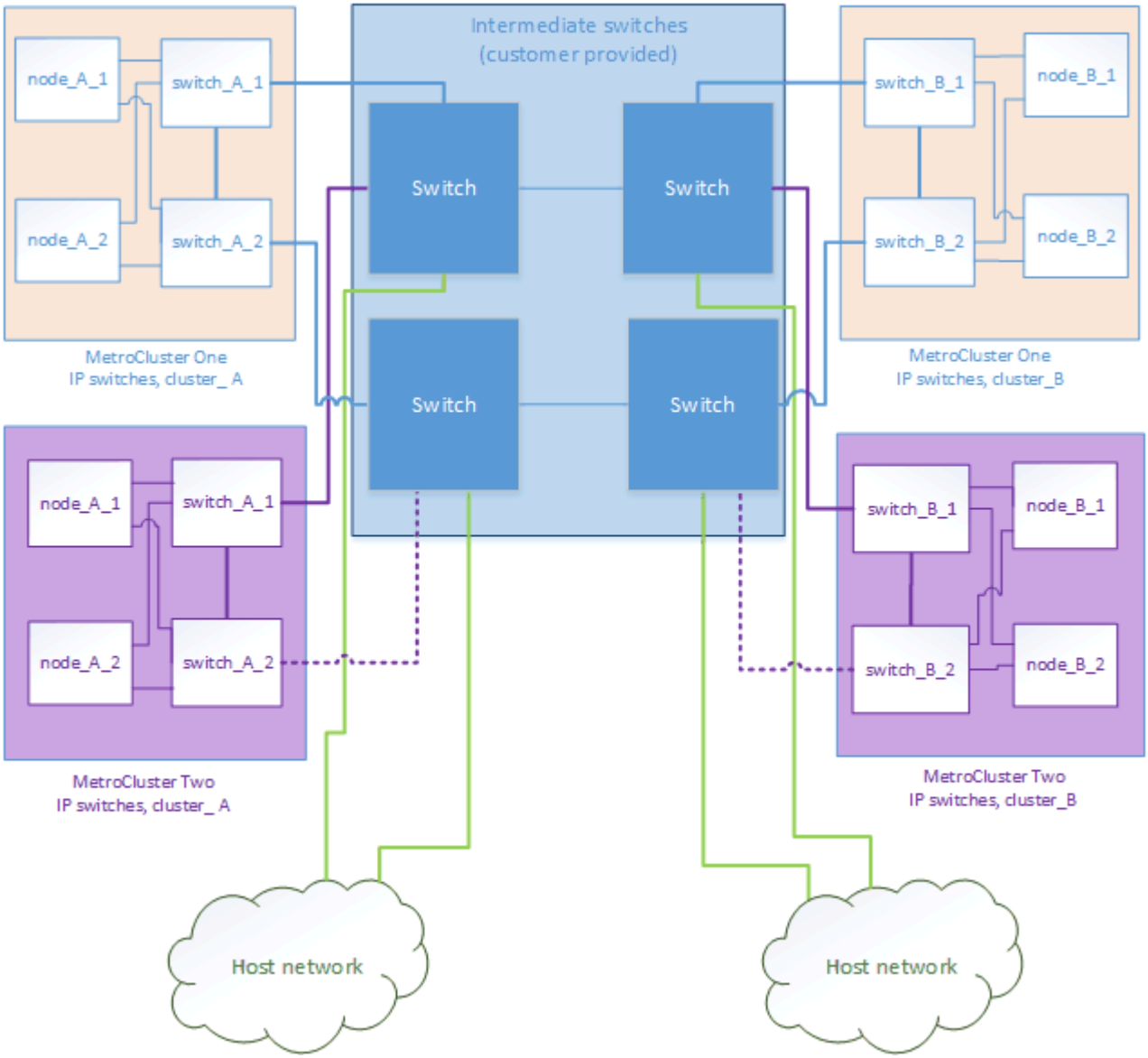
The VLAN and network xWDM or switch configuration must meet the minimum requirements.



Two MetroCluster configurations sharing an intermediate network

In this topology, two separate MetroCluster configurations are sharing the same intermediate network. In the example, MetroCluster one switch_A_1 and MetroCluster two switch_A_1 both connect to the same intermediate switch.

The example is simplified for illustration purposes only:



Two MetroCluster configurations with one connecting directly to the intermediate network

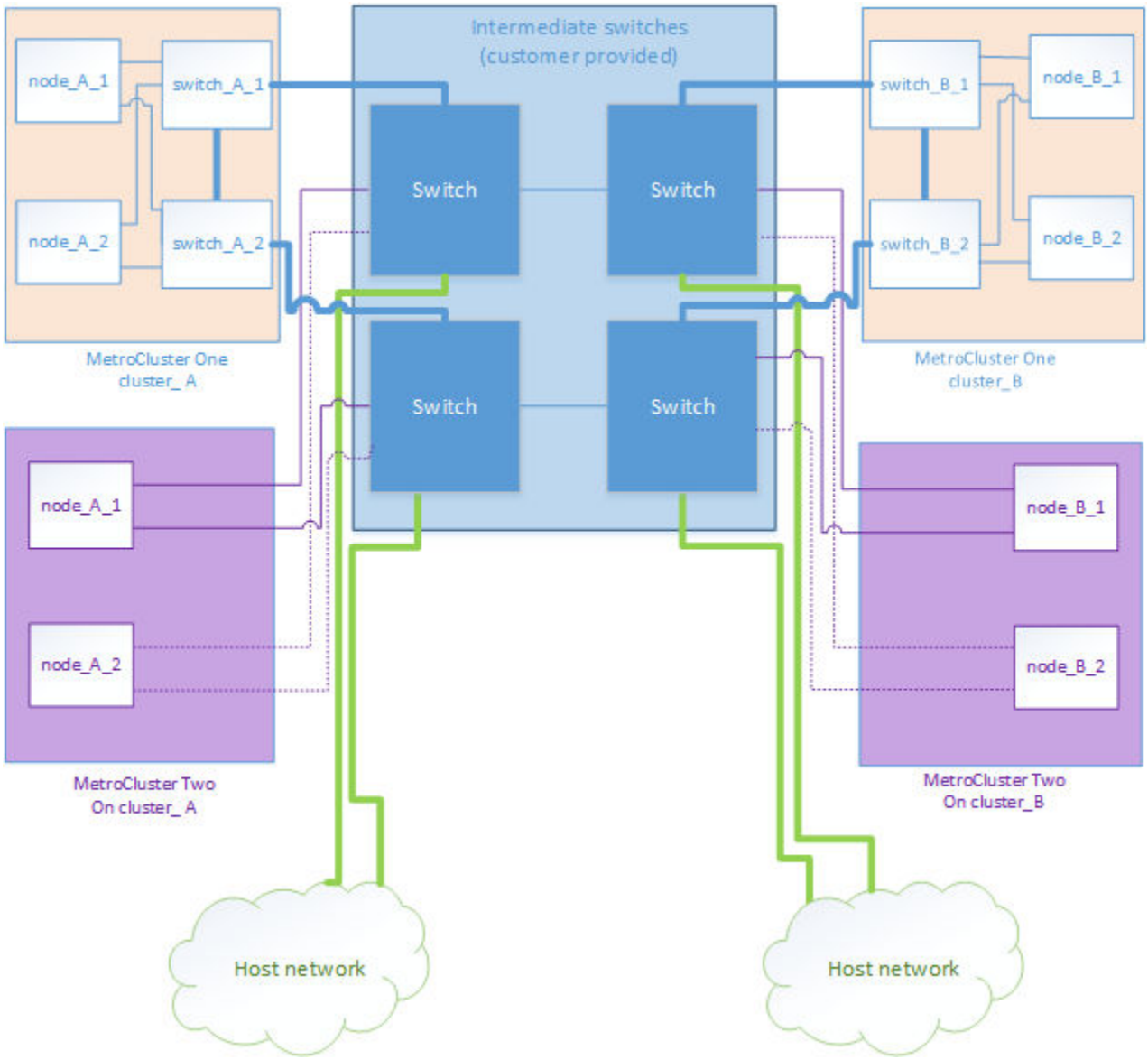
This topology is supported beginning with ONTAP 9.7. Two separate MetroCluster configurations share the same intermediate network and one MetroCluster configuration's nodes is directly connected to the intermediate switch.

MetroCluster One is a MetroCluster configuration using Lenovo validated switches, ONTAP 9.6 and a shared topology. MetroCluster Two is a MetroCluster configuration using Lenovo compliant switches and ONTAP 9.7.

Note: The intermediate switches must be compliant with Lenovo specifications.

"Considerations for using MetroCluster compliant switches" on page 20

The example is simplified for illustration purposes only:



Considerations for using MetroCluster compliant switches

MetroCluster IP switches provided by Lenovo are Lenovo validated. Beginning with ONTAP 9.7, MetroCluster IP configurations can support switches that are not Lenovo validated provided that they are compliant with Lenovo specifications.

General requirements

The requirements show how to configure MetroCluster compliant switches without using reference configuration (RCF) files.

- Only platforms that provide dedicated ports for switchless cluster interconnects are supported. Platforms such as DM5000H and AFA DM5000F are not supported because MetroCluster traffic and MetroCluster interconnect traffic share the same network ports.

  Connecting local cluster connections to a MetroCluster compliant switch is not supported.
- The MetroCluster IP interface can be connected to any switch port that can be configured to meet the requirements.

- The speed of the switch ports must be 25 Gbps for DM7000H and AFA DM7000F platforms, and at least 40 Gbps for all other platforms (40 Gbps or 100 Gbps).
- The ISLs must be 10 Gbps or higher and must be sized appropriately for the load on the MetroCluster configuration.
- The MetroCluster configuration must be connected to two networks. Connecting both the MetroCluster interfaces to the same network or switch is not supported. Each MetroCluster node must be connected to two network switches.
- The network must meet the requirements as outlined in the sections *MetroCluster ISL requirements in shared networks*, *ISL cabling requirements*, and *Required settings on intermediate switches*.
- In MetroCluster IP configurations using open networks, reverting to ONTAP 9.6 or earlier is not supported.
- The MTU of 9216 must be configured on all switches that carry MetroCluster IP traffic.

Switch and cabling requirements

- The switches must support QoS/traffic classification.
- The switches must support explicit congestion notification (ECN).
- The switches must support L4 port-vlan load-balancing policies to preserve order along the path.
- The switches must support L2 Flow Control (L2FC).
- The cables connecting the nodes to the switches must be purchased from Lenovo. The cables we provide must be supported by the switch vendor.

Platform-specific network speeds for MetroCluster compliant switches

The following table provides platform-specific network speeds for MetroCluster compliant switches.

Note: Missing values indicate that the platform is not supported.

Platform	Network Speed (Gbps)
AFA DM7000F	25
AFA DM5000F	-
DM7000H	25
DM5000H	-

Assumptions for the examples

The examples provided are valid for Cisco NX31xx and NX32xx switches. If other switches are used, these commands can be used as guidance, but the commands might be different. If a feature shown in the examples is not available on the switch, this means that the switch does not meet the minimum requirements and cannot be used to deploy a MetroCluster configuration. This is true for any switch that is connecting a MetroCluster configuration and for all switches on the path between those switches.

- The ISL ports are 15 and 16 and operate at a speed of 40 Gbps.
- The VLAN in network 1 is 10 and the VLAN in network 2 is 20. Examples might be shown for one network only.
- The MetroCluster interface is connected to port 9 on each switch and operates at a speed of 100 Gbps.
- The full context of the examples is not set or shown. You might need to enter further configuration information such as the profile, VLAN, or interface, to execute the commands.

Generic switch configuration

A VLAN in each network must be configured. The example shows how to configure a VLAN in network 10.

Example:

```
# vlan 10
```

The load balancing policy should be set so that order is preserved.

Example:

```
# port-channel load-balance src-dst ip-l4port-vlan
```

You must configure the access and class maps, which map the RDMA and iSCSI traffic to the appropriate classes.

All TCP traffic to and from the port 65200 is mapped to the storage (iSCSI) class. All TCP traffic to and from the port 10006 is mapped to the RDMA class.

Example:

```
ip access-list storage
  10 permit tcp any eq 65200 any
  20 permit tcp any any eq 65200
ip access-list rdma
  10 permit tcp any eq 10006 any
  20 permit tcp any any eq 10006

class-map type qos match-all storage
  match access-group name storage
class-map type qos match-all rdma
  match access-group name rdma
```

You must configure the ingress policy. The ingress policy maps the traffic as classified to the different COS groups. In this example, the RDMA traffic is mapped to COS group 5 and iSCSI traffic is mapped to COS group 4.

Example:

```
policy-map type qos MetroClusterIP_Ingress
class rdma
  set dscp 40
  set cos 5
  set qos-group 5
class storage
  set dscp 32
  set cos 4
  set qos-group 4
```

You must configure the egress policy on the switch. The egress policy maps the traffic to the egress queues. In this example, RDMA traffic is mapped to queue 5 and iSCSI traffic is mapped to queue 4.

Example:

```
policy-map type queuing MetroClusterIP_Egress
class type queuing c-out-8q-q7
  priority level 1
class type queuing c-out-8q-q6
  priority level 2
class type queuing c-out-8q-q5
  priority level 3
  random-detect threshold burst-optimized ecn
class type queuing c-out-8q-q4
  priority level 4
```

```
  random-detect threshold burst-optimized ecn
class type queuing c-out-8q-q3
  priority level 5
class type queuing c-out-8q-q2
  priority level 6
class type queuing c-out-8q-q1
  priority level 7
class type queuing c-out-8q-q-default
  bandwidth remaining percent 100
  random-detect threshold burst-optimized ecn
```

You need to configure a switch that has MetroCluster traffic on an ISL but does not connect to any MetroCluster interfaces. In this case, the traffic is already classified and only needs to be mapped to the appropriate queue. In the following example, all of the COS5 traffic is mapped to the class RDMA, and all of the COS4 traffic is mapped to the class iSCSI. Note that this will affect **all** of the COS5 and COS4 traffic, not only the MetroCluster traffic. If you only want to map the MetroCluster traffic, then you must use the above class maps to identify the traffic using the access groups.

Example:

```
class-map type qos match-all rdma
  match cos 5
class-map type qos match-all storage
  match cos 4
```

Configuring the ISLs

You can configure a 'trunk' mode port when setting an allowed VLAN.

There are two commands, one to **set** the allowed VLAN list, and one to **add** to the existing allowed VLAN list.

You can **set** the allowed VLANs as shown in the example.

Example:

```
switchport trunk allowed vlan 10
```

You can **add** a VLAN to the allowed list as shown in the example.

Example:

```
switchport trunk allowed vlan add 10
```

In the example, port-channel 10 is configured for VLAN 10.

Example:

```
interface port-channel10
switchport mode trunk
switchport trunk allowed vlan 10
mtu 9216
service-policy type queuing output MetroClusterIP_Egress
```

The ISL ports should be configured as part of a port-channel and be assigned the egress queues as shown in the example.

Example:

```
interface eth1/15-16
switchport mode trunk
```

```
switchport trunk allowed vlan 10
no lldp transmit
no lldp receive
mtu 9216
channel-group 10 mode active
service-policy type queuing output MetroClusterIP_Egress
no shutdown
```

Configuring the node ports

You might need to configure the node port in breakout mode. In this example, ports 25 and 26 are configured in 4 x 25 Gbps breakout mode.

Example:

```
interface breakout module 1 port 25-26 map 25g-4x
```

You might need to configure the MetroCluster interface port speed. The example shows how to configure the speed to "auto".

Example:

```
speed auto
```

The following example shows how to fix the speed at 40 Gbps.

Example:

```
speed 40000
```

You might need to configure the interface. In the following example, the interface speed is set to "auto".

The port is in access mode in VLAN 10, MTU is set to 9216 and the MetroCluster ingress policy is assigned.

Example:

```
interface eth1/9
description MetroCluster-IP Node Port
speed auto
switchport access vlan 10
spanning-tree port type edge
spanning-tree bpduguard enable
mtu 9216
flowcontrol receive on
flowcontrol send on
service-policy type qos input MetroClusterIP_Ingress
no shutdown
```

On 25-Gbps ports, the FEC setting might need to be set to "off" as shown in the example.

Example:

```
fec off
```

Note: You must always run this command **after** the interface is configured. A transceiver module might need to be inserted for the command to work.

Considerations for using TDM/xWDM and encryption equipment with MetroCluster IP configurations

You should be aware of certain considerations for using multiplexing equipment in the MetroCluster IP configuration.

These considerations apply only to direct, dedicated MetroCluster back-end links and switches, not links shared with non-MetroCluster traffic.

The Lenovo Press provides some notes about the requirements that TDM/xWDM equipment must meet to work with a MetroCluster IP configuration.

Lenovo Press

Using encryption on WDM or external encryption devices

When using encryption on WDM devices in the MetroCluster IP configuration, your environment must meet the following requirements:

- The external encryption devices or DWDM equipment must have been certified by the vendor with the switch in question.

  The certification should cover the operating mode (such as trunking and encryption).
- The overall end-to-end latency and jitter, including the encryption, cannot be above the maximum stated in the Lenovo Storage Interoperation Center (LSIC) or in this document.

SFP considerations

Any SFPs or QSFPs supported by the equipment vendor are supported for the MetroCluster ISLs. SFPs and QSFPs can be acquired from Lenovo or the equipment vendor.

Considerations for ISLs

The ISLs on one fabric should all be the same speed and length.

The ISLs on one fabric should all have the same topology. For example, they should all be direct links, or if the configuration uses WDM, then they should all use WDM.

If you are sharing ISLs with a non-MetroCluster network, you must follow the guidelines in the section “Considerations for sharing private layer 2 networks” on page 11.

Considerations for firewall usage at MetroCluster sites

If you are using a firewall at a MetroCluster site, you must ensure access for certain required ports.

The following table shows TCP/UDP port usage in an external firewall positioned between two MetroCluster sites.

Traffic type	Port/services
Cluster peering	11104 / TCP 11105 / TCP
ONTAP System Manager	443 / TCP

Traffic type	Port/services
MetroCluster IP intercluster LIFs	65200 / TCP 10006 / TCP and UDP
Hardware assist	4444 / TCP

Preconfigured settings for new MetroCluster systems from the factory

New MetroCluster nodes are preconfigured with a root aggregate. Additional hardware and software settings are configured using the detailed procedures provided in this guide.

Hardware racking and cabling

Depending on the configuration you ordered, you might need to rack the systems and complete the cabling.

Software configuration of the MetroCluster configuration

Nodes received with the new MetroCluster configuration are preconfigured with a single root aggregate. Additional configuration must be performed using the detailed procedures provided in this guide.

Hardware setup checklist

You need to know which hardware setup steps were completed at the factory and which steps you need to complete at each MetroCluster site.

Step	Completed at factory	Completed by you
Mount components in one or more cabinets.	Yes	No
Position cabinets in the desired location.	No	Yes Position them in the original order so that the supplied cables are long enough.
Connect multiple cabinets to each other, if applicable.	No	Yes Use the cabinet interconnect kit if it is included in the order. The kit box is labeled.
Secure the cabinets to the floor, if applicable.	No	Yes Use the universal bolt-down kit if it is included in the order. The kit box is labeled.
Cable the components within the cabinet.	Yes Cables 5 meters and longer are removed for shipping and placed in the accessories box.	No
Connect the cables between cabinets, if applicable.	No	Yes Cables are in the accessories box.

Step	Completed at factory	Completed by you
Connect management cables to the customer's network.	No	Yes Connect them directly or through the CN1601 management switches, if present. **Attention:** To avoid address conflicts, do not connect management ports to the customer's network until after you change the default IP addresses to the customer's values.
Connect console ports to the customer's terminal server, if applicable.	No	Yes
Connect the customer's data cables to the cluster.	No	Yes
Connect the cabinets to power and power on the components.	No	Yes Power them on in the following order: 1. PDUs 2. Disk shelves 3. Nodes
Assign IP addresses to the management ports of the cluster switches and to the management ports of the management switches, if present.	No	Yes Connect to the serial console port of each switch and log in with user name “admin” with no password. Suggested management addresses are 10.10.10.81, 10.10.10.82, 10.10.10.83, and 10.10.10.84.

Chapter 3. Configuring the MetroCluster hardware components

The MetroCluster components must be physically installed, cabled, and configured at both geographic sites.

Parts of a MetroCluster IP configuration

As you plan your MetroCluster IP configuration, you should understand the hardware components and how they interconnect.

Key hardware elements

A MetroCluster IP configuration includes the following key hardware elements:

- Storage controllers

  The storage controllers are configured as two two-node clusters.
- IP network

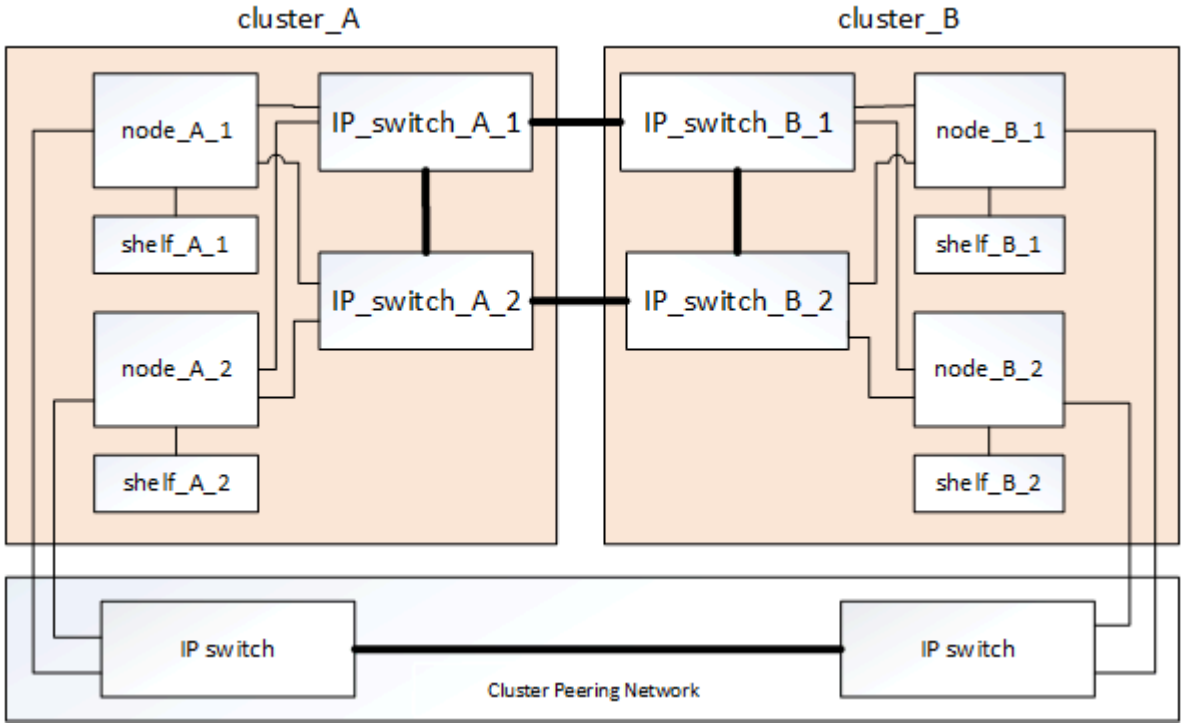
  This back-end IP network provides connectivity for two distinct uses:

  – Standard cluster connectivity for intra-cluster communications.

    This is the same cluster switch functionality used in non-MetroCluster switched ONTAP clusters.

  – MetroCluster back-end connectivity for replication of storage data and non-volatile cache.
- Cluster peering network

  The cluster peering network provides connectivity for mirroring of the cluster configuration, which includes storage virtual machine (SVM) configuration. The configuration of all of the SVMs on one cluster is mirrored to the partner cluster.



Disaster Recovery (DR) groups

A MetroCluster IP configuration consists of one DR group of four nodes.

The following illustration shows the organization of nodes in a four-node MetroCluster configuration:

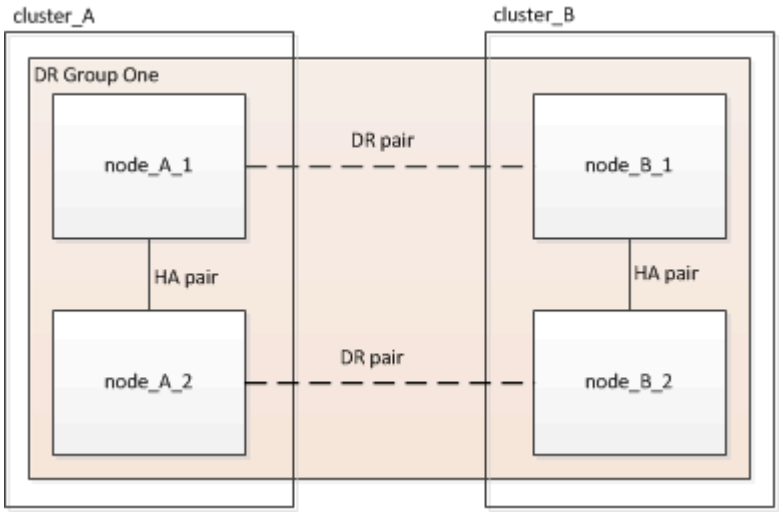


Illustration of the local HA pairs in a MetroCluster configuration

Each MetroCluster site consists of storage controllers configured as an HA pair. This allows local redundancy so that if one storage controller fails, its local HA partner can take over. Such failures can be handled without a MetroCluster switchover operation.

Local HA failover and giveback operations are performed with the `storage failover` commands, in the same manner as a non-MetroCluster configuration.

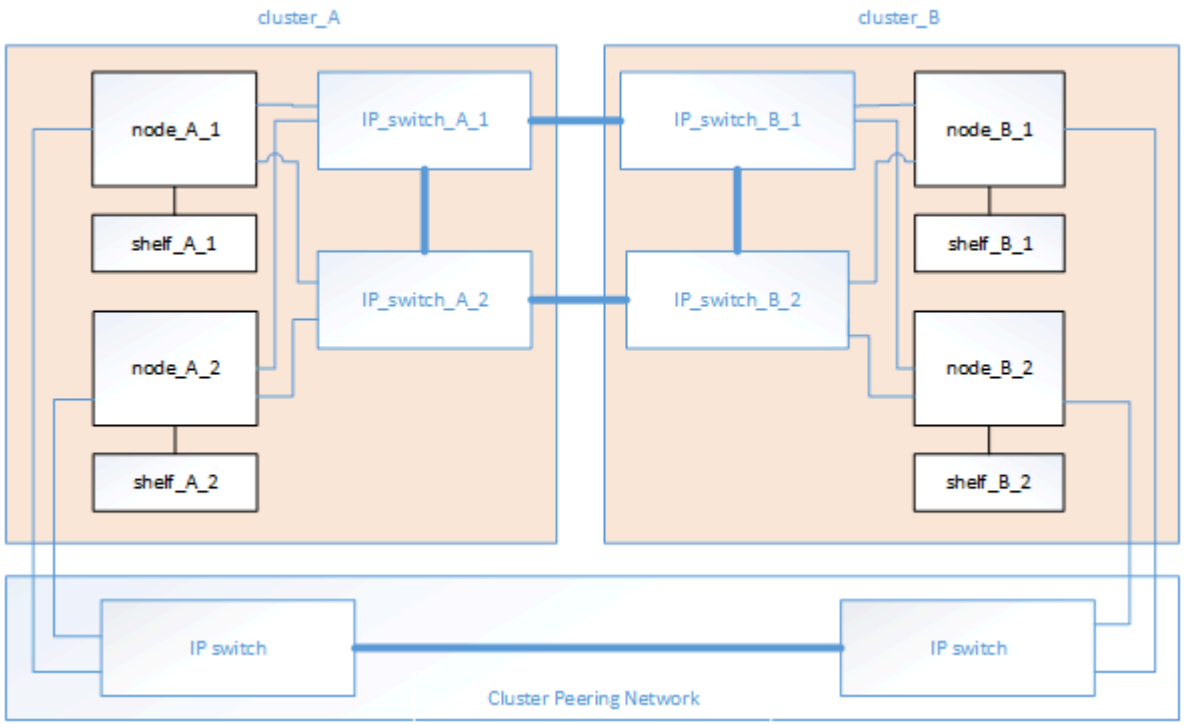
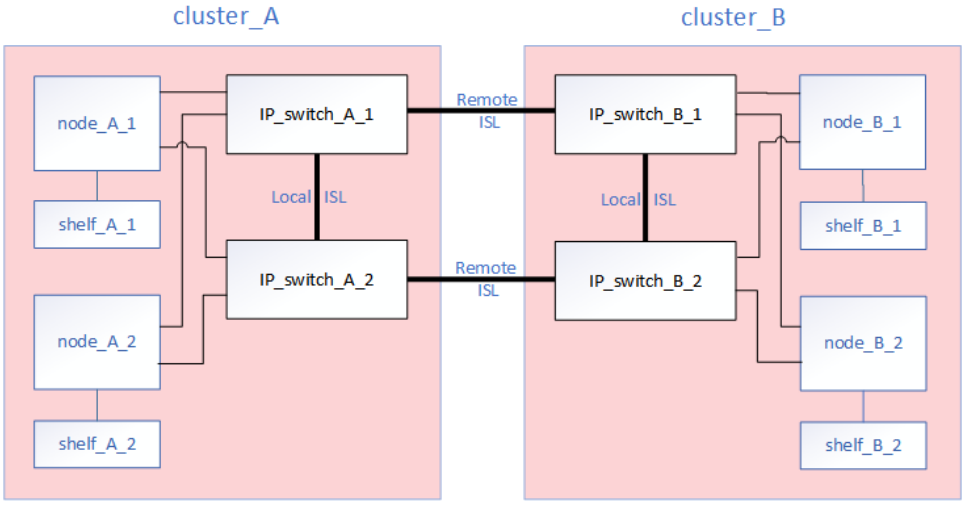


Illustration of the MetroCluster IP and cluster interconnect network

ONTAP clusters typically include a cluster interconnect network for traffic between the nodes in the cluster. In MetroCluster IP configurations, this network is also used for carrying data replication traffic between the MetroCluster sites.



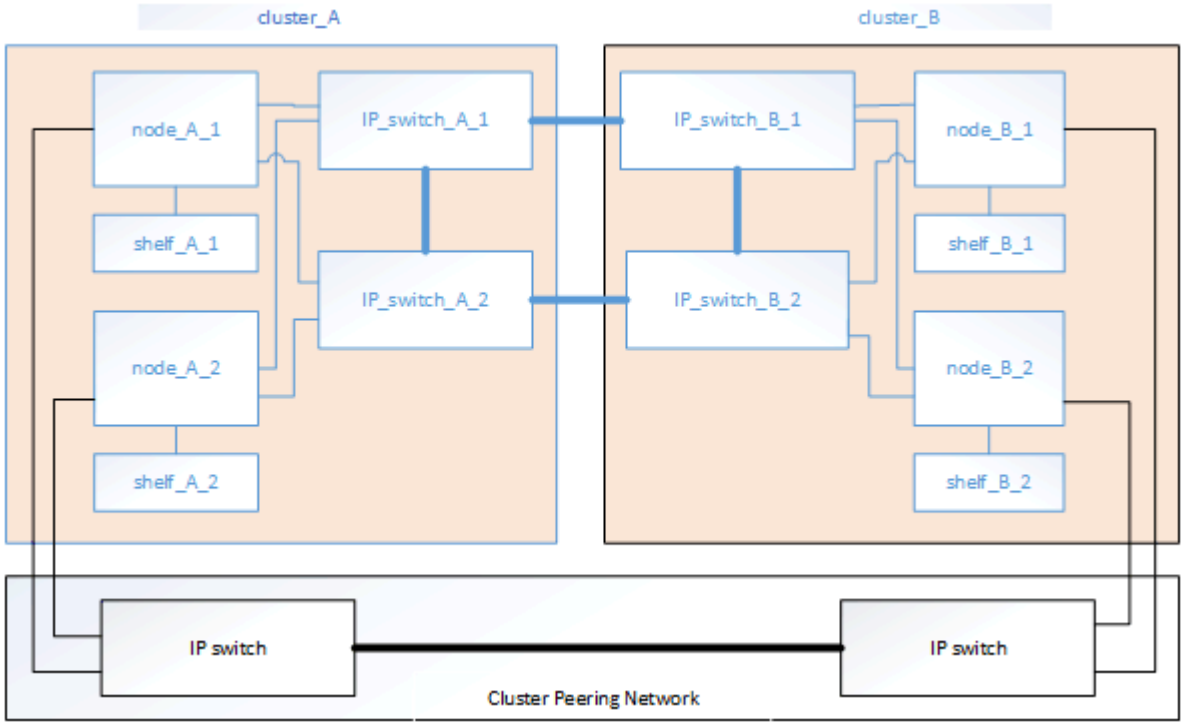
Each node in the MetroCluster IP configuration has specialized LIFs for connection to the back-end IP network:

- Two MetroCluster IP interfaces
- One intercluster LIF

Illustration of the cluster peering network

The two clusters in the MetroCluster configuration are peered through a customer-provided cluster peering network. Cluster peering supports the synchronous mirroring of storage virtual machines (SVMs, formerly known as Vservers) between the sites.

Intercluster LIFs must be configured on each node in the MetroCluster configuration, and the clusters must be configured for peering. The ports with the intercluster LIFs are connected to the customer-provided cluster peering network. Replication of the SVM configuration is carried out over this network through the Configuration Replication Service.



Required MetroCluster IP components and naming conventions

When planning your MetroCluster IP configuration, you must understand the required and supported hardware and software components. For convenience and clarity, you should also understand the naming conventions used for components in examples throughout the documentation.

Supported software and hardware

The hardware and software must be supported for the MetroCluster IP configuration.

When using AFA systems, all controller modules in the MetroCluster configuration must be configured as AFA systems.

Hardware redundancy requirements in a MetroCluster IP configuration

Because of the hardware redundancy in the MetroCluster IP configuration, there are two of each component at each site. The sites are arbitrarily assigned the letters A and B, and the individual components are arbitrarily assigned the numbers 1 and 2.

ONTAP cluster requirements in a MetroCluster IP configuration

MetroCluster IP configurations require two ONTAP clusters, one at each MetroCluster site.

Naming must be unique within the MetroCluster configuration.

Example names:

- Site A: cluster_A
- Site B: cluster_B

IP switch requirements in a MetroCluster IP configuration

MetroCluster IP configurations require four IP switches. The four switches form two switch storage fabrics that provide the ISL between each of the clusters in the MetroCluster IP configuration.

The IP switches also provide intracluster communication among the controller modules in each cluster.

Naming must be unique within the MetroCluster configuration.

Example names:

- Site A: cluster_A
  - IP_switch_A_1
  - IP_switch_A_2
- Site B: cluster_B
  - IP_switch_B_1
  - IP_switch_B_2

Controller module requirements in a MetroCluster IP configuration

MetroCluster IP configurations require four controller modules.

The controller modules at each site form an HA pair. Each controller module has a DR partner at the other site.

Each controller module must be running the same ONTAP version. Supported platform models depend on the ONTAP version:

- Starting with ONTAP 9.5, new MetroCluster IP installations on Hybrid systems are supported.
- Starting with ONTAP 9.4, controller modules configured for ADP are supported.

Example names:

- Site A: cluster_A
  - controller_A_1
  - controller_A_2
- Site B: cluster_B
  - controller_B_1
  - controller_B_2

Gigabit Ethernet adapter requirements in a MetroCluster IP configuration

MetroCluster IP configurations use a 40/100 Gbps or 10/25 Gbps Ethernet adapter for the IP interfaces to the IP switches used for the MetroCluster IP fabric.

Platform model	Required Gigabit Ethernet adapter	Required slot for adapter	Ports
AFA DM7000F and DM7000H	X1116A	Slot 1	e1a, e1b
AFA DM5000F, and DM5000H	Onboard ports	Slot 0	e0a, e0b

Pool and drive requirements (minimum supported)

Eight SAS disk shelves are recommended (four shelves at each site) to allow disk ownership on a per-shelf basis.

The MetroCluster configuration requires the minimum configuration at each site:

- Each node has at least one local pool and one remote pool at the site.

  For example, in a four-node MetroCluster configuration with two nodes at each site, four pools are required at each site.

MetroCluster configurations support RAID-DP.

Drive location considerations for half-shelf configurations

For correct auto-assignment of drives when using shelves that are half populated (12 drives in a 24-drive shelf), drives should be located in slots 0-5 and 18-23.

Installing and cabling MetroCluster components

The storage controllers must be cabled to the IP switches and the ISLs must be cabled to link the MetroCluster sites. The storage controllers must also be cabled to the storage, to each other, and to the data and management networks.

Racking the hardware components

If you have not received the equipment already installed in cabinets, you must rack the components.

About this task

This task must be performed on both MetroCluster sites.

Step 1. Plan out the positioning of the MetroCluster components.

The rack space depends on the platform model of the controller modules, the switch types, and the number of disk shelf stacks in your configuration.

Step 2. Properly ground yourself.

Step 3. Install the controller modules in the rack or cabinet.

Step 4. Install the IP switches in the rack or cabinet.

Step 5. Install the disk shelves, power them on, and then set the shelf IDs.

- You must power-cycle each disk shelf.
- Shelf IDs must be unique for each SAS disk shelf within each MetroCluster DR group (including both sites).

Cabling the IP switches

You must cable each IP switch to the local controller modules and to the ISLs.

About this task

- This task must be repeated for each switch in the MetroCluster configuration.
- The controller module Ethernet port usage depends on the model of the controller module.

Step 1. Cable the switch and node ports, using the table for your switch model and platform.

- “Port assignments for AFA DM7000F or DM7000H systems with Cisco 3232C IP switches” on page 35
- “Port assignments for AFA DM5000F and DM5000H systems with Cisco 3132Q-V IP switches” on page 36

- "Port assignments for AFA DM7000F, DM7000H, AFA DM5000F, and DM5000H systems with Broadcom supported BES-53248 IP switches" on page 38
- "Port assignments for ISLs" on page 40

Port assignments for AFA DM7000F or DM7000H systems with Cisco 3232C IP switches

The port usage in a MetroCluster IP configuration depends on the switch model and platform type.

The following tables show the port usage for site A. The same cabling is used for site B.

Note: The switches cannot be configured with ports of different speeds (for example, a mix of 100 Gbps ports and 40 Gbps ports).

ISL notes

ISL connections can be done either with the 40/100 Gbps ports (15—20) or using breakout cables with ports 21—24.

Switch port usage

A separate breakout cable is required for each MetroCluster IP interconnect connection (two breakout cables for each node).

The switch ports that support breakout mode are divided into four logical ports. For example, physical port 25 is logically split into four ports:

- 25/1
- 25/2
- 25/3
- 25/4

IP_switch_A_1 Local interconnect connections				
Switch port	Port speed	Controller module port		Usage
		Node	AFA DM7000F **and** DM7000H **systems**	
25/1	100 Gbps switch port is connected to a 25 Gbps port on controller using 4x25 Gbps breakout cable	node_A_1	e1a	MetroCluster IP interconnect
26/1		node_A_2	e1a	MetroCluster IP interconnect
27/1	-	-	-	Unused
28/1	-	-	-	Unused
29/1	100 Gbps switch port is connected to a 10 Gbps port on controller using 4x10 Gbps breakout cable	node_A_1	e0a	Local cluster interconnect
30/1		node_A_2	e0a	Local cluster interconnect

IP_switch_A_2 Local interconnect connections				
Switch port	Port speed	Controller module port		Usage
		Node	AFA DM7000F **and** DM7000H **systems**	
25/1	100 Gbps switch port is connected to a 25 Gbps port on controller using 4x25 Gbps breakout cable	node_A_1	e1b	MetroCluster IP interconnect
26/1		node_A_2	e1b	MetroCluster IP interconnect
27/1	-	-	-	Unused
28/1	-	-	-	Unused
29/1	100 Gbps switch port is connected to a 10 Gbps port on controller using 4x10 Gbps breakout cable	node_A_1	e0b	Local cluster interconnect
30/1		node_A_2	e0b	Local cluster interconnect

Port assignments for AFA DM5000F and DM5000H systems with Cisco 3132Q-V IP switches

The port usage in a MetroCluster IP configuration depends on the switch model and platform type.

The following tables show the port usage for site A. The same cabling is used for site B.

Switch port usage

These tables do not show unused ports or the ISL ports.

IP_switch_A_1 local connections			
Port	Port speed	Controller module port	Usage
	Cisco 3132Q-V	Cisco 3132Q-V	
7	40 Gbps	-	Local cluster interconnect
8	40 Gbps	-	Local cluster interconnect
9/1	-	e0a	Shared cluster / MetroCluster IP interconnect
9/2	-	-	-
9/3		-	-
9/4	10 Gbps	-	-
10/1	-	e0a	Shared cluster / MetroCluster IP interconnect
10/2	-	-	-
10/3	-	-	-
10/4	-	-	-

IP_switch_A_2 local connections				
Port	Port speed	Controller module port		Usage
	Cisco 3132Q-V	100 Gbps	40 Gbps	
7	40 Gbps	-	-	Local cluster interconnect
8	40 Gbps	-	-	Local cluster interconnect
9/1	-	node_A_1	e0b	Shared cluster / MetroCluster IP interconnect
9/2	-	-	-	-
9/3		-	-	-
9/4	10 Gbps	-	-	-
10/1	-	node_A_2	e0b	Shared cluster / MetroCluster IP interconnect
10/2	-	-	-	-
10/3	-	-	-	-
10/4	-	-	-	-

Port assignments for AFA DM5000F and DM5000H systems with Cisco 3232C IP switches

The port usage in a MetroCluster IP configuration depends on the switch model and platform type.

The following tables show the port usage for site A. The same cabling is used for site B.

Switch port usage

These tables do not show unused ports or the ISL ports.

IP_switch_A_1 local connections			
Port	Port speed	Controller module port	Usage
7	100 Gbps	-	Local cluster interconnect
8	100 Gbps	-	Local cluster interconnect
9/1	100 Gbps switch port connected to 10 Gbps port on controller using 4x10 Gbps breakout cable	node_A_1	Shared cluster / MetroCluster IP interconnect
9/2	-	-	-
9/3	-	-	-
9/4	100 Gbps switch port connected to 10 Gbps port on controller using 4x10 Gbps breakout cable	-	-

IP_switch_A_1 local connections			
Port	**Port speed**	**Controller module port**	**Usage**
10/1	100 Gbps switch port connected to 10 Gbps port on controller using 4x10 Gbps breakout cable	node_A_2	Shared cluster / MetroCluster IP interconnect
10/2	-	-	-
10/3	-	-	-
10/4	-	-	-

IP_switch_A_2 local connections				
		Controller module port		
Port	**Port speed**	**100 Gbps**	**40 Gbps**	**Usage**
7	100 Gbps	-	-	Local cluster interconnect
8	100 Gbps	-	-	Local cluster interconnect
9/1	100 Gbps switch port connected to 10 Gbps port on controller using 4x10 Gbps breakout cable	node_A_1	e0b	Shared cluster / MetroCluster IP interconnect
9/2	-	-	-	-
9/3	-	-	-	-
9/4	100 Gbps	-	-	-
10/1	100 Gbps switch port connected to 10 Gbps port on controller using 4x10 Gbps breakout cable	node_A_2	e0b	Shared cluster / MetroCluster IP interconnect
10/2	-	-	-	-
10/3	-	-	-	-
10/4	-	-	-	-

Port assignments for AFA DM7000F, DM7000H, AFA DM5000F, and DM5000H systems with Broadcom supported BES-53248 IP switches

The port usage in a MetroCluster IP configuration depends on the switch model and platform type.

The following tables show the port usage for site A. The same cabling is used for site B.

Only **one** MetroCluster configuration is supported on the switches.

Note: The switches cannot be configured with ports of different speeds (for example, a mix of 25 Gbps ports and 10 Gbps ports).

Switch port usage for AFA DM5000F, DM5000H, AFA DM7000F, and DM7000H systems

IP_switch_A_1 local connections					
		Controller module port			
Port	Port speed	Node	AFA DM5000F **and** DM5000H **systems**	AFA DM7000F **and** DM7000H **systems**	Usage
1	10 Gbps	node_A_1	-	e0a	Local cluster interconnect
2	10 Gbps	node_A_2	-	e0a	Local cluster interconnect
3		-	-	-	Local cluster interconnect
4		-	-	-	Local cluster interconnect
5	25 Gbps	node_A_1	-	e1a	MetroCluster IP interconnect
6	25 Gbps	node_A_2	-	e1a	MetroCluster IP interconnect
7		-	-	-	MetroCluster IP interconnect
8		-	-	-	MetroCluster IP interconnect
9	10 Gbps	node_A_1	e0a	-	Shared cluster / MetroCluster IP interconnect
10	10 Gbps	node_A_2	e0a	-	Shared cluster / MetroCluster IP interconnect
11				-	Shared cluster / MetroCluster IP interconnect
12				-	Shared cluster / MetroCluster IP interconnect
-			-	-	
55	100 Gbps	IP_switch_B_1	-	-	Local interconnect
56	100 Gbps	IP_switch_B_1	-	-	Local interconnect

IP_switch_A_2 local connections					
		Controller module port			
Port	Port speed	Node	AFA DM5000F **and** DM5000H **systems**	AFA DM7000F **and** DM7000H **systems**	Usage
1	10 Gbps	node_A_1	-	e0b	Local cluster interconnect
2	10 Gbps	node_A_2	-	e0b	Local cluster interconnect

IP_switch_A_2 local connections					
		Controller module port			
Port	Port speed	Node	AFA DM5000F **and** DM5000H **systems**	AFA DM7000F **and** DM7000H **systems**	Usage
3	-	-	-	-	Local cluster interconnect
4	-	-	-	-	Local cluster interconnect
5	25 Gbps	node_A_1	-	e1b	MetroCluster IP interconnect
6	25 Gbps	node_A_2	-	e1b	MetroCluster IP interconnect
7	-	-	-	-	MetroCluster IP interconnect
8	-	-	-	-	MetroCluster IP interconnect
9	10 Gbps	node_A_1	e0b	-	Shared cluster / MetroCluster IP interconnect
10	10 Gbps	node_A_2	e0b	-	Shared cluster / MetroCluster IP interconnect
11	-	-	-	-	Shared cluster / MetroCluster IP interconnect
12	-	-	-	-	Shared cluster / MetroCluster IP interconnect
-	-	-	-	-	
55	100 Gbps	IP_switch_B_2	-	-	Local interconnect
56	100 Gbps	IP_switch_B_2	-	-	Local interconnect

Port assignments for ISLs

The ISL port usage in a MetroCluster IP configuration is the same for all switch models.

ISL usage notes for Cisco 3132Q-V and Cisco 3232C switches

ISL connections can be done either with the 40 Gbps (Cisco 3132Q-V) or 40/100 Gbps (Cisco 3232C or 9336C) using ports 15—20, or using breakout cables with ports 21—24. A separate breakout cable is required for each ISL connection. The switch ports that support breakout mode are divided into four logical ports. For example, physical port 21 is logically split into four ports:

- 21/1
- 21/2
- 21/3
- 21/4

ISL port usage

When using only two ISL ports per switch (Broadcom or Cisco), you can use any four ports as long as they are the same speed.

IP_switch_A_1 ISL connections				
Switch port	**Port speed for Cisco 3232C switches**	**Port speed for Cisco 3132Q-V switches**	**Connects to switch**	**Usage**
7	100 Gbps	40 Gbps	IP_switch_A_2	Local cluster ISL
8	100 Gbps	40 Gbps	IP_switch_A_2	Local cluster ISL
9-14	-	-	-	-
15	40/100 Gbps	40 Gbps	IP_switch_B_1	MetroCluster ISL
16	40/100 Gbps	40 Gbps	IP_switch_B_1	MetroCluster ISL
17	40/100 Gbps	40 Gbps	IP_switch_B_1	MetroCluster ISL
18	40/100 Gbps	40 Gbps	IP_switch_B_1	MetroCluster ISL
19	40/100 Gbps	40 Gbps	IP_switch_B_1	MetroCluster ISL
20	40/100 Gbps	40 Gbps	IP_switch_B_1	MetroCluster ISL
21/1-4	10 Gbps using 4x10 Gbps breakout cables	10 Gbps using 4x10 Gbps breakout cables	IP_switch_B_1	MetroCluster ISL (switch in breakout mode).
22/1-4			IP_switch_B_1	
23/1-4			IP_switch_B_1	
24/1-4			IP_switch_B_1	

IP_switch_A_2 ISL connections				
Switch port	**Port Speed**	**Port speed**	**Connects to switch**	**Usage**
7	100 Gbps	40 Gbps	IP_switch_A_1	Local cluster ISL
8	100 Gbps	40 Gbps	IP_switch_A_1	Local cluster ISL
9—14	-	-	-	-
15	40/100 Gbps	40 Gbps	IP_switch_B_2	MetroCluster ISL
16	40/100 Gbps	40 Gbps	IP_switch_B_2	MetroCluster ISL
17	40/100 Gbps	40 Gbps	IP_switch_B_2	MetroCluster ISL
18	40/100 Gbps	40 Gbps	IP_switch_B_2	MetroCluster ISL
19	40/100 Gbps	40 Gbps	IP_switch_B_2	MetroCluster ISL
20	40/100 Gbps	40 Gbps	IP_switch_B_2	MetroCluster ISL
21/1-4	10 Gbps using 4x10 Gbps breakout cables	10 Gbps using 4x10 Gbps breakout cables	IP_switch_B_2	MetroCluster ISL (switch in breakout mode).
22/1-4			IP_switch_B_2	
23/1-4			IP_switch_B_2	
24/1-4			IP_switch_B_2	

ISL port usage for Broadcom supported BES-53248 IP switches

IP_switch_A_1 ISL connections			
Switch port	**Port speed**	**Connects to switch**	**Usage**
13	10 / 25 Gbps	IP_switch_B_1	MetroCluster ISL
14	10 / 25 Gbps	IP_switch_B_1	MetroCluster ISL
15	10 / 25 Gbps	IP_switch_B_1	MetroCluster ISL
16	10 / 25 Gbps	IP_switch_B_1	MetroCluster ISL
-	-	-	-
55	100 Gbps	IP_switch_A_2	Local cluster ISL
56	100 Gbps	IP_switch_A_2	Local cluster ISL

IP_switch_A_2 ISL connections			
Switch port	**Port speed**	**Connects to switch**	**Usage**
13	10 / 25 Gbps	IP_switch_B_2	MetroCluster ISL
14	10 / 25 Gbps	IP_switch_B_2	MetroCluster ISL
15	10 / 25 Gbps	IP_switch_B_2	MetroCluster ISL
16	10 / 25 Gbps	IP_switch_B_2	MetroCluster ISL
-	-	-	-
55	100 Gbps	IP_switch_A_1	Local cluster ISL
56	100 Gbps	IP_switch_A_1	Local cluster ISL

Cabling the cluster peering connections

You must cable the controller module ports used for cluster peering so that they have connectivity with the cluster on the partner site.

About this task

This task must be performed on each controller module in the MetroCluster configuration.

At least two ports on each controller module should be used for cluster peering.

The recommended minimum bandwidth for the ports and network connectivity is 1 GbE.

Step 1. Identify and cable at least two ports for cluster peering and verify they have network connectivity with the partner cluster.

Cluster peering can be done on dedicated ports or on data ports. Using dedicated ports provides higher throughput for the cluster peering traffic.

Cluster and SVM peering express configuration

Cabling the management and data connections

You must cable the management and data ports on each storage controller to the site networks.

About this task

This task must be repeated for each new controller at both MetroCluster sites.

You can connect the controller and cluster switch management ports to existing switches in your network or to new dedicated network switches such as Lenovo CN1601 cluster management switches.

Step 1. Cable the controller's management and data ports to the management and data networks at the local site.

Configuring the IP switches

You must configure the IP switches for use as the cluster interconnect and for back-end MetroCluster IP connectivity. The procedure you use depends on the switch model.

Configuring Broadcom IP switches

You must configure the Broadcom IP switches for use as the cluster interconnect and for backend MetroCluster IP connectivity.

Resetting the Broadcom IP switch to factory defaults

Before installing a new switch software version and RCFs, you must erase the Broadcom switch settings and perform basic configuration.

About this task

- You must repeat these steps on each of the IP switches in the MetroCluster IP configuration.
- You must be connected to the switch using the serial console.
- This task resets the configuration of the management network.

Step 1. Change to the elevated command prompt (#):

```
enable
```

Example

```
(Routing)> enable
(Routing) #
```

Step 2. Erase the startup configuration:

```
erase startup-config
```

Example

```
(Routing) #erase startup-config
Are you sure you want to clear the configuration? (y/n) y


(Routing) #
```

This command does not erase the banner.

Step 3. Reboot the switch:

```
reload
```

Example

```
(IP_switch_A_1) #reload
```

```
Are you sure you would like to reset the system? (y/n) y
```

Note: If the system asks whether to save the unsaved or changed configuration before reloading the switch, select **No**.

Step 4. Wait for the switch to reload, and then log in to the switch.

The default user is “admin”, and no password is set. A prompt similar to the following is displayed:

Example

```
(Routing)>
```

Step 5. Change to the elevated command prompt:
`enable`

Example

```
Routing)> enable
(Routing) #
```

Step 6. Set the serviceport protocol to **none**:
`serviceport protocol none`

Example

```
(Routing) #serviceport protocol none
Changing protocol mode will reset ip configuration.
Are you sure you want to continue? (y/n) y

(Routing) #
```

Step 7. Assign the IP address to the service port:
`serviceport ip` *ip-address netmask gateway*

Example

The following example shows a service port assigned IP address 10.10.10.10 with subnet 255.255.255.0 and gateway 10.10.10.1:

```
(Routing) #serviceport ip 10.10.10.10 255.255.255.0 10.10.10.1
```

Step 8. Verify that the service port is correctly configured:
`show serviceport`

Example

The following example shows that the port is up and the correct addresses have been assigned:

```
(Routing) #show serviceport

Interface Status.............................. Up
IP Address..................................... 10.10.10.10
Subnet Mask.................................... 255.255.255.0
Default Gateway................................ 10.10.10.1
IPv6 Administrative Mode....................... Enabled
IPv6 Prefix is ................................ fe80::dac4:97ff:fe56:87d7/64
IPv6 Default Router............................ fe80::222:bdff:fef8:19ff
Configured IPv4 Protocol....................... None
Configured IPv6 Protocol....................... None
IPv6 AutoConfig Mode........................... Disabled
```

```
Burned In MAC Address..... D8:C4:97:56:87:D7

(Routing) #
```

Step 9. If desired, configure the SSH server.

Note: The RCF file disables the Telnet protocol. If you do not configure the SSH server, you can only access the bridge using the serial port connection.

a. Generate RSA keys.
```
(Routing) #configure
(Routing) (Config)#crypto key generate rsa
```

b. Generate DSA keys.
```
(Routing) #configure
(Routing) (Config)#crypto key generate dsa
```

c. Enable the SSH server.If necessary, exit the configuration context.
```
(Routing) (Config)#end
(Routing) #ip ssh server enable
```

Note: If keys already exist, then you might be asked to overwrite them.

Step 10. If desired, configure the domain and name server:
```
configure
```

Example

The following example shows the `ip domain` and `ip name server` commands:

```
(Routing) # configure
(Routing) (Config)#ip domain name labs.Lenovo.com
(Routing) (Config)#ip name server 10.99.99.1 10.99.99.2
(Routing) (Config)#exit
(Routing) (Config)#
```

Step 11. If desired, configure the time zone and time synchronization (SNTP).

Example

The following example shows the `sntp` commands, specifying the IP address of the SNTP server and the relative timezone.

```
(Routing) #
(Routing) (Config)#sntp client mode unicast
(Routing) (Config)#sntp server 10.99.99.5
(Routing) (Config)#clock timezone -7
(Routing) (Config)#exit
(Routing) (Config)#
```

Step 12. Configure the switch name:
```
hostname IP_switch_A_1
```

Example

The switch prompt will display the new name:

```
(Routing) # hostname IP_switch_A_1

(IP_switch_A_1) #
```

Step 13. Save the configuration:
```
write memory
```

Example

You receive prompts and output similar to the following example:

```
(IP_switch_A_1) #write memory

This operation may take a few minutes.
Management interfaces will not be available during this time.

Are you sure you want to save? (y/n) y

Config file 'startup-config' created successfully .


Configuration Saved!

(IP_switch_A_1) #
```

Step 14. Repeat the previous steps on the other three switches in the MetroCluster IP configuration.

Downloading and installing the Broadcom switch EFOS software

You must download the switch operating system file and RCF file to each switch in the MetroCluster IP configuration.

About this task

This task must be repeated on each switch in the MetroCluster IP configuration.

Step 1. Copy the switch software to the switch:
`copy sftp://user@50.50.50.50/switchsoftware/efos-3.4.3.1.stk backup`

Example

In this example, the `efos-3.4.3.1.stk` operating system file is copied from the SFTP server at *50.50.50.50* to the backup partition. You need to use the IP address of your TFTP/SFTP server and the file name of the RCF file that you need to install.

```
(IP_switch_A_1) #copy sftp://user@50.50.50.50/switchsoftware/efos-3.4.3.1.stk backup
Remote Password:*************

Mode..................................... SFTP
Set Server IP.................................. 50.50.50.50
Path..................................... /switchsoftware/
Filename..................................... efos-3.4.3.1.stk
Data Type.................................... Code
Destination Filename......................... backup

Management access will be blocked for the duration of the transfer
Are you sure you want to start? (y/n) y

File transfer in progress. Management access will be blocked for the duration of the transfer. Please wait...
SFTP Code transfer starting...

File transfer operation completed successfully.

(IP_switch_A_1) #
```

Step 2. Set the switch to boot from the backup partition on the next switch reboot:
`boot system backup`

Example

```
(IP_switch_A_1) #boot system backup
Activating image backup ..

(IP_switch_A_1) #
```

Step 3. Verify that the new boot image will be active on the next boot:
`show bootvar`

Example

```
(IP_switch_A_1) #show bootvar

Image Descriptions

 active :
 backup :

 Images currently available on Flash

 ----  -----------  --------  ---------------  ------------
 unit        active    backup   current-active   next-active
 ----  -----------  --------  ---------------  ------------

1        3.4.3.0       3.4.3.1       3.4.3.0           3.4.3.1

(IP_switch_A_1) #
```

Step 4. Save the configuration:
`write memory`

Example

```
(IP_switch_A_1) #write memory

This operation may take a few minutes.
Management interfaces will not be available during this time.

Are you sure you want to save? (y/n) y

Configuration Saved!

(IP_switch_A_1) #
```

Step 5. Reboot the switch:
`reload`

Example

```
(IP_switch_A_1) #reload

Are you sure you would like to reset the system? (y/n) y
```

Step 6. Wait for the switch to reboot.

Step 7. Repeat these steps on the remaining three IP switches in the MetroCluster IP configuration.

Downloading and installing the Broadcom RCF files

You must download and install the switch RCF file to each switch in the MetroCluster IP configuration.

Before you begin

This task requires file transfer software, such as FTP, TFTP, SFTP, or SCP, to copy the files to the switches.

About this task

These steps must be repeated on each of the IP switches in the MetroCluster IP configuration.

There are four RCF files, one for each of the four switches in the MetroCluster IP configuration. You must use the correct RCF files for the switch model you are using.

Switch	RCF file		
IP_switch_A_1	BES-53248_v1.32_Switch-A1.txt		
IP_switch_A_2	BES-53248_v1.32_Switch-A2.txt		
IP_switch_B_1	BES-53248_v1.32_Switch-B1.txt		
IP_switch_B_2	BES-53248_v1.32_Switch-B2.txt		

Step 1. Download the MetroCluster IP RCF files for the Broadcom switch.

https://datacentersupport.lenovo.com/

Step 2. Copy the RCF files to the switches:

a. Copy the RCF files to the first switch:
`copy sftp://user@`*`FTP-server-IP-address`*`/RcfFiles/`*`switch-specific-RCF`* `nvram:script BES-53248_v1.32_Switch-A1.txt nvram:script BES-53248_v1.32_Switch-A1.scr`

Example

In this example, the `BES-53248_v1.32_Switch-A1.txt` RCF file is copied from the SFTP server at *50.50.50.50* to the local bootflash. You need to use the IP address of your TFTP/SFTP server and the file name of the RCF file that you need to install.

```
(IP_switch_A_1) #copy sftp://user@50.50.50.50/RcfFiles/BES-53248_v1.32_Switch-A1.txt nvram:
script BES-53248_v1.32_Switch-A1.scr

Remote Password:*************

Mode..................................... SFTP
Set Server IP.................................. 50.50.50.50
Path..................................... /RcfFiles/
Filename..................................... BES-53248_v1.32_Switch-A1.txt
Data Type.................................... Config Script
Destination Filename......................... BES-53248_v1.32_Switch-A1.scr

Management access will be blocked for the duration of the transfer
Are you sure you want to start? (y/n) y

File transfer in progress. Management access will be blocked for the duration of the transfer. Please wait...
File transfer operation completed successfully.


Validating configuration script...

config

set clibanner "*********************************************************************************
```

```
* NetApp Reference Configuration File (RCF)

*

* Switch    : BES-53248


...
The downloaded RCF is validated. Some output is being logged here.
...


Configuration script validated.
File transfer operation completed successfully.

(IP_switch_A_1) #
```

b. Verify that the RCF file is saved as a script:
`script list`

Example

```
(IP_switch_A_1) #script list

Configuration Script Name        Size(Bytes)  Date of Modification
-------------------------------  -----------  --------------------
BES-53248_v1.32_Switch-A1.scr              852   2019 01 29 18:41:25

1 configuration script(s) found.
2046 Kbytes free.
(IP_switch_A_1) #
```

c. Apply the RCF script:
`script apply BES-53248_v1.32_Switch-A1.scr`

Example

```
(IP_switch_A_1) #script apply BES-53248_v1.32_Switch-A1.scr

Are you sure you want to apply the configuration script? (y/n) y


config

set clibanner "*********************************************************************

* Lenovo Reference Configuration File (RCF)

*

* Switch    : BES-53248

...
The downloaded RCF is validated. Some output is being logged here.
...

Configuration script 'BES-53248_v1.32_Switch-A1.scr' applied.

(IP_switch_A_1) #
```

d. Save the configuration:

```
write memory
```

Example

```
(IP_switch_A_1) #write memory

This operation may take a few minutes.
Management interfaces will not be available during this time.

Are you sure you want to save? (y/n) y

Configuration Saved!

(IP_switch_A_1) #
```

e. Reboot the switch:
```
reload
```

Example

```
(IP_switch_A_1) #reload

Are you sure you would like to reset the system? (y/n) y
```

f. Repeat the previous steps for each of the other three switches, being sure to copy the matching RCF file to the corresponding switch.

Step 3. Reload the switch: `reload`

Example

```
IP_switch_A_1# reload
```

Step 4. Repeat the previous steps on the other three switches in the MetroCluster IP configuration.

Configuring Cisco IP switches

You must configure the Cisco IP switches for use as the cluster interconnect and for backend MetroCluster IP connectivity.

Resetting the Cisco IP switch to factory defaults

Before installing a new software version and RCFs, you must erase the Cisco switch configuration and perform basic configuration.

About this task

You must repeat these steps on each of the IP switches in the MetroCluster IP configuration.

Step 1. Reset the switch to factory defaults:

a. Erase the existing configuration:
```
write erase
```

b. Reload the switch software:
```
reload
```
The system reboots and enters the configuration wizard. During the boot, if you receive the prompt `Abort Auto Provisioning and continue with normal setup?(yes/no)[n]`, you should respond `yes` to proceed.

c. In the configuration wizard, enter the basic switch settings:

- Admin password

- Switch name
- Out-of-band management configuration
- Default gateway
- SSH service (RSA)

After completing the configuration wizard, the switch reboots.

d. When prompted, enter the user name and password to log in to the switch.

Example

The following example shows the prompts and system responses when configuring the switch. The angle brackets (<<<) show where you enter the information.

```
---- System Admin Account Setup ----
Do you want to enforce secure password standard (yes/no) [y]:y  <<<

    Enter the password for "admin": password  <<<
  Confirm the password for "admin": password  <<<
         ---- Basic System Configuration Dialog VDC: 1 ----

This setup utility will guide you through the basic configuration of
the system. Setup configures only enough connectivity for management
of the system.

Please register Cisco Nexus3000 Family devices promptly with your
supplier. Failure to register may affect response times for initial
service calls. Nexus3000 devices must be registered to receive
entitled support services.

Press Enter at anytime to skip a dialog. Use ctrl-c at anytime
to skip the remaining dialogs.
```

You enter basic information in the next set of prompts, including the switch name, management address, and gateway, and select SSH with RSA.

```
Would you like to enter the basic configuration dialog (yes/no): yes
  Create another login account (yes/no) [n]:
  Configure read-only SNMP community string (yes/no) [n]:
  Configure read-write SNMP community string (yes/no) [n]:
  Enter the switch name : switch-name <<<
  Continue with Out-of-band (mgmt0) management configuration? (yes/no) [y]:
    Mgmt0 IPv4 address : management-IP-address  <<<
    Mgmt0 IPv4 netmask : management-IP-netmask  <<<
  Configure the default gateway? (yes/no) [y]: y <<<
    IPv4 address of the default gateway : gateway-IP-address  <<<
  Configure advanced IP options? (yes/no) [n]:
  Enable the telnet service? (yes/no) [n]:
  Enable the ssh service? (yes/no) [y]: y  <<<
    Type of ssh key you would like to generate (dsa/rsa) [rsa]: rsa <<<
    Number of rsa key bits <1024-2048> [1024]:
  Configure the ntp server? (yes/no) [n]:
  Configure default interface layer (L3/L2) [L2]:
  Configure default switchport interface state (shut/noshut) [noshut]: shut <<<
  Configure CoPP system profile (strict/moderate/lenient/dense) [strict]:
```

The final set of prompts completes the configuration:

```
The following configuration will be applied:
  password strength-check
  switchname IP_switch_A_1
vrf context management
ip route 0.0.0.0/0 10.10.99.1
```

```
exit
  no feature telnet
  ssh key rsa 1024 force
  feature ssh
  system default switchport
  system default switchport shutdown
  copp profile strict
interface mgmt0
ip address 10.10.99.10 255.255.255.0
no shutdown

Would you like to edit the configuration? (yes/no) [n]:

Use this configuration and save it? (yes/no) [y]:
2017 Jun 13 21:24:43 A1 %$ VDC-1 %$ %COPP-2-COPP_POLICY: Control-Plane is protected with
policy copp-system-p-policy-strict.

[########################################] 100%
Copy complete.

User Access Verification
IP_switch_A_1 login: admin
Password:
Cisco Nexus Operating System (NX-OS) Software
.
.
.
IP_switch_A_1#
```

Step 2. Save the configuration:

```
IP_switch-A-1# copy running-config startup-config
```

Step 3. Reboot the switch and wait for the switch to reload:

```
IP_switch-A-1# reload
```

Step 4. Repeat the previous steps on the other three switches in the MetroCluster IP configuration.

Downloading and installing the Cisco switch NX-OS software

You must download the switch operating system file and RCF file to each switch in the MetroCluster IP configuration.

Before you begin

This task requires file transfer software, such as FTP, TFTP, SFTP, or SCP, to copy the files to the switches.

About this task

These steps must be repeated on each of the IP switches in the MetroCluster IP configuration.

You must use the supported switch software version.

Step 1. Download the supported NX-OS software file from the CISCO Web site.

Step 2. Copy the switch software to the switch:

`copy sftp://root@`*server-ip-address*`/tftpboot/`*NX-OS-file-name*` bootflash: vrf management`

Example

In this example, the `nxos.7.0.3.I4.6.bin` file is copied from SFTP server 10.10.99.99 to the local bootflash:

```
IP_switch_A_1# copy sftp://root@10.10.99.99/tftpboot/nxos.7.0.3.I4.6.bin bootflash: vrf management
root@10.10.99.99's password: password
sftp> progress
```

```
Progress meter enabled
sftp> get   /tftpboot/nxos.7.0.3.I4.6.bin  /bootflash/nxos.7.0.3.I4.6.bin
Fetching /tftpboot/nxos.7.0.3.I4.6.bin to /bootflash/nxos.7.0.3.I4.6.bin
/tftpboot/nxos.7.0.3.I4.6.bin                 100%  666MB   7.2MB/s   01:32
sftp> exit
Copy complete, now saving to disk (please wait)...
```

Step 3. Verify on each switch that the switch NX-OS files are present in each switch's `bootflash` directory: `dir bootflash:`

Example

The following example shows that the files are present on IP_switch_A_1:

```
IP_switch_A_1# dir bootflash:
                   .
                   .
                   .
  698629632    Jun 13 21:37:44 2017  nxos.7.0.3.I4.6.bin
                   .
                   .
                   .

Usage for bootflash://sup-local
 1779363840 bytes used
13238841344 bytes free
15018205184 bytes total
IP_switch_A_1#
```

Step 4. Install the switch software:
`install all nxos bootflash:nxos.`*version-number*`.bin`The switch will reload (reboot) automatically after the switch software has been installed.

Example

The following example shows the software installation on IP_switch_A_1:

```
IP_switch_A_1# install all nxos bootflash:nxos.7.0.3.I4.6.bin
Installer will perform compatibility check first. Please wait.
Installer is forced disruptive

Verifying image bootflash:/nxos.7.0.3.I4.6.bin for boot variable "nxos".
[####################] 100% -- SUCCESS

Verifying image type.
[####################] 100% -- SUCCESS

Preparing "nxos" version info using image bootflash:/nxos.7.0.3.I4.6.bin.
[####################] 100% -- SUCCESS

Preparing "bios" version info using image bootflash:/nxos.7.0.3.I4.6.bin.
[####################] 100% -- SUCCESS       [####################] 100%            -- SUCCESS

Performing module support checks.            [####################] 100%            -- SUCCESS

Notifying services about system upgrade.     [####################] 100%            -- SUCCESS



Compatibility check is done:
Module  bootable          Impact  Install-type  Reason
------  --------  --------------  ------------  ------
     1       yes      disruptive         reset  default upgrade is not hitless
```

```
Images will be upgraded according to following table:
Module       Image   Running-Version(pri:alt)          New-Version   Upg-Required
------  ----------   ------------------------  ------------------   ------------
     1        nxos                7.0(3)I4(1)          7.0(3)I4(6)   yes
     1        bios         v04.24(04/21/2016)  v04.24(04/21/2016)   no


Switch will be reloaded for disruptive upgrade.
Do you want to continue with the installation (y/n)?  [n] y


Install is in progress, please wait.

Performing runtime checks.         [####################] 100%    -- SUCCESS

Setting boot variables.
[####################] 100% -- SUCCESS

Performing configuration copy.
[####################] 100% -- SUCCESS

Module 1: Refreshing compact flash and upgrading bios/loader/bootrom.
Warning: please do not remove or power off the module at this time.
[####################] 100% -- SUCCESS


Finishing the upgrade, switch will reboot in 10 seconds.
IP_switch_A_1#
```

Step 5. Wait for the switch to reload and then log in to the switch.

After the switch has rebooted the login prompt is displayed:

```
User Access Verification
IP_switch_A_1 login: admin
Password:
Cisco Nexus Operating System (NX-OS) Software
TAC support: http://www.cisco.com/tac
Copyright (C) 2002-2017, Cisco and/or its affiliates.
All rights reserved.
.
.
.
MDP database restore in progress.
IP_switch_A_1#

The switch software is now installed.
```

Step 6. Verify that the switch software has been installed:

`show version`

The following example shows the output:

```
IP_switch_A_1# show version
Cisco Nexus Operating System (NX-OS) Software
TAC support: http://www.cisco.com/tac
Copyright (C) 2002-2017, Cisco and/or its affiliates.
All rights reserved.
.
.
.

Software
  BIOS: version 04.24
  NXOS: version 7.0(3)I4(6)   <<< switch software version
  BIOS compile time:  04/21/2016
  NXOS image file is: bootflash:///nxos.7.0.3.I4.6.bin
```

```
  NXOS compile time:  3/9/2017 22:00:00 [03/10/2017 07:05:18]


Hardware
  cisco Nexus 3132QV Chassis
  Intel(R) Core(TM) i3- CPU @ 2.50GHz with 16401416 kB of memory.
  Processor Board ID FOC20123GPS

  Device name: A1
  bootflash:   14900224 kB
  usb1:               0 kB (expansion flash)

Kernel uptime is 0 day(s), 0 hour(s), 1 minute(s), 49 second(s)

Last reset at 403451 usecs after  Mon Jun 10 21:43:52 2017

  Reason: Reset due to upgrade
  System version: 7.0(3)I4(1)
  Service:

plugin
  Core Plugin, Ethernet Plugin
IP_switch_A_1#
```

Step 7. Repeat these steps on the remaining three IP switches in the MetroCluster IP configuration.

Downloading and installing the Cisco IP RCF files

You must download the RCF file to each switch in the MetroCluster IP configuration.

Before you begin

This task requires file transfer software, such as FTP, TFTP, SFTP, or SCP, to copy the files to the switches.

About this task

These steps must be repeated on each of the IP switches in the MetroCluster IP configuration.

You must use the supported switch software version.

Lenovo Press

There are four RCF files, one for each of the four switches in the MetroCluster IP configuration. You must use the correct RCF files for the switch model you are using.

Switch	RCF file
IP_switch_A_1	NX3232_v1.8T-X1_Switch-A1.txt
IP_switch_A_2	NX3232_v1.8T-X1_Switch-A2.txt
IP_switch_B_1	NX3232_v1.8T-X1_Switch-B1.txt
IP_switch_B_2	NX3232_v1.8T-X1_Switch-B2.txt

Step 1. Download the MetroCluster IP RCF files.

Lenovo Downloads: MetroCluster IP Switch Configuration Files

Step 2. Copy the RCF files to the switches:

a. Copy the RCF files to the first switch:
`copy sftp://root@`*FTP-server-IP-address*`/tftpboot/`*switch-specific-RCF*` bootflash: vrf management`

Example

In this example, the`NX3232_v1.8T-X1_Switch-A1.txt` RCF file is copied from the SFTP server at*10.10.99.99* to the local bootflash. You must use the IP address of your TFTP/SFTP server and the file name of the RCF file that you need to install.

```
IP_switch_A_1# copy sftp://root@10.10.99.99/tftpboot/NX3232_v1.8T-X1_Switch-A1.txt bootflash: vrf management
root@10.10.99.99's password: password
sftp> progress
Progress meter enabled
sftp> get /tftpboot/NX3232_v1.8T-X1_Switch-A1.txt /bootflash/NX3232_v1.8T-X1_Switch-A1.txt
Fetching /tftpboot/NX3232_v1.8T-X1_Switch-A1.txt to /bootflash/NX3232_v1.8T-X1_Switch-A1.txt
/tftpboot/NX3232_v1.8T-X1_Switch-A1.txt 100% 5141 5.0KB/s 00:00
sftp> exit
Copy complete, now saving to disk (please wait)...
IP_switch_A_1#
```

b. Repeat the previous substep for each of the other three switches, being sure to copy the matching RCF file to the corresponding switch.

Step 3. Verify on each switch that the RCF file is present in each switch's`bootflash` directory:
`dir bootflash:`

Example

The following example shows that the files are present on IP_switch_A_1:

```
IP_switch_A_1# dir bootflash:
.
.
.
5514 Jun 13 22:09:05 2017 NX3232_v1.8T-X1_Switch-A1.txt
.
.
.
Usage for bootflash://sup-local
1779363840 bytes used
13238841344 bytes free
15018205184 bytes total
IP_switch_A_1#
```

Step 4. Copy the matching RCF file from the local bootflash to the running configuration on each switch:
`copy bootflash:`*switch-specific-RCF*`.txt running-config`

Step 5. Copy the RCF files from the running configuration to the startup configuration on each switch:
`copy running-config startup-config`

You should see output similar to the following:

```
IP_switch_A_1# copy bootflash:NX3232_v1.8T-X1_Switch-A1.txt running-config
IP_switch-A-1# copy running-config startup-config
```

Step 6. Reload the switch:
`reload`

Example

```
IP_switch_A_1# reload
```

Step 7. Repeat the previous steps on the other three switches in the MetroCluster IP configuration.

Setting Forward Error Correction for AFA DM7000F and DM7000H systems using 25-Gbps connectivity

If your AFA DM7000F or DM7000H system is configured using 25-Gbps connectivity, you need to set the Forward Error Correction (fec) parameter manually to off after applying the RCF file. The RCF file does not apply this setting.

Before you begin

The 25-Gbps ports must be cabled prior to performing this procedure.

"Port assignments for AFA DM7000F or DM7000H systems with Cisco 3232C IP switches" on page 35

About this task

This task only applies to AFA DM7000F and DM7000H platforms using 25-Gbps connectivity.

This task must be performed on all four switches in the MetroCluster IP configuration.

Step 1. Set the fec parameter to off on each 25-Gbps port that is connected to a controller module, and then copy the running configuration to the startup configuration:

a. Enter configuration mode:
`config t`

b. Specify the 25-Gbps interface to configure:
`interface` *interface-ID*

c. Set fec to off:
`fec off`

d. Repeat the previous steps for each 25-Gbps port on the switch.

e. Exit configuration mode:
`exit`

Example

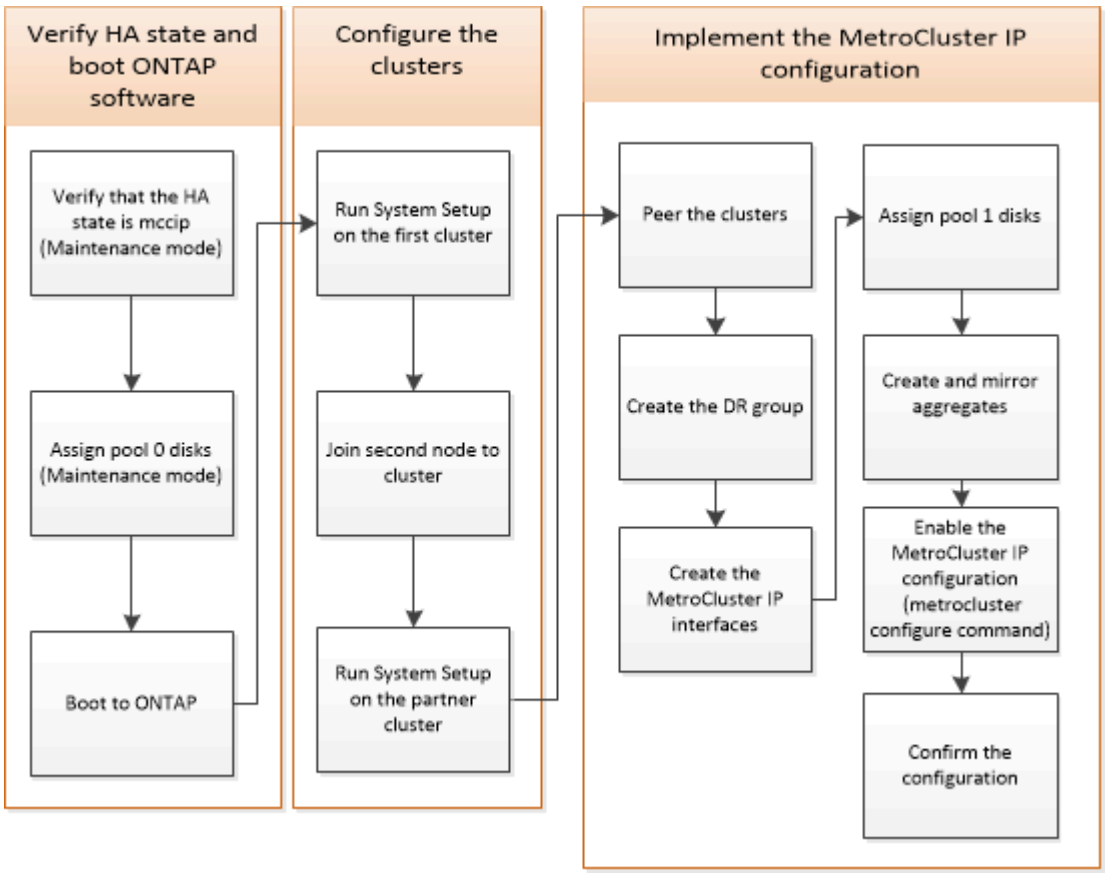
The following example shows the commands for interface Ethernet1/25/1 on switch IP_switch_A_1:

```
IP_switch_A_1# conf t
IP_switch_A_1(config)# interface Ethernet1/25/1
IP_switch_A_1(config-if)# fec off
IP_switch_A_1(config-if)# exit
IP_switch_A_1(config-if)# end
IP_switch_A_1# copy running-config startup-config
```

Step 2. Repeat the previous step on the other three switches in the MetroCluster IP configuration.

Chapter 4. Configuring the MetroCluster software in ONTAP

You must set up each node in the MetroCluster configuration in ONTAP, including the node-level configurations and the configuration of the nodes into two sites. You must also implement the MetroCluster relationship between the two sites.



Gathering required information

You need to gather the required IP addresses for the controller modules before you begin the configuration process.

IP network information worksheet for site A

You must obtain IP addresses and other network information for the first MetroCluster site (site A) from your network administrator before you configure the system.

Site A switch information

When you cable the system, you need a host name and management IP address for each cluster switch.

Cluster switch	Host name	IP address	Network mask	Default gateway
Interconnect 1				
Interconnect 2				

Cluster switch	Host name	IP address	Network mask	Default gateway
Management 1				
Management 2				

Site A cluster creation information

When you first create the cluster, you need the following information:

Type of information	Your values
Cluster name Example used in this guide: site_A	
DNS domain	
DNS name servers	
Location	
Administrator password	

Site A node information

For each node in the cluster, you need a management IP address, a network mask, and a default gateway.

Node	Port	IP address	Network mask	Default gateway
Node 1 Example used in this guide: controller_A_1				
Node 2 Example used in this guide: controller_A_2				

Site A LIFs and ports for MetroCluster IP back-end connectivity

For each node in the cluster, you need the IP addresses of two MetroCluster IP LIFs, including a network mask and a default gateway. The MetroCluster IP LIFs are used for MetroCluster IP back-end connectivity.

"Considerations for MetroCluster IP configuration" on page 4

Node	Port	IP address of MetroCluster IP LIF	Network mask	Default gateway
Node 1 MetroCluster IP LIF 1	e1a			
Node 1 MetroCluster IP LIF 2	e1b			
Node 2 MetroCluster IP LIF 1	e1a			
Node 2 MetroCluster IP LIF 2	e1b			

Site A LIFs and ports for cluster peering

For each node in the cluster, you need the IP addresses of two intercluster LIFs, including a network mask and a default gateway. The intercluster LIFs are used to peer the clusters.

Node	Port	IP address of intercluster LIF	Network mask	Default gateway
Node 1 IC LIF 1				
Node 1 IC LIF 2				
Node 2 IC LIF 1				
Node 2 IC LIF 2				

Site A time server information

You must synchronize the time, which requires one or more NTP time servers.

Node	Host name	IP address	Network mask	Default gateway
NTP server 1				
NTP server 2				

Site A AutoSupport information

You must configure AutoSupport on each node, which requires the following information:

Type of information		Your values
From email address		
Mail hosts	IP addresses or names	
Transport protocol	HTTP, HTTPS, or SMTP	
	Proxy server	
Recipient email addresses or distribution lists	Full-length messages	
	Concise messages	
	Partners	

Site A SP information

You must enable access to the Service Processor (SP) of each node for troubleshooting and maintenance, which requires the following network information for each node:

Node	IP address	Network mask	Default gateway
Node 1			

IP network information worksheet for site B

You must obtain IP addresses and other network information for the second MetroCluster site (site B) from your network administrator before you configure the system.

Site B switch information

When you cable the system, you need a host name and management IP address for each cluster switch.

Cluster switch	Host name	IP address	Network mask	Default gateway
Interconnect 1				
Interconnect 2				
Management 1				
Management 2				

Site B cluster creation information

When you first create the cluster, you need the following information:

Type of information	Your values
Cluster name Example used in this guide: site_B	
DNS domain	
DNS name servers	
Location	
Administrator password	

Site B node information

For each node in the cluster, you need a management IP address, a network mask, and a default gateway.

Node	Port	IP address	Network mask	Default gateway
Node 1 Example used in this guide: controller_B_1				
Node 2 Example used in this guide: controller_B_2				

Site B LIFs and ports for MetroCluster IP back-end connectivity

For each node in the cluster, you need the IP addresses of two MetroCluster IP LIFs, including a network mask and a default gateway. The MetroCluster IP LIFs are used for MetroCluster IP back-end connectivity.

"Considerations for MetroCluster IP configuration" on page 4

Node	Port	IP address of MetroCluster IP LIF	Network mask	Default gateway
Node 1 MetroCluster IP LIF 1	e1a			
Node 1 MetroCluster IP LIF 2	e1b			
Node 2 MetroCluster IP LIF 1	e1a			
Node 2 MetroCluster IP LIF 2	e1b			

Site B LIFs and ports for cluster peering

For each node in the cluster, you need the IP addresses of two intercluster LIFs, including a network mask and a default gateway. The intercluster LIFs are used to peer the clusters.

Node	Port	IP address of intercluster LIF	Network mask	Default gateway
Node 1 IC LIF 1				
Node 1 IC LIF 2				
Node 2 IC LIF 1				
Node 2 IC LIF 2				

Site B time server information

You must synchronize the time, which requires one or more NTP time servers.

Node	Host name	IP address	Network mask	Default gateway
NTP server 1				
NTP server 2				

Site B AutoSupport information

You must configure AutoSupport on each node, which requires the following information:

Type of information		Your values
From email address		
Mail hosts	IP addresses or names	
Transport protocol	HTTP, HTTPS, or SMTP	
	Proxy server	
Recipient email addresses or distribution lists	Full-length messages	
	Concise messages	
	Partners	

Site B SP information

You must enable access to the Service Processor (SP) of each node for troubleshooting and maintenance, which requires the following network information for each node:

Node	IP address	Network mask	Default gateway
Node 1 (controller_B_1)			

Similarities and differences between standard cluster and MetroCluster configurations

The configuration of the nodes in each cluster in a MetroCluster configuration is similar to that of nodes in a standard cluster.

The MetroCluster configuration is built on two standard clusters. Physically, the configuration must be symmetrical, with each node having the same hardware configuration, and all of the MetroCluster components must be cabled and configured. However, the basic software configuration for nodes in a MetroCluster configuration is the same as that for nodes in a standard cluster.

Configuration step	Standard cluster configuration	MetroCluster configuration
Configure management, cluster, and data LIFs on each node.	Same in both types of clusters	
Configure the root aggregate.	Same in both types of clusters	
Set up the cluster on one node in the cluster.	Same in both types of clusters	
Join the other node to the cluster.	Same in both types of clusters	
Create a mirrored root aggregate.	Optional	Required
Peer the clusters.	Optional	Required
Enable the MetroCluster configuration.	Does not apply	Required

Restoring system defaults on a previously used controller module

If your controller modules have been used previously, you must reset them for a successful MetroCluster configuration.

About this task

Important: This task is required only on controller modules that have been previously configured. You do not need to perform this task if you received the controller modules from the factory.

Step 1. At the LOADER prompt, return the environmental variables to their default setting: `set-defaults`

Step 2. Boot the node to the boot menu:
`boot_ontap menu`

After you run the command, wait until the boot menu is shown.

Step 3. Clear the node configuration:

– If you are using systems configured for ADP, select option `9a` from the boot menu, and respond **`yes`** when prompted.

Note: This process is disruptive.

The following screen shows the boot menu prompt:

```
Please choose one of the following:

    (1) Normal Boot.
    (2) Boot without /etc/rc.
    (3) Change password.
    (4) Clean configuration and initialize all disks.
    (5) Maintenance mode boot.
    (6) Update flash from backup config.
    (7) Install new software first.
    (8) Reboot node.
    (9) Configure Advanced Drive Partitioning.
```

```
    Selection (1-9)?  9a
########## WARNING ##########

    This is a disruptive operation and will result in the
    loss of all filesystem data. Before proceeding further,
    make sure that:
    1) This option (9a) has been executed or will be executed
    on the HA partner node, prior to reinitializing either
    system in the HA-pair.
    2) The HA partner node is currently in a halted state or
    at the LOADER prompt.


    Do you still want to continue (yes/no)? yes
```

– If your system is not configured for ADP, type `wipeconfig` at the boot menu prompt, and then press Enter.

The following screen shows the boot menu prompt:

```
Please choose one of the following:

    (1) Normal Boot.
    (2) Boot without /etc/rc.
    (3) Change password.
    (4) Clean configuration and initialize all disks.
    (5) Maintenance mode boot.
    (6) Update flash from backup config.
    (7) Install new software first.
    (8) Reboot node.
    (9) Configure Advanced Drive Partitioning.
    Selection (1-9)?  wipeconfig
This option deletes critical system configuration, including cluster membership.
Warning: do not run this option on a HA node that has been taken over.
Are you sure you want to continue?: yes
Rebooting to finish wipeconfig request.
```

Verifying the ha-config state of components

In a MetroCluster IP configuration that is not preconfigured at the factory, you must verify that the ha-config state of the controller and chassis components is set to `mccip` so that they boot up properly. For systems received from the factory, this value is preconfigured and you do not need to verify it.

Before you begin

The system must be in Maintenance mode.

Step 1. Display the HA state of the controller module and chassis:
`ha-config show`

The controller module and chassis should show the value `mccip`.

Step 2. If the displayed system state of the controller is not `mccip`, set the HA state for the controller:
`ha-config modify controller mccip`

Step 3. If the displayed system state of the chassis is not `mccip`, set the HA state for the chassis:
`ha-config modify chassis mccip`

Step 4. Repeat these steps on each node in the MetroCluster configuration.

Manually assigning drives to pool 0

If you did not receive the systems pre-configured from the factory, you might have to manually assign the pool 0 drives. Depending on the platform model and whether the system is using ADP, you must manually assign drives to pool 0 for each node in the MetroCluster IP configuration. The procedure you use depends on the version of ONTAP you are using.

Manually assigning drives for pool 0

If the system has not been pre-configured at the factory and does not meet the requirements for automatic drive assignment, you must manually assign the pool 0 drives.

About this task

To determine if your system requires manual disk assignment, you should review "Considerations for automatic drive assignment and ADP systems" on page 6.

You perform these steps in Maintenance mode. The procedure must be performed on each node in the configuration.

Examples in this section are based on the following assumptions:

- node_A_1 and node_A_2 own drives on:
  - site_A-shelf_1 (local)
  - site_B-shelf_2 (remote)
- node_B_1 and node_B_2 own drives on:
  - site_B-shelf_1 (local)
  - site_A-shelf_2 (remote)

Step 1. Display the boot menu:

```
boot_ontap menu
```

Step 2. Select option 9a.

Example

The following screen shows the boot menu prompt:

```
Please choose one of the following:

    (1) Normal Boot.
    (2) Boot without /etc/rc.
    (3) Change password.
    (4) Clean configuration and initialize all disks.
    (5) Maintenance mode boot.
    (6) Update flash from backup config.
    (7) Install new software first.
    (8) Reboot node.
    (9) Configure Advanced Drive Partitioning.
    Selection (1-9)?  9a
########## WARNING ##########

    This is a disruptive operation and will result in the
    loss of all filesystem data. Before proceeding further,
    make sure that:
```

```
1) This option (9a) has been executed or will be executed
on the HA partner node (and DR/DR-AUX partner nodes if
applicable), prior to reinitializing any system in the
HA-pair (or MCC setup).
2) The HA partner node (and DR/DR-AUX partner nodes if
applicable) is currently waiting at the boot menu.

Do you still want to continue (yes/no)? yes
```

Step 3. When the node restarts, press Ctrl-C when prompted to display the boot menu and then select the option for **Maintenance mode boot**.

Step 4. In Maintenance mode, manually assign drives for the local aggregates on the node:
`disk assign` *disk-id* `-p 0 -s` *local-node-sysid*

The drives should be assigned symmetrically, so each node has an equal number of drives. The following steps are for a configuration with two storage shelves at each site.

a. When configuring node_A_1, manually assign drives from slot 0 to 11 to pool0 of node A1 from site_A-shelf_1.

b. When configuring node_A_2, manually assign drives from slot 12 to 23 to pool0 of node A2 from site_A-shelf_1.

c. When configuring node_B_1, manually assign drives from slot 0 to 11 to pool0 of node B1 from site_B-shelf_1.

d. When configuring node_B_2, manually assign drives from slot 12 to 23 to pool0 of node B2 from site_B-shelf_1.

Step 5. Exit Maintenance mode:
`halt`

Step 6. Display the boot menu:
`boot_ontap menu`

Step 7. Select option 4 from the boot menu and let the system boot.

Step 8. Repeat these steps on the other nodes in the MetroCluster IP configuration.

Step 9. Proceed to “Setting up ONTAP” on page 67.

Setting up ONTAP

After you boot each node, you are prompted to perform basic node and cluster configuration. After configuring the cluster, you return to the ONTAP CLI to create aggregates and create the MetroCluster configuration.

Before you begin

- You must have cabled the MetroCluster configuration.
- You must not have configured the Service Processor.

About this task

This task must be performed on both clusters in the MetroCluster configuration.

Step 1. Power up each node at the local site if you have not already done so and let them all boot completely.

If the system is in Maintenance mode, you need to issue the `halt` command to exit Maintenance mode, and then issue the `boot_ontap` command to boot the system and get to cluster setup.

Step 2. On the first node in each cluster, proceed through the prompts to configure the cluster

a. Enable the AutoSupport tool by following the directions provided by the system.

Example

The output should be similar to the following:

```
Welcome to the cluster setup wizard.

    You can enter the following commands at any time:
    "help" or "?" - if you want to have a question clarified,
    "back" - if you want to change previously answered questions, and
    "exit" or "quit" - if you want to quit the cluster setup wizard.
    Any changes you made before quitting will be saved.

    You can return to cluster setup at any time by typing "cluster setup".
    To accept a default or omit a question, do not enter a value.

    This system will send event messages and periodic reports to Lenovo Technical
    Support. To disable this feature, enter
    autosupport modify -support disable
    within 24 hours.

    Enabling AutoSupport can significantly speed problem determination and
    resolution should a problem occur on your system.
    For further information on AutoSupport, see:
    http://support.netapp.com/autosupport/

    Type yes to confirm and continue {yes}: yes

.
.
.
```

b. Configure the node management interface by responding to the prompts.

Example

The prompts are similar to the following:

```
Enter the node management interface port [e0M]:
Enter the node management interface IP address: 172.17.8.229
Enter the node management interface netmask: 255.255.254.0
Enter the node management interface default gateway: 172.17.8.1
A node management interface on port e0M with IP address 172.17.8.229 has been created.
```

c. Create the cluster by responding to the prompts.

Example

The prompts are similar to the following:

```
Do you want to create a new cluster or join an existing cluster? {create, join}:
create


Do you intend for this node to be used as a single node cluster? {yes, no} [no]:
no

Existing cluster interface configuration found:

Port MTU IP Netmask
e0a 1500 169.254.18.124 255.255.0.0
e1a 1500 169.254.184.44 255.255.0.0
```

```
Do you want to use this configuration? {yes, no} [yes]: no

System Defaults:
Private cluster network ports [e0a,e1a].
Cluster port MTU values will be set to 9000.
Cluster interface IP addresses will be automatically generated.

Do you want to use these defaults? {yes, no} [yes]: no

Enter the cluster administrator's (username "admin") password:

Retype the password:


Step 1 of 5: Create a Cluster
You can type "back", "exit", or "help" at any question.

List the private cluster network ports [e0a,e1a]:
Enter the cluster ports' MTU size [9000]:
Enter the cluster network netmask [255.255.0.0]: 255.255.254.0
Enter the cluster interface IP address for port e0a: 172.17.10.228
Enter the cluster interface IP address for port e1a: 172.17.10.229
Enter the cluster name: cluster_A

Creating cluster cluster_A

Starting cluster support services ...

Cluster cluster_A has been created.
```

d. Add licenses, set up a Cluster Administration SVM, and enter DNS information by responding to the prompts.

Example

The prompts are similar to the following:

```
Step 2 of 5: Add Feature License Keys
You can type "back", "exit", or "help" at any question.

Enter an additional license key []:


Step 3 of 5: Set Up a Vserver for Cluster Administration
You can type "back", "exit", or "help" at any question.


Enter the cluster management interface port [e3a]:
Enter the cluster management interface IP address: 172.17.12.153
Enter the cluster management interface netmask: 255.255.252.0
Enter the cluster management interface default gateway: 172.17.12.1

A cluster management interface on port e3a with IP address 172.17.12.153 has been created.
You can use this address to connect to and manage the cluster.

Enter the DNS domain names: labs.lenovo.com
Enter the name server IP addresses: 172.19.2.30
DNS lookup for the admin Vserver will use the labs.lenovo.com domain.

Step 4 of 5: Configure Storage Failover (SFO)
You can type "back", "exit", or "help" at any question.

SFO will be enabled when the partner joins the cluster.

Step 5 of 5: Set Up the Node
You can type "back", "exit", or "help" at any question.
```

```
Where is the controller located []: svl
```

e. Enable storage failover and set up the node by responding to the prompts.

Example

The prompts are similar to the following:

```
Step 4 of 5: Configure Storage Failover (SFO)
You can type "back", "exit", or "help" at any question.

SFO will be enabled when the partner joins the cluster.

Step 5 of 5: Set Up the Node
You can type "back", "exit", or "help" at any question.

Where is the controller located []: site_A
```

f. Complete the configuration of the node, but do not create data aggregates.

You can use ThinkSystem Storage Manager for DM Series, pointing your web browser to the cluster management IP address (https://172.17.12.153).

Cluster Management Using Storage Manager

Step 3. Boot the next controller and join it to the cluster, following the prompts.

Step 4. Confirm that nodes are configured in high-availability mode:
`storage failover show -fields mode`

If not, you must configure HA mode on each node, and then reboot the nodes:
`storage failover modify -mode ha -node localhost`

This command configures high-availability mode but does not enable storage failover. Storage failover is automatically enabled when you configure the MetroCluster configuration later in the process.

Step 5. Confirm that you have four ports configured as cluster interconnects:
`network port show`

The MetroCluster IP interfaces are not configured at this time and do not appear in the command output.

Example

The following example shows two cluster ports on node_A_1:

```
cluster_A::*> network port show -role cluster

Node: node_A_1

                                                                        Ignore
                                                  Speed(Mbps) Health   Health
Port      IPspace      Broadcast Domain Link MTU  Admin/Oper  Status   Status
--------- ------------ ---------------- ---- ---- ----------- -------- ------
e4a       Cluster      Cluster          up   9000  auto/40000 healthy  false
e4e       Cluster      Cluster          up   9000  auto/40000 healthy  false
```

```
Node: node_A_2

                                                                          Ignore

                                                  Speed(Mbps) Health   Health

Port      IPspace      Broadcast Domain Link MTU  Admin/Oper  Status   Status

--------- ------------ ---------------- ---- ---- ----------- -------- ------

e4a       Cluster      Cluster          up   9000  auto/40000 healthy  false

e4e       Cluster      Cluster          up   9000  auto/40000 healthy  false

4 entries were displayed.
```

Step 6. Repeat these steps on the partner cluster.

After you finish

Return to the ONTAP command-line interface and complete the MetroCluster configuration by performing the tasks that follow.

Configuring the clusters into a MetroCluster configuration

You must peer the clusters, mirror the root aggregates, create a mirrored data aggregate, and then issue the command to implement the MetroCluster operations.

Disabling automatic drive assignment (if doing manual assignment in ONTAP 9.4)

In ONTAP 9.4, if your MetroCluster IP configuration has fewer than four external storage shelves per site, you must disable automatic drive assignment on all nodes and manually assign drives.

About this task

This task is not required in ONTAP 9.5 and later.

“Considerations for automatic drive assignment and ADP systems” on page 6

Step 1. Disable automatic drive assignment:
`storage disk option modify -node` *node_name* `-autoassign off`

You need to issue this command on all nodes in the MetroCluster IP configuration.

Verifying drive assignment of pool 0 drives

You must verify that the remote drives are visible to the nodes and have been assigned correctly.

About this task

Automatic assignment depends on the storage system platform model and drive shelf arrangement.

“Considerations for automatic drive assignment and ADP systems” on page 6

Step 1. Verify that pool 0 drives are assigned automatically:
`disk show`

Peering the clusters

The clusters in the MetroCluster configuration must be in a peer relationship so that they can communicate with each other and perform the data mirroring essential to MetroCluster disaster recovery.

Configuring intercluster LIFs

You must create intercluster LIFs on ports used for communication between the MetroCluster partner clusters. You can use dedicated ports or ports that also have data traffic.

Configuring intercluster LIFs on dedicated ports

You can configure intercluster LIFs on dedicated ports. Doing so typically increases the available bandwidth for replication traffic.

Step 1. List the ports in the cluster:

```
network port show
```

For complete command syntax, see the man page.

Example

The following example shows the network ports in cluster01 :

```
cluster01::> network port show
                                                          Speed (Mbps)
Node   Port      IPspace      Broadcast Domain Link   MTU    Admin/Oper
------ --------- ------------ ---------------- ----- ------- ------------
cluster01-01
       e0a       Cluster      Cluster          up     1500   auto/1000
       e0b       Cluster      Cluster          up     1500   auto/1000
       e0c       Default      Default          up     1500   auto/1000
       e0d       Default      Default          up     1500   auto/1000
       e0e       Default      Default          up     1500   auto/1000
       e0f       Default      Default          up     1500   auto/1000
cluster01-02
       e0a       Cluster      Cluster          up     1500   auto/1000
       e0b       Cluster      Cluster          up     1500   auto/1000
       e0c       Default      Default          up     1500   auto/1000
       e0d       Default      Default          up     1500   auto/1000
       e0e       Default      Default          up     1500   auto/1000
       e0f       Default      Default          up     1500   auto/1000
```

Step 2. Determine which ports are available to dedicate to intercluster communication:

```
network interface show -fields home-port,curr-port
```

For complete command syntax, see the man page.

Example

The following example shows that ports e0e and e0f have not been assigned LIFs:

```
cluster01::> network interface show -fields home-port,curr-port
vserver lif                  home-port curr-port
------- -------------------- --------- ---------
Cluster cluster01-01_clus1   e0a       e0a
Cluster cluster01-01_clus2   e0b       e0b
Cluster cluster01-02_clus1   e0a       e0a
Cluster cluster01-02_clus2   e0b       e0b
cluster01
        cluster_mgmt         e0c       e0c
cluster01
```

```
        cluster01-01_mgmt1   e0c       e0c
cluster01
        cluster01-02_mgmt1   e0c       e0c
```

Step 3. Create a failover group for the dedicated ports:
`network interface failover-groups create -vserver` *system_SVM* `-failover-group` *failover_group* `-targets` *physical_or_logical_ports*

Example

The following example assigns ports e0e and e0f to the failover group intercluster01 on the system SVM cluster01 :

```
cluster01::> network interface failover-groups create -vserver cluster01 -failover-group
intercluster01 -targets cluster01-01:e0e,cluster01-01:e0f,cluster01-02:e0e,cluster01-02:e0f
```

Step 4. Verify that the failover group was created:
`network interface failover-groups show`

For complete command syntax, see the man page.

Example

```
cluster01::> network interface failover-groups show
                                  Failover
Vserver          Group            Targets
---------------- ---------------- --------------------------------------------
Cluster
                 Cluster
                                  cluster01-01:e0a, cluster01-01:e0b,
                                  cluster01-02:e0a, cluster01-02:e0b
cluster01
                 Default
                                  cluster01-01:e0c, cluster01-01:e0d,
                                  cluster01-02:e0c, cluster01-02:e0d,
                                  cluster01-01:e0e, cluster01-01:e0f
                                  cluster01-02:e0e, cluster01-02:e0f
                 intercluster01
                                  cluster01-01:e0e, cluster01-01:e0f
                                  cluster01-02:e0e, cluster01-02:e0f
```

Step 5. Create intercluster LIFs on the system SVM and assign them to the failover group.

In ONTAP 9.6 and later:	`network interface create -vserver` *system_SVM* `-lif` *LIF_name* `-service-policy default-intercluster -home-node` *node* `-home-port` *port* `-address` *port_IP* `-netmask` *netmask* `-failover-group` *failover_group*
In ONTAP 9.5 and earlier:	`network interface create -vserver` *system_SVM* `-lif` *LIF_name* `-role intercluster -home-node` *node* `-home-port` *port* `-address` *port_IP* `-netmask` *netmask* `-failover-group` *failover_group*

For complete command syntax, see the man page.
The following example creates intercluster LIFs cluster01_icl01 and cluster01_icl02 in the failover group intercluster01 :

```
cluster01::> network interface create -vserver cluster01 -lif cluster01_icl01 -service-
policy default-intercluster -home-node cluster01-01 -home-port e0e -address 192.168.1.201
-netmask 255.255.255.0 -failover-group intercluster01

cluster01::> network interface create -vserver cluster01 -lif cluster01_icl02 -service-
policy default-intercluster -home-node cluster01-02 -home-port e0e -address 192.168.1.202
```

```
-netmask 255.255.255.0 -failover-group intercluster01
```

Step 6. Verify that the intercluster LIFs were created:

In ONTAP 9.6 and later:	`network interface show -service-policy default-intercluster`
In ONTAP 9.5 and earlier:	`network interface show -role intercluster`

For complete command syntax, see the man page.

```
cluster01::> network interface show -service-policy default-intercluster
            Logical    Status     Network            Current       Current Is
Vserver     Interface  Admin/Oper Address/Mask       Node          Port    Home
----------- ---------- ---------- ------------------ ------------- ------- ----
cluster01
            cluster01_icl01
                       up/up      192.168.1.201/24   cluster01-01  e0e     true
            cluster01_icl02
                       up/up      192.168.1.202/24   cluster01-02  e0f     true
```

Step 7. Verify that the intercluster LIFs are redundant:

In ONTAP 9.6 and later:	`network interface show -service-policy default-intercluster -failover`
In ONTAP 9.5 and earlier:	`network interface show -role intercluster -failover`

For complete command syntax, see the man page.
The following example shows that the intercluster LIFs cluster01_icl01 and cluster01_icl02 on the SVM e0e port will fail over to the e0f port.

```
cluster01::> network interface show -service-policy default-intercluster –failover
         Logical         Home                  Failover        Failover
Vserver  Interface       Node:Port             Policy          Group
-------- --------------- --------------------- --------------- --------
cluster01
         cluster01_icl01 cluster01-01:e0e   local-only      intercluster01
                         Failover Targets:  cluster01-01:e0e,
                                            cluster01-01:e0f
         cluster01_icl02 cluster01-02:e0e   local-only      intercluster01
                         Failover Targets:  cluster01-02:e0e,
                                            cluster01-02:e0f
```

Configuring intercluster LIFs on shared data ports

You can configure intercluster LIFs on ports shared with the data network. Doing so reduces the number of ports you need for intercluster networking.

Step 1. List the ports in the cluster:

```
network port show
```

For complete command syntax, see the man page.

Example

The following example shows the network ports in cluster01 :

```
cluster01::> network port show
                                                          Speed (Mbps)
```

```
Node   Port      IPspace      Broadcast Domain Link   MTU    Admin/Oper
------ --------- ------------ ---------------- ----- ------- ------------
cluster01-01
       e0a       Cluster      Cluster          up     1500   auto/1000
       e0b       Cluster      Cluster          up     1500   auto/1000
       e0c       Default      Default          up     1500   auto/1000
       e0d       Default      Default          up     1500   auto/1000
cluster01-02
       e0a       Cluster      Cluster          up     1500   auto/1000
       e0b       Cluster      Cluster          up     1500   auto/1000
       e0c       Default      Default          up     1500   auto/1000
       e0d       Default      Default          up     1500   auto/1000
```

Step 2. Create intercluster LIFs on the system SVM:

In ONTAP 9.6 and later:	`network interface create -vserver` *system_SVM* `-lif` *LIF_name* `-service-policy default-intercluster -home-node` *node* `-home-port` *port* `-address` *port_IP* `-netmask` *netmask*
In ONTAP 9.5 and earlier:	`network interface create -vserver` *system_SVM* `-lif` *LIF_name* `-role intercluster -home-node` *node* `-home-port` *port* `-address` *port_IP* `-netmask` *netmask*

For complete command syntax, see the man page.

Example

The following example creates intercluster LIFs cluster01_icl01 and cluster01_icl02 :

```
cluster01::> network interface create -vserver cluster01 -lif cluster01_icl01 -service-
policy default-intercluster -home-node cluster01-01 -home-port e0c -address 192.168.1.201
-netmask 255.255.255.0

cluster01::> network interface create -vserver cluster01 -lif cluster01_icl02 -service-
policy default-intercluster -home-node cluster01-02 -home-port e0c -address 192.168.1.202
-netmask 255.255.255.0
```

Step 3. Verify that the intercluster LIFs were created:

In ONTAP 9.6 and later:	`network interface show -service-policy default-intercluster`
In ONTAP 9.5 and earlier:	`network interface show -role intercluster`

For complete command syntax, see the man page.

Example

```
cluster01::> network interface show -service-policy default-intercluster
            Logical    Status     Network            Current       Current Is
Vserver     Interface  Admin/Oper Address/Mask       Node          Port    Home
----------- ---------- ---------- ------------------ ------------- ------- ----
cluster01
            cluster01_icl01
                       up/up      192.168.1.201/24   cluster01-01  e0c     true
            cluster01_icl02
                       up/up      192.168.1.202/24   cluster01-02  e0c     true
```

Step 4. Verify that the intercluster LIFs are redundant:

In ONTAP 9.6 and later:	`network interface show –service-policy default-intercluster -failover`
In ONTAP 9.5 and earlier:	`network interface show -role intercluster -failover`

For complete command syntax, see the man page.

Example

The following example shows that the intercluster LIFs cluster01_icl01 and cluster01_icl02 on the e0c port will fail over to the e0d port.

```
cluster01::> network interface show -service-policy default-intercluster –failover
         Logical         Home                  Failover        Failover
Vserver  Interface       Node:Port             Policy          Group
-------- --------------- --------------------- --------------- --------
cluster01
         cluster01_icl01 cluster01-01:e0c   local-only      192.168.1.201/24
                             Failover Targets: cluster01-01:e0c,
                                               cluster01-01:e0d
         cluster01_icl02 cluster01-02:e0c   local-only      192.168.1.201/24
                             Failover Targets: cluster01-02:e0c,
                                               cluster01-02:e0d
```

Creating a cluster peer relationship

You can use the `cluster peer create` command to create a peer relationship between a local and remote cluster. After the peer relationship has been created, you can run `cluster peer create` on the remote cluster to authenticate it to the local cluster.

Before you begin

You must have created intercluster LIFs on every node in the clusters that are being peered.

About this task

Step 1. On the destination cluster, create a peer relationship with the source cluster:
`cluster peer create -generate-passphrase -offer-expiration` *MM/DD/YYYY HH:MM:SS*`|1...7days|1...168hours -peer-addrs` *peer_LIF_IPs* `-ipspace` *ipspace*

If you specify both -generate-passphrase and -peer-addrs, only the cluster whose intercluster LIFs are specified in -peer-addrs can use the generated password.

You can ignore the -ipspace option if you are not using a custom IPspace. For complete command syntax, see the man page.

Example

The following example creates a cluster peer relationship on an unspecified remote cluster:

```
cluster02::> cluster peer create -generate-passphrase -offer-expiration 2days

                    Passphrase: UCa+6lRVICXeL/gq1WrK7ShR
              Expiration Time: 6/7/2017 08:16:10 EST
  Initial Allowed Vserver Peers: -
            Intercluster LIF IP: 192.140.112.101
              Peer Cluster Name: Clus_7ShR (temporary generated)
```

```
Warning: make a note of the passphrase - it cannot be displayed again.
```

Step 2. On source cluster, authenticate the source cluster to the destination cluster:
`cluster peer create -peer-addrs` *peer_LIF_IPs* `-ipspace` *ipspace*

For complete command syntax, see the man page.

Example

The following example authenticates the local cluster to the remote cluster at intercluster LIF IP addresses 192.140.112.101 and 192.140.112.102:

```
cluster01::> cluster peer create -peer-addrs 192.140.112.101,192.140.112.102

Notice: Use a generated passphrase or choose a passphrase of 8 or more characters.
        To ensure the authenticity of the peering relationship, use a phrase or sequence of
        characters that would be hard to guess.

Enter the passphrase:
Confirm the passphrase:

Clusters cluster02 and cluster01 are peered.
```

Enter the passphrase for the peer relationship when prompted.

Step 3. Verify that the cluster peer relationship was created:
`cluster peer show -instance`

Example

```
cluster01::> cluster peer show -instance

                                Peer Cluster Name: cluster02
                    Remote Intercluster Addresses: 192.140.112.101, 192.140.112.102
               Availability of the Remote Cluster: Available
                                Remote Cluster Name: cluster2
                                Active IP Addresses: 192.140.112.101, 192.140.112.102
                              Cluster Serial Number: 1-80-123456
                    Address Family of Relationship: ipv4
             Authentication Status Administrative: no-authentication
                Authentication Status Operational: absent
                                 Last Update Time: 02/05 21:05:41
                      IPspace for the Relationship: Default
```

Step 4. Check the connectivity and status of the nodes in the peer relationship:
`cluster peer health show`

Example

```
cluster01::> cluster peer health show
Node       cluster-Name                Node-Name
             Ping-Status               RDB-Health Cluster-Health  Avail…
---------- --------------------------- ---------  --------------- --------
cluster01-01
           cluster02                   cluster02-01
             Data: interface_reachable
             ICMP: interface_reachable true       true            true
                                         cluster02-02
             Data: interface_reachable
             ICMP: interface_reachable true       true            true
cluster01-02
```

```
cluster02                    cluster02-01
  Data: interface_reachable
  ICMP: interface_reachable true       true            true
                              cluster02-02
  Data: interface_reachable
  ICMP: interface_reachable true       true            true
```

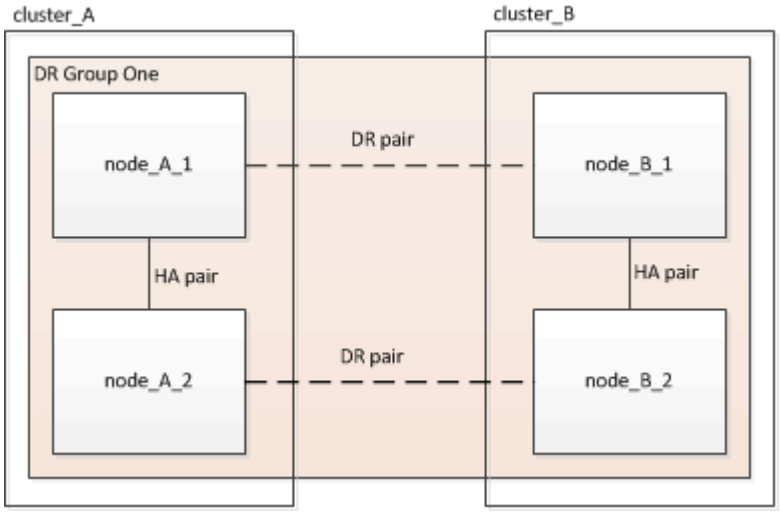
Creating the DR group

You must create the disaster recovery (DR) group relationships between the clusters.

About this task

You perform this procedure on one of the clusters in the MetroCluster configuration to create the DR relationships between the nodes in both clusters.

Note: The DR relationships cannot be changed after the DR groups are created.



Step 1. Verify that the nodes are ready for creation of the DR group by entering the following command on each:
`metrocluster configuration-settings show-status`

Example

The command output should show that the nodes are ready:

```
cluster_A::> metrocluster configuration-settings show-status
Cluster                   Node          Configuration Settings Status
------------------------- ------------- --------------------------------
cluster_A                 node_A_1      ready for DR group create
                           node_A_2      ready for DR group create
2 entries were displayed.
cluster_B::> metrocluster configuration-settings show-status
Cluster                   Node          Configuration Settings Status
------------------------- ------------- --------------------------------
cluster_B                 node_B_1      ready for DR group create
                           node_B_2      ready for DR group create
2 entries were displayed.
```

Step 2. Create the DR group:
`metrocluster configuration-settings dr-group create -partner-cluster` *partner-cluster-name* `-local-node` *local-node-name* `-remote-node` *remote-node-name*

This command is issued only once. It does not need to be repeated on the partner cluster. In the command, you specify the name of the remote cluster and the name of one local node and one node on the partner cluster.

The two nodes you specify are configured as DR partners and the other two nodes (which are not specified in the command) are configured as the second DR pair in the DR group. These relationships cannot be changed after you enter this command.

The following command creates these DR pairs:

- node_A_1 and node_B_1
- node_A_2 and node_B_2

```
Cluster_A::> metrocluster configuration-settings dr-group create -partner-cluster cluster_B
-local-node node_A_1 -remote-node node_B_1
[Job 27] Job succeeded: DR Group Create is successful.
```

Configuring and connecting the MetroCluster IP interfaces

You must configure the MetroCluster IP (MCCIP) interfaces that are used for replication of each node's storage and nonvolatile cache. You then establish the connections using the MCCIP interfaces. This creates iSCSI connections for storage replication.

About this task

Note: You must choose the MetroCluster IP addresses carefully because you cannot change them after initial configuration. "Considerations for MetroCluster IP configuration" on page 4

You must create two interfaces for each node. The interfaces must be associated with the VLANs defined in the MetroCluster RCF file.

The following IP addresses and subnets are used in the examples:

Node	Interface	IP address	Subnet
node_A_1	MetroCluster IP interface 1	10.1.1.1	10.1.1/24
	MetroCluster IP interface 2	10.1.2.1	10.1.2/24
node_A_2	MetroCluster IP interface 1	10.1.1.2	10.1.1/24
	MetroCluster IP interface 2	10.1.2.2	10.1.2/24
node_B_1	MetroCluster IP interface 1	10.1.1.3	10.1.1/24
	MetroCluster IP interface 2	10.1.2.3	10.1.2/24
node_B_2	MetroCluster IP interface 1	10.1.1.4	10.1.1/24
	MetroCluster IP interface 2	10.1.2.4	10.1.2/24

- AFA DM7000F and DM7000H systems use ports e1a and e1b for the MetroCluster IP interfaces.
- AFA DM5000F and DM5000H systems use VLAN ports e0a-10 and e0b-20 for the MetroCluster IP interfaces.

  These physical ports are also used as cluster interfaces. The VLANs are configured automatically.

Step 1. Confirm that each node has disk autoassignment enabled:

```
storage disk option show
```

Disk autoassignment will assign pool 0 and pool 1 disks on a shelf-by-shelf basis.

The `Auto Assign` column indicates whether disk autoassignment is enabled.

```
Node         BKg. FW. Upd.  Auto Copy   Auto Assign  Auto Assign Policy
-----------  -------------  ----------  -----------  ------------------
node_A_1              on           on           on           default
node_A_2              on           on           on           default
2 entries were displayed.
```

Step 2. Verify you can create MetroCluster IP interfaces on the nodes:
`metrocluster configuration-settings show-status`

Example

All nodes should be ready:

```
Cluster    Node Configuration Settings Status
----------    -----------  --------------------------------
cluster_A
node_A_1  ready for interface create
node_A_2  ready for interface create
cluster_B
node_B_1  ready for interface create
node_B_2  ready for interface create
4 entries were displayed.
```

Step 3. Create the interfaces on node_A_1.

a. Configure the interface on port e5a on node_A_1:
`metrocluster configuration-settings interface create -cluster-name` *cluster-name* `-home-node` *node-name* `-home-port e5a -address` *ip-address* `-netmask` *netmask*

Example

The following example shows the creation of the interface on port e5a on node_A_1 with IP address 10.1.1.1:

```
cluster_A::> metrocluster configuration-settings interface create -cluster-name cluster_A
-home-node node_A_1
-home-port e1a -address 10.1.1.1 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

b. Configure the interface on port e5b on node_A_1:
`metrocluster configuration-settings interface create -cluster-name` *cluster-name* `-home-node` *node-name* `-home-port e5b -address` *ip-address* `-netmask` *netmask*

Example

The following example shows the creation of the interface on port e5b on node_A_1 with IP address 10.1.2.1:

```
cluster_A::> metrocluster configuration-settings interface create -cluster-name cluster_A
-home-node node_A_1
-home-port e1b -address 10.1.2.1 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

Note: You can verify that these interfaces are present using the `metrocluster configuration-settings interface show` command.

Step 4. Create the interfaces on node_A_2.

a. Configure the interface on port e5a on node_A_2:

`metrocluster configuration-settings interface create -cluster-name` *cluster-name* `-home-node` *node-name* `-home-port e5a -address` *ip-address* `-netmask` *netmask*

Example

The following example shows the creation of the interface on port e5a on node_A_2 with IP address 10.1.1.2:

```
cluster_A::> metrocluster configuration-settings interface create -cluster-name cluster_A
-home-node node_A_2
-home-port e1a -address 10.1.1.2 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

b. Configure the interface on port e5b on node_A_2:
`metrocluster configuration-settings interface create -cluster-name` *cluster-name* `-home-node` *node-name* `-home-port e5b -address` *ip-address* `-netmask` *netmask*

Example

The following example shows the creation of the interface on port e5b on node_A_2 with IP address 10.1.2.2:

```
cluster_A::> metrocluster configuration-settings interface create -cluster-name cluster_A
-home-node node_A_2
-home-port e15b -address 10.1.2.2 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

Step 5. Create the interfaces on node_B_1.

a. Configure the interface on port e5a on node_B_1:
`metrocluster configuration-settings interface create -cluster-name`*cluster-name*`-home-node`*node-name*`-home-port e5a -address`*ip-address*`-netmask` *netmask*

Example

The following example shows the creation of the interface on port e5a on node_B_1 with IP address 10.1.1.3:

```
cluster_A::> metrocluster configuration-settings interface create -cluster-name cluster_A -home-node
 node_B_1 -home-port e5a -address 10.1.1.3 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.cluster_A::>
```

b. Configure the interface on port e5b on node_B_1:
`metrocluster configuration-settings interface create -cluster-name` *cluster-name* `-home-node` *node-name* `-home-port e5a -address`*ip-address* `-netmask` *netmask*

Example

The following example shows the creation of the interface on port e5b on node_B_1 with IP address 10.1.2.3:

```
cluster_A::> metrocluster configuration-settings interface create -cluster-name cluster_A -home-node node_B_1
 -home-port e5b -address 10.1.2.3 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.cluster_A::>
```

Step 6. Create the interfaces on node_B_2.

Note: The ports used in the following examples are e5a and e5b. You must configure the interfaces on the correct ports for your platform model, as given in above.

a. Configure the interface on port e5a on node_B_2:
`metrocluster configuration-settings interface create -cluster-name` *cluster-name* `-home-node` *node-name* `-home-port e5a -addressip-addresip-addresss` *ip-address* `-netmask` *netmask*

Example

The following example shows the creation of the interface on port e5a on node_B_2 with IP address 10.1.1.4:

```
cluster_B::>metrocluster configuration-settings interface create -cluster-name cluster_B -home-node node_B_2
 -home-port e5a -address 10.1.1.4 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.cluster_A::>
```

b. Configure the interface on port e5b on node_B_2:
`metrocluster configuration-settings interface create -cluster-name` *cluster-name* `-home-node` *node-name* `-home-port e5b -address` *ip-address* `-netmask` *netmask*

Example

The following example shows the creation of the interface on port e5b on node_B_2 with IP address 10.1.2.4:

```
cluster_B::> metrocluster configuration-settings interface create -cluster-name cluster_B -home-node node_B_2
 -home-port e5b -address 10.1.2.4 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

Step 7. Verify that the interfaces have been configured:
`metrocluster configuration-settings interface show`

Example

The following example shows that the configuration state for each interface is completed.

```
cluster_A::> metrocluster configuration-settings interface show
DR                                                               Config
Group Cluster Node    Network Address Netmask         Gateway   State
----- ------- ------- --------------- --------------- --------- ----------
1     cluster_A  node_A_1
                  Home Port: e1a
                       10.1.1.1     255.255.255.0   -         completed
                  Home Port: e1b
                       10.1.2.1     255.255.255.0   -         completed
                  node_A_2
                  Home Port: e1a
                       10.1.1.2     255.255.255.0   -         completed
                  Home Port: e1b
                       10.1.2.2     255.255.255.0   -         completed
      cluster_B  node_B_1
                  Home Port: e1a
                       10.1.1.3     255.255.255.0   -         completed
                  Home Port: e1b
                       10.1.2.3     255.255.255.0   -         completed
                  node_B_2
                  Home Port: e1a
                       10.1.1.4     255.255.255.0   -         completed
                  Home Port: e1b
                       10.1.2.4     255.255.255.0   -         completed
8 entries were displayed.
cluster_A::>
```

Step 8. Verify that the nodes are ready to connect the MetroCluster interfaces:
`metrocluster configuration-settings show-status`

Example

The following example shows all nodes in the ready for connection state:

```
Cluster     Node        Configuration Settings Status
----------    -----------  --------------------------------
cluster_A
node_A_1  ready for connection connect
node_A_2  ready for connection connect
cluster_B
node_B_1  ready for connection connect
node_B_2  ready for connection connect
4 entries were displayed.
```

Step 9. Establish the connections:
`metrocluster configuration-settings connection connect`

The IP addresses cannot be changed after you issue this command.

Example

The following example shows cluster_A is successfully connected:

```
cluster_A::> metrocluster configuration-settings connection connect
[Job 53] Job succeeded: Connect is successful.
cluster_A::>
```

Step 10. Verify that the connections have been established:
`metrocluster configuration-settings show-status`

Example

The configuration settings status for all nodes should be completed:

```
Cluster    Node     Configuration Settings Status
----------    -----------  --------------------------------
cluster_A
node_A_1  completed
node_A_2  completed
cluster_B
node_B_1  completed
node_B_2  completed
4 entries were displayed.
```

Step 11. Verify that the iSCSI connections have been established:

a. Change to the advanced privilege level:
`set -privilege advanced`

You need to respond with **y** when you are prompted to continue into advanced mode and you see the advanced mode prompt (***>**).

b. Display the connections:
`storage iscsi-initiator show`

On systems running ONTAP 9.5, there are eight MCCIP initiators on each cluster that should appear in the output.

On systems running ONTAP 9.4 and earlier, there are four MCCIP initiators on each cluster that should appear in the output.

Example

The following example shows the eight MCCIP initiators on a cluster running ONTAP 9.5:

```
cluster_A::*> storage iscsi-initiator show
Node Type Label    Target Portal           Target Name                      Admin/Op
---- ---- -------- ------------------      --------------------------------- --------

cluster_A-01
     dr_auxiliary
               mccip-aux-a-initiator
                    10.227.16.113:65200     prod506.com.company:abab44       up/up
               mccip-aux-a-initiator2
                    10.227.16.113:65200     prod507.com.company:abab44       up/up
               mccip-aux-b-initiator
                    10.227.95.166:65200     prod506.com.company:abab44       up/up
               mccip-aux-b-initiator2
                    10.227.95.166:65200     prod507.com.company:abab44       up/up
     dr_partner
               mccip-pri-a-initiator
                    10.227.16.112:65200     prod506.com.company:cdcd88       up/up
               mccip-pri-a-initiator2
                    10.227.16.112:65200     prod507.com.company:cdcd88       up/up
               mccip-pri-b-initiator
                    10.227.95.165:65200     prod506.com.company:cdcd88       up/up
               mccip-pri-b-initiator2
                    10.227.95.165:65200     prod507.com.company:cdcd88       up/up
cluster_A-02
     dr_auxiliary
               mccip-aux-a-initiator
                    10.227.16.112:65200     prod506.com.company:cdcd88       up/up
               mccip-aux-a-initiator2
                    10.227.16.112:65200     prod507.com.company:cdcd88       up/up
               mccip-aux-b-initiator
                    10.227.95.165:65200     prod506.com.company:cdcd88       up/up
               mccip-aux-b-initiator2
                    10.227.95.165:65200     prod507.com.company:cdcd88       up/up
     dr_partner
               mccip-pri-a-initiator
                    10.227.16.113:65200     prod506.com.company:abab44       up/up
               mccip-pri-a-initiator2
                    10.227.16.113:65200     prod507.com.company:abab44       up/up
               mccip-pri-b-initiator
                    10.227.95.166:65200     prod506.com.company:abab44       up/up
               mccip-pri-b-initiator2
                    10.227.95.166:65200     prod507.com.company:abab44       up/up
16 entries were displayed.
```

c. Return to the admin privilege level:
   `set -privilege admin`

Step 12. Verify that the nodes are ready for final implementation of the MetroCluster configuration:
`metrocluster node show`

Example

```
cluster_A::> metrocluster node show
DR                              Configuration  DR
Group Cluster Node              State          Mirroring Mode
----- ------- ----------------- -------------- --------- ----
-     cluster_A
              node_A_1          ready to configure -     -
```

```
              node_A_2           ready to configure -      -
2 entries were displayed.
cluster_A::>
cluster_B::> metrocluster node show
DR                               Configuration  DR
Group Cluster Node               State          Mirroring Mode
----- ------- ------------------ -------------- --------- ----
-     cluster_B
              node_B_1           ready to configure -      -
              node_B_2           ready to configure -      -
2 entries were displayed.
cluster_B::>
```

Verifying or manually performing pool 1 drives assignment

Depending on the storage configuration, you must either verify pool 1 drive assignment or manually assign drives to pool 1 for each node in the MetroCluster IP configuration. The procedure you use depends on the version of ONTAP you are using.

About this task

Configuration type	Procedure
The configuration includes either three shelves, or, if it contains more than four shelves, has an uneven multiple of four shelves (for example, seven shelves), and is running ONTAP 9.5.	“Manually assigning drives for pool 1 (ONTAP 9.4 or later)” on page 88
The configuration does not include four storage shelves per site and is running ONTAP 9.4	“Manually assigning drives for pool 1 (ONTAP 9.4 or later)” on page 88

Verifying disk assignment for pool 1 disks

You must verify that the remote disks are visible to the nodes and have been assigned correctly.

Before you begin

You must wait at least ten minutes for disk auto-assignment to complete after the MetroCluster IP interfaces and connections were created with the `metrocluster configuration-settings connection connect` command.

About this task

Command output will show disk names in the form: *node-name*:0m.i1.0L1

“Considerations for automatic drive assignment and ADP systems” on page 6

Step 1. Verify pool 1 disks are auto-assigned:
`disk show`

Example

Drive autoassignment has assigned one quarter (8 drives) to node_A_1 and one quarter to node_A_2. The remaining drives will be remote (pool1) disks for node_B_1 and node_B_2.

```
cluster_B::> diskshow-host-adapter 0m -owner node_B_2
                    Usable      Disk              Container   Container
Disk                Size        Shelf Bay Type    Type        Name      Owner
----------------    ----------  ----- --- ------- ----------- --------- --------
node_B_2:0m.i0.2L4  894.0GB     0     29  SSD-NVM shared      -         node_B_2
```

```
node_B_2:0m.i0.2L10 894.0GB    0     25  SSD-NVM shared      -          node_B_2
node_B_2:0m.i0.3L3  894.0GB    0     28  SSD-NVM shared      -          node_B_2
node_B_2:0m.i0.3L9  894.0GB    0     24  SSD-NVM shared      -          node_B_2
node_B_2:0m.i0.3L11 894.0GB    0     26  SSD-NVM shared      -          node_B_2
node_B_2:0m.i0.3L12 894.0GB    0     27  SSD-NVM shared      -          node_B_2
node_B_2:0m.i0.3L15 894.0GB    0     30  SSD-NVM shared      -          node_B_2
node_B_2:0m.i0.3L16 894.0GB    0     31  SSD-NVM shared      -          node_B_2
8 entries were displayed.

cluster_B::> disk show -host-adapter 0m -owner node_B_1
                     Usable     Disk              Container   Container
Disk                 Size       Shelf Bay Type    Type        Name      Owner
-----------------    ---------- ----- --- ------- ----------- --------- --------
node_B_1:0m.i2.3L19 1.75TB     0     42  SSD-NVM shared      -          node_B_1
node_B_1:0m.i2.3L20 1.75TB     0     43  SSD-NVM spare       Pool1     node_B_1
node_B_1:0m.i2.3L23 1.75TB     0     40  SSD-NVM shared       -         node_B_1
node_B_1:0m.i2.3L24 1.75TB     0     41  SSD-NVM spare       Pool1     node_B_1
node_B_1:0m.i2.3L29 1.75TB     0     36  SSD-NVM shared       -         node_B_1
node_B_1:0m.i2.3L30 1.75TB     0     37  SSD-NVM shared       -         node_B_1
node_B_1:0m.i2.3L31 1.75TB     0     38  SSD-NVM shared       -         node_B_1
node_B_1:0m.i2.3L32 1.75TB     0     39  SSD-NVM shared       -         node_B_1
8 entries were displayed.

cluster_B::> disk show
                     Usable     Disk              Container   Container
Disk                 Size       Shelf Bay Type    Type        Name      Owner
-----------------    ---------- ----- --- ------- ----------- --------- --------
node_B_1:0m.i1.0L6  1.75TB     0     1   SSD-NVM shared      -          node_A_2
node_B_1:0m.i1.0L8  1.75TB     0     3   SSD-NVM shared      -          node_A_2
node_B_1:0m.i1.0L17 1.75TB     0     18  SSD-NVM shared      -          node_A_1
node_B_1:0m.i1.0L22 1.75TB     0     17 SSD-NVM shared - node_A_1
node_B_1:0m.i1.0L25 1.75TB     0     12 SSD-NVM shared - node_A_1
node_B_1:0m.i1.2L2  1.75TB     0     5 SSD-NVM shared - node_A_2
node_B_1:0m.i1.2L7  1.75TB     0     2 SSD-NVM shared - node_A_2
node_B_1:0m.i1.2L14 1.75TB     0     7 SSD-NVM shared - node_A_2
node_B_1:0m.i1.2L21 1.75TB     0     16 SSD-NVM shared - node_A_1
node_B_1:0m.i1.2L27 1.75TB     0     14 SSD-NVM shared - node_A_1
node_B_1:0m.i1.2L28 1.75TB     0     15 SSD-NVM shared - node_A_1
node_B_1:0m.i2.1L1  1.75TB     0     4 SSD-NVM shared - node_A_2
node_B_1:0m.i2.1L5  1.75TB     0     0 SSD-NVM shared - node_A_2
node_B_1:0m.i2.1L13 1.75TB     0     6 SSD-NVM shared - node_A_2
node_B_1:0m.i2.1L18 1.75TB     0     19 SSD-NVM shared - node_A_1
node_B_1:0m.i2.1L26 1.75TB     0     13 SSD-NVM shared - node_A_1
node_B_1:0m.i2.3L19 1.75TB     0 42 SSD-NVM shared - node_B_1
node_B_1:0m.i2.3L20 1.75TB     0 43 SSD-NVM shared - node_B_1
node_B_1:0m.i2.3L23 1.75TB     0 40 SSD-NVM shared - node_B_1
node_B_1:0m.i2.3L24 1.75TB     0 41 SSD-NVM shared - node_B_1
node_B_1:0m.i2.3L29 1.75TB     0 36 SSD-NVM shared - node_B_1
node_B_1:0m.i2.3L30 1.75TB     0 37 SSD-NVM shared - node_B_1
node_B_1:0m.i2.3L31 1.75TB     0 38 SSD-NVM shared - node_B_1
node_B_1:0m.i2.3L32 1.75TB     0 39 SSD-NVM shared - node_B_1
node_B_1:0n.12      1.75TB     0 12 SSD-NVM shared aggr0 node_B_1
node_B_1:0n.13      1.75TB     0 13 SSD-NVM shared aggr0 node_B_1
node_B_1:0n.14      1.75TB     0 14 SSD-NVM shared aggr0 node_B_1
node_B_1:0n.15      1.75TB 0 15 SSD-NVM shared aggr0 node_B_1
node_B_1:0n.16      1.75TB 0 16 SSD-NVM shared aggr0 node_B_1
node_B_1:0n.17      1.75TB 0 17 SSD-NVM shared aggr0 node_B_1
node_B_1:0n.18      1.75TB 0 18 SSD-NVM shared aggr0 node_B_1
node_B_1:0n.19      1.75TB 0 19 SSD-NVM shared - node_B_1
node_B_1:0n.24      894.0GB 0 24 SSD-NVM shared - node_A_2
node_B_1:0n.25      894.0GB 0 25 SSD-NVM shared - node_A_2
node_B_1:0n.26      894.0GB 0 26 SSD-NVM shared - node_A_2
node_B_1:0n.27      894.0GB 0 27 SSD-NVM shared - node_A_2
node_B_1:0n.28      894.0GB 0 28 SSD-NVM shared - node_A_2
node_B_1:0n.29      894.0GB 0 29 SSD-NVM shared - node_A_2
node_B_1:0n.30      894.0GB 0 30 SSD-NVM shared - node_A_2
node_B_1:0n.31      894.0GB 0 31 SSD-NVM shared - node_A_2
node_B_1:0n.36      1.75TB 0 36 SSD-NVM shared - node_A_1
```

```
node_B_1:0n.37      1.75TB 0 37 SSD-NVM shared - node_A_1
node_B_1:0n.38      1.75TB 0 38 SSD-NVM shared - node_A_1
node_B_1:0n.39      1.75TB 0 39 SSD-NVM shared - node_A_1
node_B_1:0n.40      1.75TB 0 40 SSD-NVM shared - node_A_1
node_B_1:0n.41      1.75TB 0 41 SSD-NVM shared - node_A_1
node_B_1:0n.42      1.75TB 0 42 SSD-NVM shared - node_A_1
node_B_1:0n.43      1.75TB 0 43 SSD-NVM shared - node_A_1
node_B_2:0m.i0.2L4  894.0GB 0 29 SSD-NVM shared - node_B_2
node_B_2:0m.i0.2L10 894.0GB 0 25 SSD-NVM shared - node_B_2
node_B_2:0m.i0.3L3  894.0GB 0 28 SSD-NVM shared - node_B_2
node_B_2:0m.i0.3L9  894.0GB 0 24 SSD-NVM shared - node_B_2
node_B_2:0m.i0.3L11 894.0GB 0 26 SSD-NVM shared - node_B_2
node_B_2:0m.i0.3L12 894.0GB 0 27 SSD-NVM shared - node_B_2
node_B_2:0m.i0.3L15 894.0GB 0 30 SSD-NVM shared - node_B_2
node_B_2:0m.i0.3L16 894.0GB 0 31 SSD-NVM shared - node_B_2
node_B_2:0n.0       1.75TB 0 0 SSD-NVM shared aggr0_rha12_b1_cm_02_0 node_B_2
node_B_2:0n.1 1.75TB 0 1 SSD-NVM shared aggr0_rha12_b1_cm_02_0 node_B_2
node_B_2:0n.2 1.75TB 0 2 SSD-NVM shared aggr0_rha12_b1_cm_02_0 node_B_2
node_B_2:0n.3 1.75TB 0 3 SSD-NVM shared aggr0_rha12_b1_cm_02_0 node_B_2
node_B_2:0n.4 1.75TB 0 4 SSD-NVM shared aggr0_rha12_b1_cm_02_0 node_B_2
node_B_2:0n.5 1.75TB 0 5 SSD-NVM shared aggr0_rha12_b1_cm_02_0 node_B_2
node_B_2:0n.6 1.75TB 0 6 SSD-NVM shared aggr0_rha12_b1_cm_02_0 node_B_2
node_B_2:0n.7 1.75TB 0 7 SSD-NVM shared - node_B_2
64 entries were displayed.

cluster_B::>


cluster_A::> disk show
Usable Disk Container Container
Disk Size Shelf Bay Type Type Name Owner
---------------- ---------- ----- --- ------- ----------- --------- --------
node_A_1:0m.i1.0L2 1.75TB 0 5 SSD-NVM shared - node_B_2
node_A_1:0m.i1.0L8 1.75TB 0 3 SSD-NVM shared - node_B_2
node_A_1:0m.i1.0L18 1.75TB 0 19 SSD-NVM shared - node_B_1
node_A_1:0m.i1.0L25 1.75TB 0 12 SSD-NVM shared - node_B_1
node_A_1:0m.i1.0L27 1.75TB 0 14 SSD-NVM shared - node_B_1
node_A_1:0m.i1.2L1 1.75TB 0 4 SSD-NVM shared - node_B_2
node_A_1:0m.i1.2L6 1.75TB 0 1 SSD-NVM shared - node_B_2
node_A_1:0m.i1.2L7 1.75TB 0 2 SSD-NVM shared - node_B_2
node_A_1:0m.i1.2L14 1.75TB 0 7 SSD-NVM shared - node_B_2
node_A_1:0m.i1.2L17 1.75TB 0 18 SSD-NVM shared - node_B_1
node_A_1:0m.i1.2L22 1.75TB 0 17 SSD-NVM shared - node_B_1
node_A_1:0m.i2.1L5 1.75TB 0 0 SSD-NVM shared - node_B_2
node_A_1:0m.i2.1L13 1.75TB 0 6 SSD-NVM shared - node_B_2
node_A_1:0m.i2.1L21 1.75TB 0 16 SSD-NVM shared - node_B_1
node_A_1:0m.i2.1L26 1.75TB 0 13 SSD-NVM shared - node_B_1
node_A_1:0m.i2.1L28 1.75TB 0 15 SSD-NVM shared - node_B_1
node_A_1:0m.i2.3L19 1.75TB 0 42 SSD-NVM shared - node_A_1
node_A_1:0m.i2.3L20 1.75TB 0 43 SSD-NVM shared - node_A_1
node_A_1:0m.i2.3L23 1.75TB 0 40 SSD-NVM shared - node_A_1
node_A_1:0m.i2.3L24 1.75TB 0 41 SSD-NVM shared - node_A_1
node_A_1:0m.i2.3L29 1.75TB 0 36 SSD-NVM shared - node_A_1
node_A_1:0m.i2.3L30 1.75TB 0 37 SSD-NVM shared - node_A_1
node_A_1:0m.i2.3L31 1.75TB 0 38 SSD-NVM shared - node_A_1
node_A_1:0m.i2.3L32 1.75TB 0 39 SSD-NVM shared - node_A_1
node_A_1:0n.12 1.75TB 0 12 SSD-NVM shared aggr0 node_A_1
node_A_1:0n.13 1.75TB 0 13 SSD-NVM shared aggr0 node_A_1
node_A_1:0n.14 1.75TB 0 14 SSD-NVM shared aggr0 node_A_1
node_A_1:0n.15 1.75TB 0 15 SSD-NVM shared aggr0 node_A_1
node_A_1:0n.16 1.75TB 0 16 SSD-NVM shared aggr0 node_A_1
node_A_1:0n.17 1.75TB 0 17 SSD-NVM shared aggr0 node_A_1
node_A_1:0n.18 1.75TB 0 18 SSD-NVM shared aggr0 node_A_1
node_A_1:0n.19 1.75TB 0 19 SSD-NVM shared - node_A_1
node_A_1:0n.24 894.0GB 0 24 SSD-NVM shared - node_B_2
node_A_1:0n.25 894.0GB 0 25 SSD-NVM shared - node_B_2
node_A_1:0n.26 894.0GB 0 26 SSD-NVM shared - node_B_2
node_A_1:0n.27 894.0GB 0 27 SSD-NVM shared - node_B_2
```

```
node_A_1:0n.28 894.0GB 0 28 SSD-NVM shared - node_B_2
node_A_1:0n.29 894.0GB 0 29 SSD-NVM shared - node_B_2
node_A_1:0n.30 894.0GB 0 30 SSD-NVM shared - node_B_2
node_A_1:0n.31 894.0GB 0 31 SSD-NVM shared - node_B_2
node_A_1:0n.36 1.75TB 0 36 SSD-NVM shared - node_B_1
node_A_1:0n.37 1.75TB 0 37 SSD-NVM shared - node_B_1
node_A_1:0n.38 1.75TB 0 38 SSD-NVM shared - node_B_1
node_A_1:0n.39 1.75TB 0 39 SSD-NVM shared - node_B_1
node_A_1:0n.40 1.75TB 0 40 SSD-NVM shared - node_B_1
node_A_1:0n.41 1.75TB 0 41 SSD-NVM shared - node_B_1
node_A_1:0n.42 1.75TB 0 42 SSD-NVM shared - node_B_1
node_A_1:0n.43 1.75TB 0 43 SSD-NVM shared - node_B_1
node_A_2:0m.i2.3L3 894.0GB 0 28 SSD-NVM shared - node_A_2
node_A_2:0m.i2.3L4 894.0GB 0 29 SSD-NVM shared - node_A_2
node_A_2:0m.i2.3L9 894.0GB 0 24 SSD-NVM shared - node_A_2
node_A_2:0m.i2.3L10 894.0GB 0 25 SSD-NVM shared - node_A_2
node_A_2:0m.i2.3L11 894.0GB 0 26 SSD-NVM shared - node_A_2
node_A_2:0m.i2.3L12 894.0GB 0 27 SSD-NVM shared - node_A_2
node_A_2:0m.i2.3L15 894.0GB 0 30 SSD-NVM shared - node_A_2
node_A_2:0m.i2.3L16 894.0GB 0 31 SSD-NVM shared - node_A_2
node_A_2:0n.0 1.75TB 0 0 SSD-NVM shared aggr0_node_A_2_0 node_A_2
node_A_2:0n.1 1.75TB 0 1 SSD-NVM shared aggr0_node_A_2_0 node_A_2
node_A_2:0n.2 1.75TB 0 2 SSD-NVM shared aggr0_node_A_2_0 node_A_2
node_A_2:0n.3 1.75TB 0 3 SSD-NVM shared aggr0_node_A_2_0 node_A_2
node_A_2:0n.4 1.75TB 0 4 SSD-NVM shared aggr0_node_A_2_0 node_A_2
node_A_2:0n.5 1.75TB 0 5 SSD-NVM shared aggr0_node_A_2_0 node_A_2
node_A_2:0n.6 1.75TB 0 6 SSD-NVM shared aggr0_node_A_2_0 node_A_2
node_A_2:0n.7 1.75TB 0 7 SSD-NVM shared - node_A_2
64 entries were displayed.

cluster_A::>
```

Manually assigning drives for pool 1 (ONTAP 9.4 or later)

If the system was not preconfigured at the factory and does not meet the requirements for automatic drive assignment, you must manually assign the remote pool 1 drives.

About this task

This procedure applies to configurations running ONTAP 9.4 or later.

Details for determining whether your system requires manual disk assignment are included in "Considerations for automatic drive assignment and ADP systems" on page 6.

If the configuration has fewer than four external shelves per site, you must check that automatic drive assignment on all nodes is disabled and manually assign the drives.

When the configuration includes only two external shelves per site,pool 1 drives for each site should be shared from the same shelf as shown in the following examples:

- node_A_1 is assigned drives inbays 0-11 on site_B-shelf_2 (remote)
- node_A_2 is assigned drives inbays 12-23onsite_B-shelf_2 (remote)

Step 1. From each node in the MetroCluster IP configuration, assign remote drives to pool 1.

a. Display the list of unassigned drives:

```
disk show -host-adapter 0m -container-type unassigned
```

Example

```
cluster_A::> disk show -host-adapter 0m -container-type unassigned
                     Usable           Disk    Container   Container
Disk                   Size Shelf Bay Type    Type        Name      Owner
---------------- ---------- ----- --- ------- ----------- --------- --------
```

```
6.23.0                    -    23   0 SSD     unassigned  -          -
6.23.1                    -    23   1 SSD     unassigned  -          -
.
.
.
node_A_2:0m.i1.2L51       -    21  14 SSD     unassigned  -          -
node_A_2:0m.i1.2L64       -    21  10 SSD     unassigned  -          -
.
.
.
48 entries were displayed.

cluster_A::>
```

b. Assign ownership of remote drives (0m) to pool 1 of the first node (for example, node_A_1):
`disk assign -disk` *disk-id* `-pool 1 -owner` *owner-node-name*

disk-id must identify a drive on a remote shelf of *owner-node-name*.

c. Confirm that the drives were assigned to pool 1:
`disk show -host-adapter 0m -container-type unassigned`

Note: The iSCSI connection used to access the remote drives appears as device 0m.

Example

The following output shows that the drives on shelf 23 were assigned because they no longer appear in the list of unassigned drives:

```
cluster_A::> disk show -host-adapter 0m -container-type unassigned
                      Usable           Disk    Container   Container
Disk                    Size Shelf Bay Type    Type        Name      Owner
---------------- ---------- ----- --- ------- ----------- --------- --------
node_A_2:0m.i1.2L51       -    21  14 SSD     unassigned  -          -
node_A_2:0m.i1.2L64       -    21  10 SSD     unassigned  -          -
.
.
.
node_A_2:0m.i2.1L90       -    21  19 SSD     unassigned  -          -
24 entries were displayed.

cluster_A::>
```

d. Repeat these steps to assign pool 1 drives to the second node on site A (for example, node_A_2).

e. Repeat these steps on site B.

Enabling automatic drive assignment in ONTAP 9.4

In ONTAP 9.4, if you disabled automatic drive assignment as directed previously in this procedure, you must reenable it on all nodes.

About this task

"Considerations for automatic drive assignment and ADP systems" on page 6

Step 1. Enable automatic drive assignment:
`storage disk option modify -node` *node_name* `-autoassign on`

You must issue this command on all nodes in the MetroCluster IP configuration.

Mirroring the root aggregates

You must mirror the root aggregates to provide data protection.

About this task

By default, the root aggregate is created as RAID-DP type aggregate. You can change the root aggregate from RAID-DP to RAID4 type aggregate. The following command modifies the root aggregate for RAID4 type aggregate:

```
storage aggregate modify –aggregate aggr_name -raidtype raid4
```

Note: On non-ADP systems, the RAID type of the aggregate can be modified from the default RAID-DP to RAID4 before or after the aggregate is mirrored.

Step 1. Mirror the root aggregate:
`storage aggregate mirror` *aggr_name*
The following command mirrors the root aggregate for controller_A_1:
```
controller_A_1::> storage aggregate mirror aggr0_controller_A_1
```
This mirrors the aggregate, so it consists of a local plex and a remote plex located at the remote MetroCluster site.

Step 2. Repeat the previous step for each node in the MetroCluster configuration.

Creating a mirrored data aggregate on each node

You must create a mirrored data aggregate on each node in the DR group.

Before you begin

- You should know what drives will be used in the new aggregate.
- If you have multiple drive types in your system (heterogeneous storage), you should understand how you can ensure that the correct drive type is selected.

About this task

- Drives are owned by a specific node; when you create an aggregate, all drives in that aggregate must be owned by the same node, which becomes the home node for that aggregate.

  In systems using ADP, aggregates are created using partitions in which each drive is partitioned in to P1, P2 and P3 partitions.
- Aggregate names should conform to the naming scheme you determined when you planned your MetroCluster configuration.

  Disk and Aggregate Management Guide

Step 1. Display a list of available spares:
`storage disk show -spare -owner` *node_name*

Step 2. Create the aggregate by using the `storage aggregate create –mirror true` command.

If you are logged in to the cluster on the cluster management interface, you can create an aggregate on any node in the cluster. To ensure that the aggregate is created on a specific node, use the -node parameter or specify drives that are owned by that node.

You can specify the following options:

- Aggregate's home node (that is, the node that owns the aggregate in normal operation)
- List of specific drives that are to be added to the aggregate

- Number of drives to include

  Note: In the minimum supported configuration, in which a limited number of drives are available, you must use the force-small-aggregate option to allow the creation of a three disk RAID-DP aggregate.
- Checksum style to use for the aggregate
- Type of drives to use
- Size of drives to use
- Drive speed to use
- RAID type for RAID groups on the aggregate
- Maximum number of drives that can be included in a RAID group
- Whether drives with different RPM are allowed

For more information about these options, see the `storage aggregate create` man page.

Example

The following command creates a mirrored aggregate with 10 disks:

```
cluster_A::> storage aggregate create aggr1_node_A_1 -diskcount 10 -node node_A_1 -mirror true
[Job 15] Job is queued: Create aggr1_node_A_1.
[Job 15] The job is starting.
[Job 15] Job succeeded: DONE
```

Step 3. Verify the RAID group and drives of your new aggregate:
`storage aggregate show-status -aggregate` *aggregate-name*

Implementing the MetroCluster configuration

You must run the `metrocluster configure` command to start data protection in a MetroCluster configuration.

Before you begin

- There should be at least two non-root mirrored data aggregates on each cluster.

  You can verify this with the `storage aggregate show` command.

  Note: If you want to use a single mirrored data aggregate, then see Step 1 on page 91 for instructions.
- The ha-config state of the controllers and chassis must be mccip .

About this task

You issue the `metrocluster configure` command once, on any of the nodes, to enable the MetroCluster configuration. You do not need to issue the command on each of the sites or nodes, and it does not matter which node or site you choose to issue the command on.

The `metrocluster configure` command automatically pairs the two nodes with the lowest system IDs in each of the two clusters as disaster recovery (DR) partners. In a four-node MetroCluster configuration, there are two DR partner pairs. The second DR pair is created from the two nodes with higher system IDs.

Step 1. Configure the MetroCluster in the following format:

If your MetroCluster configuration has...	Then do this...
Multiple data aggregates	From any node's prompt, configure MetroCluster: `metrocluster configure` *node-name*

A single mirrored data aggregate	1. From any node's prompt, change to the advanced privilege level:
	`set -privilege advanced`
	You need to respond with `y` when you are prompted to continue into advanced mode and you see the advanced mode prompt (`*>`).
	2. Configure the MetroCluster with the -allow-with-one-aggregate true parameter:
	`metrocluster configure -allow-with-one-aggregate true` *node-name*
	3. Return to the admin privilege level:
	`set -privilege admin`

Example

Note: The best practice is to have multiple data aggregates. If the first DR group has only one aggregate and you want to add a DR group with one aggregate, you must move the metadata volume off the single data aggregate. For more information on this procedure, see Moving a metadata volume in MetroCluster configurations.

The following command enables the MetroCluster configuration on all of the nodes in the DR group that contains controller_A_1:

```
cluster_A::*> metrocluster configure -node-name controller_A_1

[Job 121] Job succeeded: Configure is successful.
```

Step 2. Verify the networking status on site A:
`network port show`

Example

The following example shows the network port usage on a four-node MetroCluster configuration:

```
cluster_A::> network port show
                                                         Speed (Mbps)
Node   Port      IPspace   Broadcast Domain Link   MTU    Admin/Oper
------ --------- --------- ---------------- ----- ------- ------------
controller_A_1
       e0a       Cluster   Cluster          up     9000  auto/1000
       e0b       Cluster   Cluster          up     9000  auto/1000
       e0c       Default   Default          up     1500  auto/1000
       e0d       Default   Default          up     1500  auto/1000
       e0e       Default   Default          up     1500  auto/1000
       e0f       Default   Default          up     1500  auto/1000
       e0g       Default   Default          up     1500  auto/1000
controller_A_2
       e0a       Cluster   Cluster          up     9000  auto/1000
       e0b       Cluster   Cluster          up     9000  auto/1000
       e0c       Default   Default          up     1500  auto/1000
       e0d       Default   Default          up     1500  auto/1000
       e0e       Default   Default          up     1500  auto/1000
       e0f       Default   Default          up     1500  auto/1000
       e0g       Default   Default          up     1500  auto/1000
```

```
14 entries were displayed.
```

Step 3. Verify the MetroCluster configuration from both sites in the MetroCluster configuration.

a. Verify the configuration from site A:
`metrocluster show`

Example

```
cluster_A::> metrocluster show

Configuration: IP fabric

Cluster                   Entry Name          State
------------------------- ------------------- -----------
 Local: cluster_A         Configuration state configured
                           Mode                normal
Remote: cluster_B         Configuration state configured
                           Mode                normal
```

b. Verify the configuration from site B:
`metrocluster show`

Example

```
cluster_B::> metrocluster show

Configuration: IP fabric

Cluster                   Entry Name          State
------------------------- ------------------- -----------
 Local: cluster_B         Configuration state configured
                           Mode                normal
Remote: cluster_A         Configuration state configured
                           Mode                normal
```

Checking the MetroCluster configuration

You can check that the components and relationships in the MetroCluster configuration are working correctly. You should do a check after initial configuration and after making any changes to the MetroCluster configuration. You should also do a check before a negotiated (planned) switchover or a switchback operation.

About this task

If the `metrocluster check run` command is issued twice within a short time on either or both clusters, a conflict can occur and the command might not collect all data. Subsequent `metrocluster check show` commands do not show the expected output.

Step 1. Check the configuration:
`metrocluster check run`

Example

The command runs as a background job and might not be completed immediately.

```
cluster_A::> metrocluster check run
The operation has been started and is running in the background. Wait for
it to complete and run "metrocluster check show" to view the results. To
check the status of the running metrocluster check operation, use the command,
"metrocluster operation history show -job-id 2245"
cluster_A::> metrocluster check show
Last Checked On: 9/13/2018 20:41:37
```

```
Component           Result
------------------- ---------
nodes               ok
lifs                ok
config-replication  ok
aggregates          ok
clusters            ok
connections         ok
6 entries were displayed.
```

Step 2. Display more detailed results from the most recent `metrocluster check run` command:

```
metrocluster check aggregate show
metrocluster check cluster show
metrocluster check config-replication show
metrocluster check lif show
metrocluster check node show
```

The `metrocluster check show` commands show the results of the most recent `metrocluster check run` command. You should always run the `metrocluster check run` command prior to using the `metrocluster check show` commands so that the information displayed is current.

Example

The following example shows the `metrocluster check aggregate show` command output for a healthy four-node MetroCluster configuration:

```
cluster_A::> metrocluster check aggregate show

Last Checked On: 8/5/2014 00:42:58

Node                  Aggregate                Check                      Result
---------------       --------------------     ---------------------      ---------
controller_A_1        controller_A_1_aggr0
                                               mirroring-status           ok
                                               disk-pool-allocation       ok
                                               ownership-state            ok
                      controller_A_1_aggr1
                                               mirroring-status           ok
                                               disk-pool-allocation       ok
                                               ownership-state            ok
                      controller_A_1_aggr2
                                               mirroring-status           ok
                                               disk-pool-allocation       ok
                                               ownership-state            ok


controller_A_2        controller_A_2_aggr0
                                               mirroring-status           ok
                                               disk-pool-allocation       ok
                                               ownership-state            ok
                      controller_A_2_aggr1
                                               mirroring-status           ok
                                               disk-pool-allocation       ok
                                               ownership-state            ok
                      controller_A_2_aggr2
                                               mirroring-status           ok
                                               disk-pool-allocation       ok
                                               ownership-state            ok

18 entries were displayed.
```

The following example shows the `metrocluster check cluster show` command output for a healthy four-node MetroCluster configuration. It indicates that the clusters are ready to perform a negotiated switchover if necessary.

```
Last Checked On: 9/13/2017 20:47:04

Cluster               Check                          Result
--------------------- ------------------------------ ---------
mccint-fas9000-0102
                      negotiated-switchover-ready    not-applicable
                      switchback-ready                not-applicable
                      job-schedules                   ok
                      licenses                        ok
                      periodic-check-enabled          ok
mccint-fas9000-0304
                      negotiated-switchover-ready    not-applicable
                      switchback-ready                not-applicable
                      job-schedules                   ok
                      licenses                        ok
                      periodic-check-enabled          ok
10 entries were displayed.
```

Completing ONTAP configuration

After configuring, enabling, and checking the MetroCluster configuration, you can proceed to complete the cluster configuration by adding additional SVMs, network interfaces and other ONTAP functionality as needed.

Verifying switchover, healing, and switchback

You should verify the switchover, healing, and switchback operations of the MetroCluster configuration.

Step 1. Use the procedures for negotiated switchover, healing, and switchback that are mentioned in the *MetroCluster Management and Disaster Recovery Guide*.

MetroCluster management and disaster recovery

Configuring the MetroCluster Tiebreaker or ONTAP Mediator software

You can download and install on a third site either the MetroCluster Tiebreaker software, or, starting with ONTAP 9.7, the ONTAP Mediator.

Before you begin

You must have a Linux host available that has network connectivity to both clusters in the MetroCluster configuration. The specific requirements are in the MetroCluster Tiebreaker or ONTAP Mediator documentation.

If you are connecting to an existing Tiebreaker or ONTAP Mediator instance, you need the username, password, and IP address of the Tiebreaker or Mediator service.

About this task

If you must install a new instance of the ONTAP Mediator, follow the directions to install and configure the software.

Chapter 5 “Configuring the ONTAP Mediator service for unplanned automatic switchover” on page 97

If you must install a new instance of the Tiebreaker software, follow the directions to install and configure the software.

Tiebreaker software installation and configuration

You cannot use both the MetroCluster Tiebreaker software and the ONTAP Mediator with the same MetroCluster configuration.

"Considerations for using ONTAP Mediator or MetroCluster Tiebreaker" on page 3

Step 1. Configure the ONTAP Mediator service or the Tiebreaker software:

– If you are using an existing instance of the ONTAP Mediator, add the ONTAP Mediator service to ONTAP using the following command:
`metrocluster configuration-settings mediator add -mediator-address` *ip-address-of-mediator-host*

– If you are using the Tiebreaker software, refer to the Tiebreaker documentation.

Tiebreaker software installation and configuration

Protecting configuration backup files

You can provide additional protection for the cluster configuration backup files by specifying a remote URL (either HTTP or FTP) where the configuration backup files will be uploaded in addition to the default locations in the local cluster.

Step 1. Set the URL of the remote destination for the configuration backup files:
`system configuration backup settings modify` *URL-of-destination*

The System Administration Guide contains additional information under the section *Managing configuration backups*.

System Administration Guide

Chapter 5. Configuring the ONTAP Mediator service for unplanned automatic switchover

Starting with ONTAP 9.7, the ONTAP Mediator service can assist the MetroCluster IP configuration in performing an automatic unplanned switchover by providing a physically separate repository for status information.

- The ONTAP Mediator service and MetroCluster Tiebreaker software should not both be used with the same MetroCluster configuration.
- The ONTAP Mediator service can support up to five MetroCluster configurations simultaneously.

Installing and configuring the ONTAP Mediator service

The ONTAP Mediator must be installed on the Linux host and then configured in ONTAP from the MetroCluster nodes.

Network requirements for using Mediator in a MetroCluster configuration

To install the ONTAP Mediator service in a MetroCluster configuration, you must ensure that the configuration meets several network requirements.

- Round trip latency

  Round trip latency must be no more than 25 ms.
- MTU

  The MTU size must be at least 1400.
- Packet loss

  Packet loss must be less than or equal to 0.01%.
- Bandwidth

  The link between the Mediator service and the MetroCluster configuration must have at least 1 Gbps of bandwidth.

Guidelines for upgrading the ONTAP Mediator in a MetroCluster configuration

If you are upgrading the ONTAP Mediator you must meet the Linux version requirements and follow guidelines for the upgrade.

- The Mediator service can be upgraded from version 1.0 to 1.1.
- All Mediator versions are supported on MetroCluster IP configurations running ONTAP 9.7 or later.

Upgrading the host operating system and Mediator together

The following table provides the upgrade guidelines if you are upgrading from RHEL/CentOS 7.6 to a later RHEL/CentOS release in addition upgrading the Mediator version.

Target Linux version	Target Mediator version	Upgrade notes
RHEL/CentOS 7.7	1.1	• The upgrade must be performed in the following order: 1. Upgrade the operating system from RHEL/CentOS version 7.6 to 7.7. **Note:** The ONTAP Mediator and Mediator-assisted automatic unplanned switchover is not available during the operating system upgrade. The Mediator is offline until the Mediator version 1.0 to 1.1 upgrade process is complete. 2. Reboot the host to apply the kernel module changes. 3. Upgrade the Mediator from version 1.0 to 1.1. "Installing or upgrading the ONTAP Mediator service" on page 100 • The `storage iscsi-initiator show` command will report that the connection to the Mediator service is down during the upgrade. • The ONTAP operating system will generate `cf.mccip.med.auso.stDisabled` EMS events during the upgrade. • The ONTAP operating system will generate a `cf.mccip.med.auso.stEnabled` EMS event when automatic unplanned switchover is re-enabled.
RHEL/CentOS 8.0 or 8.1	1.1	There is no direct upgrade path. You must remove the 1.0 version and install the 1.1 version after the operating system upgrade: 1. Delete the Mediator service from the ONTAP configuration: `metrocluster configuration-settings mediator remove` 2. Uninstall the 1.0 version of the Mediator service. "Uninstalling the ONTAP Mediator service " on page 111 3. Upgrade the Linux operating system to version 8.0 or 8.1. 4. Install the 1.1 version of the Mediator service. "Installing or upgrading the ONTAP Mediator service" on page 100 5. Add the Mediator service to the ONTAP configuration: `metrocluster configuration-settings add -address` *mediator-1.1-ip-address*

After the upgrade

After the Mediator and operating system upgrade is complete, you should issue the `storage iscsi-initiator show` command to confirm that the Mediator connections are up.

Reverting from a Mediator 1.1 installation

A direct revert from Mediator version 1.1 to 1.0 is not supported. You must remove the 1.1 version and reinstall the 1.0 version.

1. Delete the Mediator service from the ONTAP configuration: `metrocluster configuration-settings mediator remove`
2. Uninstall the 1.1 version of the Mediator service.

   "Uninstalling the ONTAP Mediator service " on page 111
3. Install the 1.0 version of the Mediator service.

"Installing or upgrading the ONTAP Mediator service" on page 100

4. Add the Mediator service to the ONTAP configuration: `metrocluster configuration-settings add -address` *mediator-1.0-ip-address*

Recovering from Linux kernel upgrades

The ONTAP Mediator requires the SCST kernel module. If the Linux kernel is updated, this dependency may lead to a loss of service. It is highly recommended that you rebuild the SCST kernel module when any kernel package changes are made.

Notes:

- Upgrading from ONTAP Mediator version 1.0 to 1.1 rebuilds the SCST module.
- Kernel module changes are applied after the Linux kernel is rebooted.

You can use either of the following procedures to recover from a kernel upgrade that has resulted in loss of service for the Mediator.

Procedure	Steps

Remove and reinstall the SCST kernel module	**You must have the SCST tar bundle used by your version of Mediator:**
	• **ONTAP Mediator 1.0 requires `scst-3.3.0.tar.bz2`**
	• **ONTAP Mediator 1.1 requires `scst-3.4.0.tar.bz2`**
	1. Uninstall the SCST module:
	a. Download and untar the SCST tar bundle required by your version of Mediator.
	b. Run the following commands inside of the `scst` directory:
	`systemctl stop mediator-scst`
	`make scstadm_uninstall`
	`make iscsi_uninstall`
	`make usr_uninstall`
	`make scst_uninstall`
	`depmod`
	2. Reinstall the SCST module for your version of Mediator by issuing the following commands inside of the `scst` directory:
	`make scst_install`
	`make usr_install`
	`make iscsi_install`
	`make scstadm_install`
	`depmod`
	`patch /etc/init.d/scst < /opt/netapp/lib/ontap_mediator/systemd/scst.patch`
	`reboot`

Remove and reinstall ONTAP Mediator Note: This requires a reconfiguration of the Mediator in ONTAP.	**1. Delete the Mediator service from the ONTAP configuration:** `metrocluster configuration-settings mediator remove` **2. Uninstall the Mediator service.** **"Uninstalling the ONTAP Mediator service " on page 111** **3. Reinstall the Mediator service.** **"Installing or upgrading the ONTAP Mediator service" on page 100** **4. Add the Mediator service to the ONTAP configuration:** `metrocluster configuration-settings add -address` *mediator-ip-address*

Installing or upgrading the ONTAP Mediator service

To install the ONTAP Mediator service, you must ensure all prerequisites are met, get the installation package and run the installer on the host.

Before you begin

Mediator version	Supported Linux versions
1.1	Red Hat Enterprise Linux or CentOS 7.6, 7.7, 8.0, 8.1

- 64-bit physical installation or virtual machine
  - 8 GB RAM
  - User: Root access

  The best practices for installing Red Hat Enterprise Linux or CentOS and the associated repositories on your system are listed below. Systems installed or configured differently might require additional steps.
  - You must install Red Hat Enterprise Linux or CentOS according to Red Hat best practices.
  - While installing the ONTAP Mediator service on Red Hat Enterprise Linux or CentOS, the system must have access to the appropriate repository so that the installation program can access and install all the required software dependencies.
  - For the `yum` installer to find dependent software in the Red Hat Enterprise Linux repositories, you must have registered the system during the Red Hat Enterprise Linux installation or afterwards by using a valid Red Hat subscription.
- The following ports must be unused and available for the Mediator:
  - 3260
  - 31784
- If using a third-party firewall:
  - HTTPS access must be enabled.
  - It must be configured to allow access on ports 31784 and 3260.

  When using the default Red Hat or CentOS firewall, the firewall is automatically configured during Mediator installation.
- If the Linux host is in a location without access to the internet, you can either install the packages manually or you must ensure that the required packages are available in a local repository.

  You can use the following link for information about setting up a repository.

  Redhat Solution 3176811: Creating a Local Repository and Sharing With Disconnected/Offline/Air-gapped Systems

  The following packages are required by the ONTAP Mediator service version 1.1:

All RHEL/CentOS versions	Additional packages for RHEL/ CentOS 7.x	Additional packages for RHEL/ CentOS 8.x
– openssl – openssl-devel – kernel-devel – gcc – libselinux-utils – make – redhat-lsb-core – patch – bzip2 – python36 – python36-devel – perl-Data-Dumper – perl-ExtUtils-MakeMaker – python3-pip	– policycoreutils-python – python36-pip	– elfutils-libelf-devel – policycoreutils-python-utils

- If signature verification is configured, it must be disabled. This can be done in one of two ways:
  - If the UEFI SecureBoot mechanism is configured, disable it.
  - Disable the signature verification mechanism by updating and regenerating the `grub.cfg` file:
    1. Open the `/etc/default/grub` file.
    2. Add the string `module.sig_enforce=0` to the end of the `GRUB_CMDLINE_LINUX` statement.
    3. Regenerate the `grub.cfg` file to implement the change:
       `update-bootloader || update-grub || grub2-mkconfig -o /boot/grub2/grub.cfg`
    4. Reboot the host.

About this task

The Mediator installation package is a self-extracting compressed tar file that includes:

- An RPM file containing all dependencies that cannot be obtained from the supported release's repository.
- An install script.

A valid SSL certification is recommended, as documented in this procedure.

This procedure is used for an installation or an upgrade of an existing installation.

"Guidelines for upgrading the ONTAP Mediator in a MetroCluster configuration" on page 97

Step 1. Enable access to the repositories listed above so Mediator can access the required packages during the installation process.

If your operating system is...	You must provide access to these repositories...
RHEL 7.x	rhel-7-server-optional-rpms
CentOS 7.x	C7.6.1810 - Base repository

If your operating system is...	You must provide access to these repositories...
RHEL 8.x	• rhel-8-for-x86_64-baseos-rpms • rhel-8-for-x86_64-appstream-rpms
CentOS 8.0	kernel-devel

If your operating system is...	Use these steps...

RHEL 7.x	1. Subscribe to the required repository:
	`subscription-manager repos --enable rhel-7-server-optional-rpms`
	The following example shows the execution of this command:
	`[root@localhost ~]# subscription-manager repos --enable rhel-7-server-optional-rpms`
	`Repository 'rhel-7-server-optional-rpms' is enabled for this system.`
	2. Run the `yum repolist` command.
	The following example shows the execution of this command. The rhel-7-server-optional-rpms repository should appear in the list.
	`[root@localhost ~]# yum repolist`
	`Loaded plugins: product-id, search-disabled-repos, subscription-manager`
	`rhel-7-server-optional-rpms | 3.2 kB 00:00:00`
	`rhel-7-server-rpms | 3.5 kB 00:00:00`
	`(1/3): rhel-7-server-optional-rpms/7Server/x86_64/group | 26 kB 00:00:00`
	`(2/3): rhel-7-server-optional-rpms/7Server/x86_64/updateinfo | 2.5 MB 00:00:00`
	`(3/3): rhel-7-server-optional-rpms/7Server/x86_64/primary_db | 8.3 MB 00:00:01`
	`repo id repo name status`
	`rhel-7-server-optional-rpms/7Server/x86_64 Red Hat Enterprise Linux 7 Server - Optional (RPMs) 19,447`
	`rhel-7-server-rpms/7Server/x86_64 Red Hat Enterprise Linux 7 Server (RPMs) 26,758`
	`repolist: 46,205`
	`[root@localhost ~]#`

RHEL 8.x	1. Subscribe to the required repository:
	`subscription-manager repos --enable rhel-8-for-x86_64-baseos-rpms`
	`subscription-manager repos --enable rhel-8-for-x86_64-appstream-rpms`
	The following example shows the execution of this command:
	`[root@localhost ~]# subscription-manager repos --enable rhel-8-for-x86_64-baseos-rpms`
	`[root@localhost ~]# subscription-manager repos --enable rhel-8-for-x86_64-appstream-rpms`
	`Repository 'rhel-8-for-x86_64-baseos-rpms' is enabled for this system.`
	`Repository 'rhel-8-for-x86_64-appstream-rpms' is enabled for this system.`
	2. Run the `yum repolist` command.
	The newly subscribed repositories should appear in the list.

<table>
<tr><td>CentOS 7.x</td><td>Add the C7.6.1810 - Base repository. The C7.6.1810 - Base vault repository contains the kernel-devel package needed for ONTAP Mediator.
1. Add the following lines to <code>/etc/yum.repos.d/CentOS-Vault.repo</code>.
<pre>[C7.6.1810-base]
name=CentOS-7.6.1810 - Base
baseurl=http://vault.centos.org/7.6.1810/os/$
basearch/gpgcheck=1
gpgkey=file:///etc/pki/rpm-gpg/RPM-GPG-KEY-CentOS-7
enabled=1</pre>2. Run the <code>yum repolist</code> command.
The following example shows the execution of this command. The CentOS-7.6.1810 - Base repository should appear in the list.
<pre>Loaded plugins: fastestmirror
Loading mirror speeds from cached hostfile
 * base: distro.ibiblio.org
 * extras: distro.ibiblio.org
 * updates: ewr.edge.kernel.org
C7.6.1810-base | 3.6 kB 00:00:00
(1/2): C7.6.1810-base/x86_64/group_gz | 166 kB 00:00:00
(2/2): C7.6.1810-base/x86_64/primary_db | 6.0 MB 00:00:04
repo id   repo name   status
C7.6.1810-base/x86_64   CentOS-7.6.1810 - Base   10,019
base/7/x86_64   CentOS-7 - Base   10,097
extras/7/x86_64   CentOS-7 - Ex   307
updates/7/x86_64   CentOS-7 - Updates   1,010
repolist: 21,433
[root@localhost ~]#</pre></td></tr>
<tr><td>CentOS 8.0.1905 or later builds</td><td>Because the latest version of the 8.0 (CentOS 8.0.1905) core resides in the CentOS Vault, you must provide access to the matching kernel-devel package to compile the needed kernel module.
Issue the following command to directly install the kernel-devel package: <code>rpm -I http://vault.centos.org/8.0.1905/BaseOS/x86_64/os/Packages/kernel-devel-$(uname -r).rpm</code>

If the system displays an error indicating that the package is already installed, remove the package and try again:
1. Remove the kernel-devel package: <code>yum remove kernel-devel</code>
2. Repeat the <code>rpm</code> command shown above.</td></tr>
</table>

Step 2. Download the Mediator installation package from the ONTAP Mediator page.

Lenovo Downloads: ONTAP Mediator

Step 3. Confirm that the Mediator installation package is in the target directory: `ls`

Example

```
[root@mediator-host ~]#ls
./ontap-mediator_1.1
```

If you are at a location without access to the internet, you must ensure that the installer has access to the required packages.

Step 4. If necessary, move the Mediator installation package from the download directory to the installation directory on the Linux Mediator host.

Step 5. Install the Mediator installation package and respond to the prompts as required:
`./ontap-mediator_1.1`

The installation process proceeds to create the required accounts and install required packages. If you have a previous version of Mediator installed on the host, you will be prompted to confirm that you want to upgrade.

Example

The following example shows a fresh installation of the Mediator service:

```
[root@red-hat-enterprise-linux ~]# ./ontap-mediator_1.1

ONTAP Mediator: Self Extracting Installer


ONTAP Mediator requires two user accounts. One for the service (netapp), and one for use by ONTAP to
 the mediator API (mediatoradmin).

Would you like to use the default account names: netapp + mediatoradmin? (Y(es)/n(o)): y



Enter ONTAP Mediator system service account (mediatoradmin) password:

Re-Enter ONTAP Mediator system service account (mediatoradmin) password:


Checking for default Linux firewall

Linux firewall is running. Open ports 31784 and 3260? y(es)/n(o): y
success
success

################################################################

Preparing for installation of ONTAP Mediator packages.

Do you wish to continue? y(es)/n(o): y

+ Installing required packages.

Loaded plugins: product-id, search-disabled-repos, subscription-manager
epel/x86_64/metalink                 |  17 kB  00:00:00
epel-extra                           | 4.9 kB  00:00:00
ius                                  | 1.3 kB  00:00:00
rhel-7-server-rpms                   | 3.5 kB  00:00:00
(1/3): ius/x86_64/primary            | 129 kB  00:00:00
(2/3): epel-extra/group_gz           |  88 kB  00:00:01
(3/3): epel-extra/primary_db         | 6.7 MB  00:00:06
ius                                                                                                 538/538
Package 1:make-3.82-23.el7.x86_64 already installed and latest version
                                .
                                .
                                .
                                .

================================================================================================================
 Package                    Arch                    Version                     Repository                  Size
Installing:
 bzip2                       x86_64                  1.0.6-13.el7              rhel-7-server-rpms         52 k
 gcc                         x86_64                  4.8.5-36.el7_6.2          rhel-7-server-rpms         16 M
 kernel-devel               x86_64                  3.10.0-957.el7            rhel-7-server-rpms         17 M
 openssl-devel              x86_64                  1:1.0.2k-16.el7_6.1       rhel-7-server-rpms        1.5 M
 patch                       x86_64                  2.7.1-10.el7_5            rhel-7-server-rpms         110 k
 perl-Compress-Raw-Zlib     x86_64                 1:2.061-4.el7             rhel-7-server-rpms         57 k
 perl-Digest-MD5            x86_64                  2.52-3.el7                rhel-7-server-rpms         30 k
 perl-Digest-SHA            x86_64                  1:5.85-4.el7              rhel-7-server-rpms         58 k
 perl-ExtUtils-CBuilder     noarch                 1:0.28.2.6-294.el7_6      rhel-7-server-rpms        68 k
 perl-ExtUtils-MakeMaker    noarch                  6.68-3.el7                rhel-7-server-rpms         275 k
```

```
 perl-JSON-PP               noarch               2.27202-2.el7             rhel-7-server-rpms        55 k
 python36                    x86_64               3.6.8-2.el7.ius           ius                        60 k
 python36-devel              x86_64               3.6.8-2.el7.ius           ius                       206 k
 python36-pip                noarch               9.0.1-2.el7.ius           ius                       1.7 M
 redhat-lsb-core            x86_64               4.1-27.el7                rhel-7-server-rpms        37 k

Updating:
 openssl                     x86_64               1:1.0.2k-16.el7_6.1       rhel-7-server-rpms       493 k

Installing for dependencies:
 at                           x86_64                3.1.13-24.el7             rhel-7-server-rpms        51 k
 avahi-libs                  x86_64               0.6.31-19.el7             rhel-7-server-rpms        62 k
 bc                           x86_64                1.06.95-13.el7            rhel-7-server-rpms       115 k
 cpp                          x86_64                4.8.5-36.el7_6.2          rhel-7-server-rpms       6.0 M
 cups-client                 x86_64               1:1.6.3-35.el7            rhel-7-server-rpms       151 k
 cups-libs                   x86_64               1:1.6.3-35.el7            rhel-7-server-rpms       357 k
 dwz                          x86_64                0.11-3.el7                rhel-7-server-rpms        99 k
 ed                           x86_64                1.9-4.el7                 rhel-7-server-rpms        72 k
 gdbm-devel                   x86_64                1.10-8.el7                rhel-7-server-rpms        47 k
 glibc-devel                 x86_64               2.17-260.el7_6.6          rhel-7-server-rpms        1.1 M
 glibc-headers               x86_64               2.17-260.el7_6.6          rhel-7-server-rpms       684 k
 kernel-headers              x86_64               3.10.0-957.27.2.el7       rhel-7-server-rpms       8.0 M
 keyutils-libs-devel         x86_64               1.5.8-3.el7               rhel-7-server-rpms       37 k
 krb5-devel                   x86_64                1.15.1-37.el7_6           rhel-7-server-rpms       271 k
 libcom_err-devel            x86_64               1.42.9-13.el7             rhel-7-server-rpms       31 k
 libdb-devel                 x86_64               5.3.21-24.el7             rhel-7-server-rpms       38 k
 libkadm5                     x86_64                1.15.1-37.el7_6           rhel-7-server-rpms       178 k
 libmpc                       x86_64                1.0.1-3.el7               rhel-7-server-rpms       51 k
 libselinux-devel            x86_64               2.5-14.1.el7              rhel-7-server-rpms       187 k
 libsepol-devel              x86_64               2.5-10.el7                rhel-7-server-rpms        77 k
 libverto-devel              x86_64               0.2.5-4.el7               rhel-7-server-rpms        12 k
 m4                             x86_64                1.4.16-10.el7             rhel-7-server-rpms       256 k
 mailx                        x86_64                12.5-19.el7               rhel-7-server-rpms       245 k
 mpfr                         x86_64                3.1.1-4.el7               rhel-7-server-rpms       203 k
 pcre-devel                  x86_64               8.32-17.el7               rhel-7-server-rpms       480 k
 perl                        x86_64               4:5.16.3-294.el7_6        rhel-7-server-rpms       8.0 M
 perl-Carp                   noarch               1.26-244.el7              rhel-7-server-rpms       19 k
 perl-Data-Dumper            x86_64               2.145-3.el7               rhel-7-server-rpms       47 k
 perl-Digest                 noarch               1.17-245.el7              rhel-7-server-rpms      23 k
 perl-Encode                 x86_64               2.51-7.el7                rhel-7-server-rpms       1.5 M
 perl-Exporter               noarch               5.68-3.el7                rhel-7-server-rpms      28 k
 perl-ExtUtils-Install       noarch               1.58-294.el7_6            rhel-7-server-rpms      75 k
 perl-ExtUtils-Manifest      noarch               1.61-244.el7              rhel-7-server-rpms       31 k
 perl-ExtUtils-ParseXS       noarch               1:3.18-3.el7              rhel-7-server-rpms        77 k
 perl-File-Path              noarch               2.09-2.el7                rhel-7-server-rpms       27 k
 perl-File-Temp              noarch               0.23.01-3.el7             rhel-7-server-rpms       56 k
 perl-Filter                 x86_64               1.49-3.el7                rhel-7-server-rpms         76 k
 perl-Getopt-Long              noarch               2.40-3.el7                rhel-7-server-rpms         56 k
 perl-HTTP-Tiny                noarch               0.033-3.el7               rhel-7-server-rpms         38 k
 perl-IPC-Cmd                  noarch               1:0.80-4.el7              rhel-7-server-rpms         34 k
 perl-Locale-Maketext          noarch               1.23-3.el7                rhel-7-server-rpms         93 k
 perl-Locale-Maketext-Simple   noarch               1:0.21-294.el7_6          rhel-7-server-rpms         50 k
 perl-Module-CoreList          noarch               1:2.76.02-294.el7_6       rhel-7-server-rpms         86 k
 perl-Module-Load              noarch               1:0.24-3.el7              rhel-7-server-rpms         11 k
 perl-Module-Load-Conditional  noarch               0.54-3.el7                rhel-7-server-rpms         18 k
 perl-Module-Metadata          noarch               1.000018-2.el7            rhel-7-server-rpms         26 k
 perl-Params-Check             noarch               1:0.38-2.el7              rhel-7-server-rpms         18 k
 perl-PathTools              x86_64               3.40-5.el7                rhel-7-server-rpms         83 k
 perl-Perl-OSType            noarch               1.003-3.el7               rhel-7-server-rpms         20 k
 perl-Pod-Escapes              noarch               1:1.04-294.el7_6          rhel-7-server-rpms         51 k
 perl-Pod-Perldoc            noarch               3.20-4.el7                rhel-7-server-rpms         87 k
 perl-Pod-Simple             noarch               1:3.28-4.el7              rhel-7-server-rpms         216 k
 perl-Pod-Usage              noarch               1.63-3.el7                rhel-7-server-rpms         27 k
 perl-Scalar-List-Utils      x86_64               1.27-248.el7              rhel-7-server-rpms        36 k
 perl-Socket                   x86_64                2.010-4.el7               rhel-7-server-rpms         49 k
 perl-Storable               x86_64               2.45-3.el7                rhel-7-server-rpms         77 k
 perl-Test-Harness           noarch               3.28-3.el7                rhel-7-server-rpms         302 k
 perl-Text-ParseWords        noarch               3.29-4.el7                rhel-7-server-rpms         14 k
```

```
 perl-Time-HiRes               x86_64              4:1.9725-3.el7            rhel-7-server-rpms        45 k
 perl-Time-Local               noarch              1.2300-2.el7              rhel-7-server-rpms        24 k
 perl-constant                 noarch              1.27-2.el7                rhel-7-server-rpms        19 k
perl-devel                     x86_64              4:5.16.3-294.el7_6        rhel-7-server-rpms       453 k
 perl-libs                     x86_64              4:5.16.3-294.el7_6        rhel-7-server-rpms       688 k
 perl-macros                   x86_64              4:5.16.3-294.el7_6        rhel-7-server-rpms        44 k
 perl-parent                   noarch              1:0.225-244.el7           rhel-7-server-rpms        12 k
 perl-podlators                noarch              2.5.1-3.el7               rhel-7-server-rpms       112 k
 perl-srpm-macros              noarch              1-8.el7                   rhel-7-server-rpms       4.7 k
 perl-threads                  x86_64              1.87-4.el7                rhel-7-server-rpms        49 k
 perl-threads-shared           x86_64              1.43-6.el7                rhel-7-server-rpms        39 k
 perl-version                  x86_64              3:0.99.07-3.el7           rhel-7-server-rpms        84 k
 psmisc                        x86_64              22.20-15.el7              rhel-7-server-rpms       141 k
 pyparsing                     noarch              1.5.6-9.el7               rhel-7-server-rpms        94 k
 python-rpm-macros             noarch              3-24.el7                  epel-extra                7.9 k
 python-srpm-macros            noarch              3-24.el7                  epel-extra                7.3 k
 python3-rpm-macros            noarch              3-24.el7                  epel-extra                6.9 k
 python36-libs                 x86_64              3.6.8-2.el7.ius           ius                       8.6 M
 python36-setuptools           noarch              39.2.0-4.el7.ius          ius                       621 k
 redhat-lsb-submod-security    x86_64              4.1-27.el7                rhel-7-server-rpms        15 k
 redhat-rpm-config             noarch              9.1.0-87.el7              rhel-7-server-rpms        81 k
 spax                          x86_64              1.5.2-13.el7              rhel-7-server-rpms       260 k
 systemtap-sdt-devel           x86_64              3.3-3.el7                 rhel-7-server-rpms        74 k
 time                          x86_64              1.7-45.el7                rhel-7-server-rpms        30 k
 zip                           x86_64              3.0-11.el7                rhel-7-server-rpms       260 k
 zlib-devel                    x86_64              1.2.7-18.el7              rhel-7-server-rpms        50 k

Updating for dependencies:
 glibc                         x86_64              2.17-260.el7_6.6          rhel-7-server-rpms       3.6 M
 glibc-common                  x86_64              2.17-260.el7_6.6          rhel-7-server-rpms        11 M
 krb5-libs                     x86_64              1.15.1-37.el7_6           rhel-7-server-rpms       803 k
 libgcc                        x86_64              4.8.5-36.el7_6.2          rhel-7-server-rpms       102 k
 libgomp                       x86_64              4.8.5-36.el7_6.2          rhel-7-server-rpms       158 k
 openssl-libs                  x86_64              1:1.0.2k-16.el7_6.1       rhel-7-server-rpms       1.2 M

Transaction Summary
================================================================================================================
Install  15 Packages (+84 Dependent packages)
Upgrade   1 Package  (+ 6 Dependent packages)

Total download size: 97 M

Is this ok [y/d/N]: y

Downloading packages:

Delta RPMs disabled because /usr/bin/applydeltarpm not installed.
                    .
                    .
                    .
                    .
+ Installing ONTAP Mediator. (Log: /tmp/ontap_mediator.jx6UrF/ontap-mediator/install.log)
+ Install successful. (Moving log to /opt/netapp/lib/ontap_mediator/log/install.log)

[root@red-hat-enterprise-linux ~]#
```

Example

The following example shows an upgrade of the Mediator service:

```
[root@scspr1845484003 ontap-mediator]# ./ontap_mediator_1.1
ONTAP Mediator will be upgraded from version 1.0.231618 to 1.1.5590287.


Do you wish to continue? Y(es)/n(o): y
Capturing support_bundle
    Mediator API User Name: mediatoradmin
```

```
                Password:
Running plugins, please wait...

Running 11/11: sysctl...
Creating tar archive...
Support bundle has been generated at /opt/netapp/data/support_bundles/scspr1845484003_1582830730_BASIC.tgz
Testing the DB can be upgraded
Verifying database is up-to-date...
Database out-of-date.Migrations are necessary:
Current revision: bdd9252c33ee, Head revision: 6b2a053cd598
Cloning database...
Cloning successful.
Performing database migrations against clone...
/opt/netapp/lib/ontap_mediator/pyenv/lib64/python3.6/site-packages/alembic/util/messaging.py:70: UserWarning:
 Skipping unsupported ALTER for creation of implicit constraint
  warnings.warn(msg)
Migrations successful.
Database upgrade operation complete.
+ Upgrading ONTAP Mediator. (Log: /root/ontap-mediator/upgrade_20200227141143.log)
+ Upgrade successful. (Moving log to /opt/netapp/lib/ontap_mediator/log/upgrade_20200227141143.log)
```

Step 6. Use the following steps to configure third-part certification. Third-party certification is recommended.

a. The certificate must be placed in the following directory: `/opt/netapp/lib/ontap_mediator/ontap_mediator/server_config` You must overwrite the following files with your certificate, renaming the certificate files if necessary to `ontap_mediator_server.crt` and `ontap_mediator_server.key`.

```
# SSL Certificates
cert_path: '/opt/netapp/lib/ontap_mediator/ontap_mediator/server_config/ontap_mediator_server.crt'
key_path: '/opt/netapp/lib/ontap_mediator/ontap_mediator/server_config/ontap_mediator_server.key'
```

Note: You must be sure to follow security best practices for your operating system.If you are not using a third-party certificate, the Mediator installation process will create a self-signed certificate.

b. Along with the server certificates, you must update the CA certificates as well. You must overwrite the following files with your certificate, renaming the certificate files if necessary to :

- `ontap_mediator_server.crt`
- `ontap_mediator_server.key`
- `ca.crt`
- `ca.key`
- `ca.srl`

```
# SSL Certificates
cert_path: '/opt/netapp/lib/ontap_mediator/ontap_mediator/server_config/ontap_mediator_server.crt'
key_path: '/opt/netapp/lib/ontap_mediator/ontap_mediator/server_config/ontap_mediator_server.key'
ca_cert_path: '/opt/netapp/lib/ontap_mediator/ontap_mediator/server_config/ca.crt'
ca_key_path: '/opt/netapp/lib/ontap_mediator/ontap_mediator/server_config/ca.key'
ca_serial_path: '/opt/netapp/lib/ontap_mediator/ontap_mediator/server_config/ca.srl'
```

Step 7. Verify the installation.

a. Run the following command to view the status of the ONTAP Mediator services: `systemctl`

```
[root@scspr1915530002 ~]# systemctl status ontap_mediator mediator-scst

● ontap_mediator.service - ONTAP Mediator
   Loaded: loaded (/opt/netapp/lib/ontap_mediator/systemd/ontap_mediator.service; enabled; vendor preset: disabled)

   Active: active (running) since Thu 2020-06-18 09:55:02 EDT; 3 days ago

Main PID: 3559 (uwsgi)

   Status: "uWSGI is ready"
```

```
   CGroup: /system.slice/ontap_mediator.service

           \u251c\u25003559 /opt/netapp/lib/ontap_mediator/pyenv/bin/uwsgi --ini /opt/netapp/lib/ontap_mediator
/uwsgi/ontap_mediator.ini

           \u251c\u25004510 /opt/netapp/lib/ontap_mediator/pyenv/bin/uwsgi --ini /opt/netapp/lib/ontap_mediator
/uwsgi/ontap_mediator.ini

           \u2514\u25004512 /opt/netapp/lib/ontap_mediator/pyenv/bin/uwsgi --ini /opt/netapp/lib/ontap_mediator
/uwsgi/ontap_mediator.ini



Jun 18 09:54:43 scspr1915530002 systemd[1]: Starting ONTAP Mediator...

Jun 18 09:54:45 scspr1915530002 ontap_mediator[3559]: [uWSGI] getting INI configuration from /opt/netapp/lib
/ontap_mediator/uwsgi/ontap_mediator.ini

Jun 18 09:55:02 scspr1915530002 systemd[1]: Started ONTAP Mediator.



● mediator-scst.service
   Loaded: loaded (/opt/netapp/lib/ontap_mediator/systemd/mediator-scst.service; enabled; vendor preset: disabled)

   Active: active (running) since Thu 2020-06-18 09:54:51 EDT; 3 days ago

  Process: 3564 ExecStart=/etc/init.d/scst start (code=exited, status=0/SUCCESS)

Main PID: 4202 (iscsi-scstd)

   CGroup: /system.slice/mediator-scst.service

           \u2514\u25004202 /usr/local/sbin/iscsi-scstd



Jun 18 09:54:43 scspr1915530002 systemd[1]: Starting mediator-scst.service...

Jun 18 09:54:48 scspr1915530002 iscsi-scstd[4200]: max_data_seg_len 1048576, max_queued_cmds 2048

Jun 18 09:54:51 scspr1915530002 scst[3564]: Loading and configuring SCST[  OK  ]

Jun 18 09:54:51 scspr1915530002 systemd[1]: Started mediator-scst.service.

[root@scspr1915530002 ~]#
```

b. To view the ports the ONTAP Mediator service is using, run: `netstat`

```
[root@scspr1905507001 ~]# netstat -anlt | grep -E '3260|31784'

tcp        0      0 0.0.0.0:31784           0.0.0.0:*               LISTEN

tcp        0      0 0.0.0.0:3260            0.0.0.0:*               LISTEN

tcp6       0      0 :::3260                 :::*                    LISTEN
```

The ONTAP Mediator service is now installed and running. Further configuration must be performed in the ONTAP storage system to use the Mediator features.

Configuring the ONTAP Mediator service from a MetroCluster IP configuration

The ONTAP Mediator service must be configured on the ONTAP node for use in a MetroCluster IP configuration.

Before you begin

- The ONTAP Mediator must have been successfully installed on a network location that can be reached by both MetroCluster sites.
- You must have the IP address of the host running the ONTAP Mediator service.
- You must have the username and password for the ONTAP Mediator service.
- All nodes of the MetroCluster IP configuration must be online.

About this task

- This task enables automatic unplanned switchover by default.
- This task can be performed on the ONTAP interface of any node in the MetroCluster IP configuration.
- A single installation of the ONTAP Mediator service can be configured with up to five MetroCluster IP configurations.

Step 1. Add the ONTAP Mediator service to ONTAP using the following command:
`metrocluster configuration-settings mediator add -mediator-address` *ip-address-of-mediator-host*

Note: You will be prompted for the username and password for the Mediator admin user account.

Step 2. Verify that the automatic switchover feature in enabled: `metrocluster show`

Step 3. Verify that the Mediator is now running.

a. Run:
`storage disk show -container-type mediator`

```
ClusterA::> storage disk show -container-type mediator
                     Usable           Disk    Container   Container
Disk                   Size Shelf Bay Type    Type        Name      Owner
---------------- ---------- ----- --- ------- ----------- --------- --------
NET-1.5                   -     -   - VMDISK  mediator    -         ClusterA-02
NET-1.6                   -     -   - VMDISK  mediator    -         ClusterB-01
NET-1.7                   -     -   - VMDISK  mediator    -         ClusterB-02
NET-1.8                   -     -   - VMDISK  mediator    -         ClusterA-01
```

b. Run the following command:
`set advanced`

```
ClusterA::> set advanced
```

c. Run:
`storage iscsi-initiator show -label mediator`

```
ClusterA::*> storage iscsi-initiator show -label mediator
  (storage iscsi-initiator show)
                                                                        Status
Node Type Label    Target Portal      Target Name                      Admin/Op
---- ---- -------- ------------------ -------------------------------- --------
ClusterA-01
     mailbox
          mediator 1.1.1.1       iqn.2012-05.local:mailbox.target.6616cd3f-9ef1-11e9-aada-00a098ccf5d8:a05e1ffb
-9ef1-11e9-8f68-00a098cbca9e:1 up/up
ClusterA-02
     mailbox
          mediator 1.1.1.1       iqn.2012-05.local:mailbox.target.6616cd3f-9ef1-11e9-aada-00a098ccf5d8:a05e1ffb
-9ef1-11e9-8f68-00a098cbca9e:1 up/up
```

Connecting a MetroCluster configuration to a different ONTAP Mediator instance

If you want to connect the MetroCluster nodes to a different ONTAP Mediator instance, you must unconfigure and then reconfigure the Mediator connection in the ONTAP software.

Before you begin

You need the username, password, and IP address of the new ONTAP Mediator instance.

About this task

These commands can be issued from any node in the MetroCluster configuration.

Step 1. Remove the current ONTAP Mediator from the MetroCluster configuration:
`metrocluster configuration-settings mediator remove`

Step 2. Establish the new ONTAP Mediator connection to the MetroCluster configuration:
`metrocluster configuration-settings mediator add -mediator-address` *ip-address-of-mediator-host*

Changing the ONTAP Mediator password

After you have installed ONTAP Mediator service, you might want to change the password. You can change the password in two ways.

About this task

This task is performed on the Linux host on which the ONTAP Mediator service is installed.

If you are unable to reach this command, you might need to run the command using the full path as shown in the following example:
`/usr/local/bin/mediator_change_password`

Step 1. Change the password by choosing one of the following options:

– Run the `mediator_change_password` command and respond to the prompts as shown in the following example:

```
[root@mediator-host ~]# mediator_change_password
Change the Mediator API password by entering the following values:
    Mediator API User Name: mediatoradmin
               Old Password:
               New Password:
          Confirm Password:
The password has been updated successfully.
[root@mediator-host ~]#
```

– Run the following command: `MEDIATOR_USERNAME=` *mediatoradmin* `MEDIATOR_PASSWORD=` *mediator1* `MEDIATOR_NEW_PASSWORD=` *mediator2* `mediator_change_password`

The example shows the password is changed from 'mediator1' to 'mediator2'.

```
[root@mediator-host ~]# MEDIATOR_USERNAME=mediatoradmin MEDIATOR_PASSWORD=mediator1 MEDIATOR_NEW_PASSWORD
=mediator2 mediator_change_password
The password has been updated successfully.
[root@mediator-host ~]#
```

Changing the ONTAP Mediator user name

After the ONTAP Mediator service is installed, you might want to change the user name.

About this task

This task is performed on the Linux host on which the ONTAP Mediator service is installed.

If you are unable to reach this command, you might need to run the command using the full path as shown in the following example:

```
/usr/local/bin/mediator_username
```

Step 1. Change the ONTAP Mediator user name by choosing from one of the following options:

– Run the command `mediator_change_user` and respond to the prompts as shown in the following example:

```
[root@mediator-host ~]# mediator_change_user
Modify the Mediator API username by entering the following values:
    Mediator API User Name: mediatoradmin
                  Password:
New Mediator API User Name: mediator
The account username has been modified successfully.
[root@mediator-host ~]#
```

– Run the following command: `MEDIATOR_USERNAME=` *mediator* `MEDIATOR_PASSWORD=` *mediator2* `MEDIATOR_NEW_USERNAME=` *mediatoradmin* `mediator_change_user` as shown in the following example:

```
[root@mediator-host ~]# MEDIATOR_USERNAME= mediator MEDIATOR_PASSWORD='mediator2' MEDIATOR_NEW_USERNAME=
 mediatoradmin mediator_change_user
The account username has been modified successfully.
[root@mediator-host ~]#
```

Uninstalling the ONTAP Mediator service

If necessary, you can remove the ONTAP Mediator service.

Before you begin

The Mediator must be disconnected from ONTAP before you remove the Mediator service.

About this task

This task is performed on the Linux host on which the ONTAP Mediator service is installed.

If you are unable to reach this command, you might need to run the command using the full path as shown in the following example:

```
/usr/local/bin/uninstall_ontap_mediator
```

Step 1. Uninstall the ONTAP Mediator service as shown in the following example using the command: `uninstall_ontap_mediator`

```
[root@mediator-host ~]# uninstall_ontap_mediator

ONTAP Mediator: Self Extracting Uninstaller

+ Removing ONTAP Mediator. (Log: /tmp/ontap_mediator.GmRGdA/uninstall_ontap_mediator/remove.log)
+ Remove successful.
[root@mediator-host ~]#
```

Chapter 6. Testing the MetroCluster configuration

You can test failure scenarios to confirm the correct operation of the MetroCluster configuration.

Verifying negotiated switchover

You can test the negotiated (planned) switchover operation to confirm uninterrupted data availability.

About this task

This test validates that data availability is not affected (except for Microsoft Server Message Block (SMB)) by switching the cluster over to the second data center.

This test should take about 30 minutes.

This procedure has the following expected results:

- The `metrocluster switchover` command will present a warning prompt.

  If you respond `yes` to the prompt, the site the command from is issued will switch over to the partner site.
- Nodes at the partner site should shut down gracefully and remain at the LOADER> prompt.

Step 1. Confirm that all nodes are in the configured state and normal mode:
`metrocluster node show`

Example

```
cluster_A::>  metrocluster node show

Cluster                           Configuration State    Mode
--------------------------------- ---------------------- ------------------------
 Local: cluster_A                 configured             normal
Remote: cluster_B                  configured             normal
```

Step 2. Begin the switchover operation:
`metrocluster switchover`

Example

```
cluster_A::> metrocluster switchover
Warning: negotiated switchover is about to start. It will stop all the data Vservers
on cluster "cluster_B" and automatically re-start them on cluster "cluster_A". It will
finally gracefully shutdown cluster "cluster_B".
```

Step 3. Confirm that the local cluster is in the configured state and switchover mode:
`metrocluster node show`

Example

```
cluster_A::>  metrocluster node show

Cluster                           Configuration State    Mode
--------------------------------- ---------------------- ------------------------
Local: cluster_A                  configured             switchover
Remote: cluster_B                  not-reachable          -
               configured              normal
```

Step 4. Confirm that the switchover operation was successful:
`metrocluster operation show`

Example

```
cluster_A::>  metrocluster operation show

cluster_A::> metrocluster operation show
  Operation: switchover
      State: successful
 Start Time: 2/6/2016 13:28:50
   End Time: 2/6/2016 13:29:41
     Errors: -
```

Step 5. Use the `vserver show` and `network interface show` commands to verify that DR SVMs and LIFs have come online.

Verifying healing and manual switchback

You can test the healing and manual switchback operations to verify that data availability is not affected (except for SMB configurations) by switching back the cluster to the original data center after a negotiated switchover.

About this task

This test should take about 30 minutes.

The expected result of this procedure is that services should be switched back to their home nodes.

The healing steps are not required on systems running ONTAP 9.5 or later, on which healing is performed automatically after a negotiated switchover. On systems running ONTAP 9.6 and later, healing is also performed automatically after unscheduled switchover.

Step 1. If the system is running ONTAP 9.4 or earlier, heal the data aggregate:
`metrocluster heal aggregates`

Example

The following example shows the successful completion of the command:

```
cluster_A::> metrocluster heal aggregates
[Job 936] Job succeeded: Heal Aggregates is successful.
```

Step 2. If the system is running ONTAP 9.4 or earlier, heal the root aggregate:
`metrocluster heal root-aggregates`
This step is required on the following configurations:

- MetroCluster FC configurations.
- MetroCluster IP configurations running ONTAP 9.4 or earlier.

Example

The following example shows the successful completion of the command:

```
cluster_A::> metrocluster heal root-aggregates
[Job 937] Job succeeded: Heal Root Aggregates is successful.
```

Step 3. Verify that healing is completed:
`metrocluster node show`

Example

The following example shows the successful completion of the command:

```
cluster_A::> metrocluster node show
DR                               Configuration  DR
Group Cluster Node               State          Mirroring Mode
----- ------- ------------------ -------------- --------- --------------------
1     cluster_A
              node_A_1          configured     enabled   heal roots completed
      cluster_B
              node_B_2          unreachable    -         switched over
42 entries were displayed.metrocluster operation show
```

Example

If the automatic healing operation fails for any reason, you must issue the metrocluster heal commands manually as done in ONTAP versions prior to ONTAP 9.5. You can use the `metrocluster operation show` and `metrocluster operation history show -instance` commands to monitor the status of healing and determine the cause of a failure.

Step 4. Verify that all aggregates are mirrored:

```
storage aggregate show
```

Example

The following example shows that all aggregates have a RAID Status of `mirrored`:

```
cluster_A::> storage aggregate show
cluster Aggregates:
Aggregate Size     Available Used% State   #Vols  Nodes       RAID Status
--------- -------- --------- ----- ------- ------ ----------- ------------
data_cluster
             4.19TB    4.13TB    2% online       8 node_A_1    raid_dp,
                                                               mirrored,
                                                               normal
root_cluster
            715.5GB   212.7GB   70% online       1 node_A_1    raid4,
                                                               mirrored,
                                                               normal
cluster_B Switched Over Aggregates:
Aggregate Size     Available Used% State   #Vols  Nodes       RAID Status
--------- -------- --------- ----- ------- ------ ----------- ------------
data_cluster_B
             4.19TB    4.11TB    2% online       5 node_A_1    raid_dp,
                                                               mirrored,
                                                               normal
root_cluster_B    -          -     - unknown      - node_A_1   -
```

Step 5. Boot nodes from the disaster site.

Step 6. Check the status of switchback recovery:

```
metrocluster node show
```

Example

```
cluster_A::> metrocluster node show
DR                               Configuration  DR
Group Cluster Node               State          Mirroring Mode
----- ------- ------------------ -------------- --------- --------------------
1     cluster_A
              node_A_1           configured     enabled   heal roots completed
      cluster_B
              node_B_2           configured     enabled   waiting for switchback
                                                           recovery
2 entries were displayed.
```

Step 7. Perform the switchback:

```
metrocluster switchback
```

Example

```
cluster_A::> metrocluster switchback
[Job 938] Job succeeded: Switchback is successful.Verify switchback
```

Step 8. Confirm status of the nodes:
`metrocluster node show`

Example

```
cluster_A::> metrocluster node show
DR                               Configuration  DR
Group Cluster Node               State          Mirroring Mode
----- ------- ------------------ -------------- --------- --------------------
1     cluster_A
              node_A_1         configured     enabled   normal
      cluster_B
              node_B_2         configured     enabled   normal

2 entries were displayed.
```

Step 9. Confirm status of the metrocluster operation:
`metrocluster operation show`

Example

The output should show a successful state.

```
cluster_A::> metrocluster operation show
  Operation: switchback
      State: successful
 Start Time: 2/6/2016 13:54:25
   End Time: 2/6/2016 13:56:15
     Errors: -
```

Verifying operation after power line disruption

You can test the MetroCluster configuration's response to the failure of a PDU.

About this task

The best practice is for each power supply unit (PSU) in a component to be connected to separate power supplies. If both PSUs are connected to the same power distribution unit (PDU) and an electrical disruption occurs, the site could down or a complete shelf might become unavailable. Failure of one power line is tested to confirm that there is no cabling mismatch that could cause a service disruption.

This test should take about 15 minutes.

This test requires turning off power to all left-hand PDUs and then all right-hand PDUs on all of the racks containing the MetroCluster components.

This procedure has the following expected results:

- Errors should be generated as the PDUs are disconnected.
- No failover or loss of service should occur.

Step 1. Turn off the power of the PDUs on the left-hand side of the rack containing the MetroCluster components.

Step 2. Monitor the result on the console by using the `system environment sensors show -state fault` and `storage shelf show -errors` commands.

Example

```
cluster_A::> system environment sensors show -state fault

Node Sensor State Value/Units Crit-Low Warn-Low Warn-Hi Crit-Hi
---- --------------------- ------ ----------- -------- -------- ------- -------
node_A_1
PSU1 fault
PSU_OFF
PSU1 Pwr In OK fault
FAULT
node_A_2
PSU1 fault
PSU_OFF
PSU1 Pwr In OK fault
FAULT
4 entries were displayed.

cluster_A::> storage shelf show -errors
    Shelf Name: 1.1
     Shelf UID: 50:0a:09:80:03:6c:44:d5
 Serial Number: SHFHU1443000059

Error Type           Description
------------------   --------------------------
Power                Critical condition is detected in storage shelf power supply unit "1". The unit might fail.
                     Reconnect PSU1
```

Step 3. Turn the power back on to the left-hand PDUs.

Step 4. Make sure that ONTAP clears the error condition.

Step 5. Repeat the previous steps with the right-hand PDUs.

Verifying operation after loss of a single storage shelf

You can test the failure of a single storage shelf to verify that there is no single point of failure.

About this task

This procedure has the following expected results:

- An error message should be reported by the monitoring software.
- No failover or loss of service should occur.
- Mirror resynchronization starts automatically after the hardware failure is restored.

Step 1. Check the storage failover status:
`storage failover show`

Example

```
cluster_A::> storage failover show

Node            Partner        Possible State Description
--------------- -------------- -------- -------------------------------------
node_A_1        node_A_2       true     Connected to node_A_2
node_A_2        node_A_1       true     Connected to node_A_1
2 entries were displayed.
```

Step 2. Check the aggregate status:
`storage aggregate show`

Example

```
cluster_A::> storage aggregate show

cluster Aggregates:
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
node_A_1data01_mirrored
             4.15TB    3.40TB   18% online       3 node_A_1       raid_dp,
                                                                      mirrored,
                                                                      normal
node_A_1root
            707.7GB   34.29GB   95% online       1 node_A_1       raid_dp,
                                                                      mirrored,
                                                                      normal
node_A_2_data01_mirrored
             4.15TB    4.12TB    1% online       2 node_A_2       raid_dp,
                                                                      mirrored,
                                                                      normal
node_A_2_data02_unmirrored
             2.18TB    2.18TB    0% online       1 node_A_2       raid_dp,
                                                                      normal
node_A_2_root
            707.7GB   34.27GB   95% online       1 node_A_2       raid_dp,
                                                                      mirrored,
                                                                      normal
```

Step 3. Verify that all data SVMs and data volumes are online and serving data:
`vserver show -type data`
`network interface show -fields is-home false`
`volume show !vol0,!MDV*`

Example

```
cluster_A::> vserver show -type data

cluster_A::> vserver show -type data
                               Admin      Operational Root
Vserver     Type    Subtype    State      State        Volume     Aggregate
----------- ------- ---------- ---------- ------------ ---------- ----------
SVM1        data    sync-source           running      SVM1_root  node_A_1_data01_mirrored
SVM2        data    sync-source          running     SVM2_root  node_A_2_data01_mirrored

cluster_A::> network interface show -fields is-home false
There are no entries matching your query.

cluster_A::> volume show !vol0,!MDV*
Vserver   Volume       Aggregate    State      Type       Size  Available Used%
--------- ------------ ------------ ---------- ---- ---------- ---------- -----
SVM1
          SVM1_root
                       node_A_1data01_mirrored
                                    online     RW         10GB     9.50GB    5%
SVM1
          SVM1_data_vol
                       node_A_1data01_mirrored
                                    online     RW         10GB     9.49GB    5%
SVM2
          SVM2_root
                       node_A_2_data01_mirrored
                                    online     RW         10GB     9.49GB    5%
SVM2
          SVM2_data_vol
                       node_A_2_data02_unmirrored
                                    online     RW          1GB    972.6MB    5%
```

Step 4. Identify a shelf in Pool 1 for node node_A_2 to power off to simulate a sudden hardware failure:
`storage aggregate show -r -node` *node-name* `!*root`

The shelf you select must contain drives that are part of a mirrored data aggregate.

Example

In the following example, shelf ID 31 is selected to fail.

```
cluster_A::> storage aggregate show -r -node node_A_2 !*root
Owner Node: node_A_2
 Aggregate: node_A_2_data01_mirrored (online, raid_dp, mirrored) (block checksums)
  Plex: /node_A_2_data01_mirrored/plex0 (online, normal, active, pool0)
   RAID Group /node_A_2_data01_mirrored/plex0/rg0 (normal, block checksums)
                                                                      Usable Physical
     Position Disk                        Pool Type     RPM     Size     Size Status
     -------- --------------------------- ---- ----- ------ -------- -------- -----------
     dparity  2.30.3                       0   BSAS    7200  827.7GB  828.0GB (normal)
     parity   2.30.4                       0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.6                       0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.8                       0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.5                       0   BSAS    7200  827.7GB  828.0GB (normal)

  Plex: /node_A_2_data01_mirrored/plex4 (online, normal, active, pool1)
   RAID Group /node_A_2_data01_mirrored/plex4/rg0 (normal, block checksums)
                                                                      Usable Physical
     Position Disk                        Pool Type     RPM     Size     Size Status
     -------- --------------------------- ---- ----- ------ -------- -------- -----------
     dparity  1.31.7                       1   BSAS    7200  827.7GB  828.0GB (normal)
     parity   1.31.6                       1   BSAS    7200  827.7GB  828.0GB (normal)
     data     1.31.3                       1   BSAS    7200  827.7GB  828.0GB (normal)
     data     1.31.4                       1   BSAS    7200  827.7GB  828.0GB (normal)
     data     1.31.5                       1   BSAS    7200  827.7GB  828.0GB (normal)

 Aggregate: node_A_2_data02_unmirrored (online, raid_dp) (block checksums)
  Plex: /node_A_2_data02_unmirrored/plex0 (online, normal, active, pool0)
   RAID Group /node_A_2_data02_unmirrored/plex0/rg0 (normal, block checksums)
                                                                      Usable Physical
     Position Disk                        Pool Type     RPM     Size     Size Status
     -------- --------------------------- ---- ----- ------ -------- -------- -----------
     dparity  2.30.12                      0   BSAS    7200  827.7GB  828.0GB (normal)
     parity   2.30.22                      0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.21                      0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.20                      0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.14                      0   BSAS    7200  827.7GB  828.0GB (normal)
15 entries were displayed.
```

Step 5. Physically power off the shelf that you selected.

Step 6. Check the aggregate status again:

```
storage aggregate show
storage aggregate show -r -node node_A_2 !*root
```

Example

The aggregate with drives on the powered-off shelf should have a **degraded** RAID status, and drives on the affected plex should have a **failed** status, as shown in the following example:

```
cluster_A::> storage aggregate show
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
node_A_1data01_mirrored
             4.15TB    3.40TB   18% online       3 node_A_1       raid_dp,
                                                                    mirrored,
                                                                    normal
node_A_1root
            707.7GB   34.29GB   95% online       1 node_A_1       raid_dp,
                                                                    mirrored,
                                                                    normal
```

```
node_A_2_data01_mirrored
             4.15TB    4.12TB    1% online       2 node_A_2       raid_dp,
                                                                      mirror
                                                                      degraded
node_A_2_data02_unmirrored
             2.18TB    2.18TB    0% online       1 node_A_2       raid_dp,
                                                                      normal
node_A_2_root
            707.7GB   34.27GB   95% online       1 node_A_2       raid_dp,
                                                                      mirror
                                                                      degraded
cluster_A::> storage aggregate show -r -node node_A_2 !*root
Owner Node: node_A_2
 Aggregate: node_A_2_data01_mirrored (online, raid_dp, mirror degraded) (block checksums)
  Plex: /node_A_2_data01_mirrored/plex0 (online, normal, active, pool0)
   RAID Group /node_A_2_data01_mirrored/plex0/rg0 (normal, block checksums)
                                                                    Usable Physical
     Position Disk                      Pool Type     RPM     Size     Size Status
     -------- ------------------------- ---- ------ ------ -------- -------- ----------
     dparity  2.30.3                     0   BSAS    7200  827.7GB  828.0GB (normal)
     parity   2.30.4                     0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.6                     0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.8                     0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.5                     0   BSAS    7200  827.7GB  828.0GB (normal)

  Plex: /node_A_2_data01_mirrored/plex4 (offline, failed, inactive, pool1)
   RAID Group /node_A_2_data01_mirrored/plex4/rg0 (partial, none checksums)
                                                                    Usable Physical
     Position Disk                      Pool Type     RPM     Size     Size Status
     -------- ------------------------- ---- ------ ------ -------- -------- ----------
     dparity  FAILED                     -   -           -  827.7GB        - (failed)
     parity   FAILED                     -   -           -  827.7GB        - (failed)
     data     FAILED                     -   -           -  827.7GB        - (failed)
     data     FAILED                     -   -           -  827.7GB        - (failed)
     data     FAILED                     -   -           -  827.7GB        - (failed)

 Aggregate: node_A_2_data02_unmirrored (online, raid_dp) (block checksums)
  Plex: /node_A_2_data02_unmirrored/plex0 (online, normal, active, pool0)
   RAID Group /node_A_2_data02_unmirrored/plex0/rg0 (normal, block checksums)
                                                                    Usable Physical
     Position Disk                      Pool Type     RPM     Size     Size Status
     -------- ------------------------- ---- ------ ------ -------- -------- ----------
     dparity  2.30.12                    0   BSAS    7200  827.7GB  828.0GB (normal)
     parity   2.30.22                    0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.21                    0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.20                    0   BSAS    7200  827.7GB  828.0GB (normal)
     data     2.30.14                    0   BSAS    7200  827.7GB  828.0GB (normal)
15 entries were displayed.
```

Step 7. Verify that the data is being served and that all volumes are still online:

```
vserver show -type data
network interface show -fields is-home false
volume show !vol0,!MDV*
```

Example

```
cluster_A::> vserver show -type data

cluster_A::> vserver show -type data
                               Admin      Operational Root
Vserver     Type    Subtype    State      State       Volume     Aggregate
----------- ------- ---------- ---------- ----------- ---------- ----------
SVM1        data    sync-source           running     SVM1_root  node_A_1_data01_mirrored
SVM2        data    sync-source          running     SVM2_root  node_A_1_data01_mirrored

cluster_A::> network interface show -fields is-home false
There are no entries matching your query.
```

```
cluster_A::> volume show !vol0,!MDV*
Vserver   Volume       Aggregate    State      Type       Size  Available Used%
--------- ------------ ------------ ---------- ---- ---------- ---------- -----
SVM1
          SVM1_root
                       node_A_1data01_mirrored
                                    online     RW         10GB     9.50GB    5%
SVM1
          SVM1_data_vol
                       node_A_1data01_mirrored
                                    online     RW         10GB     9.49GB    5%
SVM2
          SVM2_root
                       node_A_1data01_mirrored
                                    online     RW         10GB     9.49GB    5%
SVM2
          SVM2_data_vol
                       node_A_2_data02_unmirrored
                                    online     RW          1GB    972.6MB    5%
```

Step 8. Physically power on the shelf.

Resynchronization starts automatically.

Step 9. Verify that resynchronization has started:
`storage aggregate show`

Example

The affected aggregate should have a **resyncing** RAID status, as shown in the following example:

```
cluster_A::> storage aggregate show
cluster Aggregates:
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
node_A_1_data01_mirrored
            4.15TB    3.40TB   18% online       3 node_A_1       raid_dp,
                                                                     mirrored,
                                                                     normal
node_A_1_root
           707.7GB   34.29GB   95% online       1 node_A_1       raid_dp,
                                                                     mirrored,
                                                                     normal
node_A_2_data01_mirrored
            4.15TB    4.12TB    1% online       2 node_A_2       raid_dp,
                                                                     resyncing
node_A_2_data02_unmirrored
            2.18TB    2.18TB    0% online       1 node_A_2       raid_dp,
                                                                     normal
node_A_2_root
           707.7GB   34.27GB   95% online       1 node_A_2       raid_dp,
                                                                     resyncing
```

Step 10. Monitor the aggregate to confirm that resynchronization is complete:
`storage aggregate show`

Example

The affected aggregate should have a **normal** RAID status, as shown in the following example:

```
cluster_A::> storage aggregate show
cluster Aggregates:
Aggregate     Size Available Used% State   #Vols  Nodes            RAID Status
--------- -------- --------- ----- ------- ------ ---------------- ------------
node_A_1data01_mirrored
            4.15TB    3.40TB   18% online       3 node_A_1       raid_dp,
```

```
                                                                    mirrored,
                                                                    normal
node_A_1root
            707.7GB    34.29GB    95% online       1 node_A_1       raid_dp,
                                                                    mirrored,
                                                                    normal
node_A_2_data01_mirrored
             4.15TB     4.12TB     1% online       2 node_A_2       raid_dp,
                                                                    normal
node_A_2_data02_unmirrored
             2.18TB     2.18TB     0% online       1 node_A_2       raid_dp,
                                                                    normal
node_A_2_root
            707.7GB    34.27GB    95% online       1 node_A_2       raid_dp,
                                                                    resyncing
```

Chapter 7. Considerations when removing MetroCluster configurations

After removing the MetroCluster configuration, all disk connectivity and interconnects should be adjusted to be in a supported state. If you need to remove the MetroCluster configuration, contact technical support.

Attention: You cannot reverse the MetroCluster unconfiguration. This process should only be done with the assistance of technical support.

Chapter 8. Considerations when using ONTAP in a MetroCluster configuration

When using ONTAP in a MetroCluster configuration, you should be aware of certain considerations for licensing, peering to clusters outside the MetroCluster configuration, performing volume operations, NVFAIL operations, and other ONTAP operations.

Licensing considerations

- Both sites should be licensed for the same site-licensed features.
- All nodes should be licensed for the same node-locked features.

SnapMirror consideration

- SnapMirror SVM disaster recovery is only supported on MetroCluster configurations running versions of ONTAP 9.5 or later.

FlexCache support in a MetroCluster configuration

Starting with ONTAP 9.7, FlexCache volumes are supported on MetroCluster configurations. You should be aware of requirements for manual repeering after switchover or switchback operations.

SVM repeering after switchover when FlexCache origin and cache are within the same MetroCluster site

After a negotiated or unplanned switchover, any SVM FlexCache peering relationship within the cluster must be manually configured.

For example, SVMs vs1 (cache) and vs2 (origin) are on site_A. These SVMs are peered.

After switchover, SVMs vs1-mc and vs2-mc are activated at the partner site (site_B). They must be manually repeered for FlexCache to work using the `vserver peer repeer` command.

SVM repeering after switchover or switchback when a FlexCache destination is on a third cluster and in disconnected mode

For FlexCache relationships to a cluster outside of the MetroCluster configuration, the peering must always be manually reconfigured after a switchover if the involved clusters are in disconnected mode during switchover.

For example:

- One end of the FlexCache (cache_1 on vs1) resides on MetroCluster site_A has one end of the FlexCache
- The other end of the FlexCache (origin_1 on vs2) resides on site_C (not in the MetroCluster configuration)

When switchover is triggered, and if site_A and site_C are not connected, you must manually repeer the SVMs on site_B (the switchover cluster) and site_C using the `vserver peer repeer` command after the switchover.

When switchback is performed, you must again repeer the SVMs on site_A (the original cluster) and site_C.

FabricPool support in MetroCluster configurations

Starting with ONTAP 9.7, MetroCluster configurations support FabricPool storage tiers.

For general information on using FabricPools, see the *Disks and Aggregates Power Guide*.

Disk and Aggregate Management Guide

Considerations when using FabricPools

- The clusters must have FabricPool licenses with matching capacity limits.
- The clusters must have IPSpaces with matching names.

  This can be the default IPSpace, or an IP space an administer has created. This IPSpace will be used for FabricPool object store configuration setups.
- For the selected IPSpace, each cluster must have an intercluster LIF defined that can reach the external object store.

Configuring an aggregate for use in a mirrored FabricPool

Note: Before you configure the aggregate you must set up object stores as described in "Setting up object stores for FabricPool in a MetroCluster configuration" in the *Disks and Aggregates Power Guide*.

Disk and Aggregate Management Guide

To configure an aggregate for use in a FabricPool:

1. Create the aggregate or select an existing aggregate.
2. Mirror the aggregate as a typical mirrored aggregate within the MetroCluster configuration.
3. Create the FabricPool mirror with the aggregate, as described in the *Disks and Aggregates Power Guide*.

   Disk and Aggregate Management Guide

   a. Attach a primary object store.

      This object store is physically closer to the cluster.

   b. Add a mirror object store.

      This object store is physically further distant to the cluster than the primary object store.

FlexGroup support in MetroCluster configurations

Starting with ONTAP 9.6 MetroCluster configurations support FlexGroup volumes.

Job schedules in a MetroCluster configuration

In ONTAP 9.4 and later, user-created job schedules are automatically replicated between clusters in a MetroCluster configuration. If you create, modify, or delete a job schedule on a cluster, the same schedule is automatically created on the partner cluster, using Configuration Replication Service (CRS).

Note: System-created schedules are not replicated and you must manually perform the same operation on the partner cluster so that job schedules on both clusters are identical.

Cluster peering from the MetroCluster site to a third cluster

Because the peering configuration is not replicated, if you peer one of the clusters in the MetroCluster configuration to a third cluster outside of that configuration, you must also configure the peering on the partner MetroCluster cluster. This is so that peering can be maintained if a switchover occurs.

The non-MetroCluster cluster must be running ONTAP 9.4 or later. If not, peering is lost if a switchover occurs even if the peering has been configured on both MetroCluster partners.

LDAP client configuration replication in a MetroCluster configuration

An LDAP client configuration created on a storage virtual machine (SVM) on a local cluster is replicated to its partner data SVM on the remote cluster. For example, if the LDAP client configuration is created on the admin SVM on the local cluster, then it is replicated to all the admin data SVMs on the remote cluster. This MetroCluster feature is intentional so that the LDAP client configuration is active on all the partner SVMs on the remote cluster.

Networking and LIF creation guidelines for MetroCluster configurations

You should be aware of how LIFs are created and replicated in a MetroCluster configuration. You must also know about the requirement for consistency so that you can make proper decisions when configuring your network.

IPspace object replication and subnet configuration requirements

You should be aware of the requirements for replicating IPspace objects to the partner cluster and for configuring subnets and IPv6 in a MetroCluster configuration.

IPspace replication

You must consider the following guidelines while replicating IPspace objects to the partner cluster:

- The IPspace names of the two sites must match.
- IPspace objects must be manually replicated to the partner cluster.

  Any storage virtual machines (SVMs) that are created and assigned to an IPspace before the IPspace is replicated will not be replicated to the partner cluster.

Subnet configuration

You must consider the following guidelines while configuring subnets in a MetroCluster configuration:

- Both clusters of the MetroCluster configuration must have a subnet in the same IPspace with the same subnet name, subnet, broadcast domain, and gateway.
- The IP ranges of the two clusters must be different.

  In the following example, the IP ranges are different:

```
cluster_A::> network subnet show

IPspace: Default
Subnet                       Broadcast                    Avail/
Name      Subnet             Domain    Gateway            Total    Ranges
--------- ---------------- --------- ------------       -------  ---------------
subnet1   192.168.2.0/24   Default   192.168.2.1        10/10    192.168.2.11-192.168.2.20

cluster_B::> network subnet show
 IPspace: Default
Subnet                       Broadcast                    Avail/
Name      Subnet             Domain    Gateway            Total    Ranges
--------- ---------------- --------- ------------      --------  ---------------
subnet1   192.168.2.0/24   Default   192.168.2.1        10/10    192.168.2.21-192.168.2.30
```

IPv6 configuration

If IPv6 is configured on one site, IPv6 must be configured on the other site as well.

Requirements for LIF creation in a MetroCluster configuration

You should be aware of the requirements for creating LIFs when configuring your network in a MetroCluster configuration.

You must consider the following guidelines when creating LIFs:

- Fibre Channel: You must use stretched VSAN or stretched fabrics
- IP/iSCSI: You must use layer 2 stretched network
- ARP broadcasts: You must enable ARP broadcasts between the two clusters
- Duplicate LIFs: You must not create multiple LIFs with the same IP address (duplicate LIFs) in an IPspace

Verify LIF creation

You can confirm the successful creation of a LIF in a MetroCluster configuration by running the `metrocluster check lif show` command. If you encounter any issues while creating the LIF, you can use the `metrocluster check lif repair-placement` command to fix the issues.

LIF replication and placement requirements and issues

You should be aware of the LIF replication requirements in a MetroCluster configuration. You should also know how a replicated LIF is placed on a partner cluster, and you should be aware of the issues that occur when LIF replication or LIF placement fails.

Replication of LIFs to the partner cluster

When you create a LIF on a cluster in a MetroCluster configuration, the LIF is replicated on the partner cluster. LIFs are not placed on a one-to-one name basis. For availability of LIFs after a switchover operation, the LIF placement process verifies that the ports are able to host the LIF based on reachability and port attribute checks.

The system must meet the following conditions to place the replicated LIFs on the partner cluster:

Condition	LIF type: FC	LIF type: IP/iSCSI
Node identification	ONTAP attempts to place the replicated LIF on the disaster recovery (DR) partner of the node on which it was created. If the DR partner is unavailable, the DR auxiliary partner is used for placement.	ONTAP attempts to place the replicated LIF on the DR partner of the node on which it was created. If the DR partner is unavailable, the DR auxiliary partner is used for placement.
Port identification	ONTAP identifies the connected FC target ports on the DR cluster.	The ports on the DR cluster that are in the same IPspace as the source LIF are selected for a reachability check. If there are no ports in the DR cluster in the same IPspace, the LIF cannot be placed. All of the ports in the DR cluster that are already hosting a LIF in the same IPspace and subnet are automatically marked as reachable; and can be used for placement. These ports are not included in the reachability check.

Condition	LIF type: FC	LIF type: IP/iSCSI
Reachability check	Reachability is determined by checking for the connectivity of the source fabric WWN on the ports in the DR cluster. If the same fabric is not present at the DR site, the LIF is placed on a random port on the DR partner.	Reachability is determined by the response to an Address Resolution Protocol (ARP) broadcast from each previously identified port on the DR cluster to the source IP address of the LIF to be placed. For reachability checks to succeed, ARP broadcasts must be allowed between the two clusters. Each port that receives a response from the source LIF will be marked as possible for placement.
Port selection	ONTAP categorizes the ports based on attributes such as adapter type and speed, and then selects the ports with matching attributes. If no ports with matching attributes are found, the LIF is placed on a random connected port on the DR partner.	From the ports that are marked as reachable during the reachability check, ONTAP prefers ports that are in the broadcast domain that is associated with the subnet of the LIF. If there are no network ports available on the DR cluster that are in the broadcast domain that is associated with the subnet of the LIF, then ONTAP selects ports that have reachability to the source LIF. If there are no ports with reachability to the source LIF, a port is selected from the broadcast domain that is associated with the subnet of the source LIF, and if no such broadcast domain exists, a random port is selected. ONTAP categorizes the ports based on attributes such as adapter type, interface type, and speed, and then selects the ports with matching attributes.
LIF placement	From the reachable ports, ONTAP selects the least loaded port for placement.	From the selected ports, ONTAP selects the least loaded port for placement.

Placement of replicated LIFs when the DR partner node is down

When an iSCSI or FC LIF is created on a node whose DR partner has been taken over, the replicated LIF is placed on the DR auxiliary partner node. After a subsequent giveback operation, the LIFs are not automatically moved to the DR partner. This can lead to LIFs being concentrated on a single node in the partner cluster. During a MetroCluster switchover operation, subsequent attempts to map LUNs belonging to the storage virtual machine (SVM) fail.

You should run the `metrocluster check lif show` command after a takeover operation or giveback operation to verify that the LIF placement is correct. If errors exist, you can run the `metrocluster check lif repair-placement` command to resolve the issues.

LIF placement errors

LIF placement errors that are displayed by the `metrocluster check lif show` command are retained after a switchover operation. If the `network interface modify`, `network interface rename`, or `network interface delete` command is issued for a LIF with a placement error, the error is removed and does not appear in the output of the `metrocluster check lif show` command.

LIF replication failure

You can also check whether LIF replication was successful by using the `metrocluster check lif show` command. An EMS message is displayed if LIF replication fails.

You can correct a replication failure by running the `metrocluster check lif repair-placement` command for any LIF that fails to find a correct port. You should resolve any LIF replication failures as soon as possible to verify the availability of LIF during a MetroCluster switchover operation.

Note: Even if the source SVM is down, LIF placement might proceed normally if there is a LIF belonging to a different SVM in a port with the same IPspace and network in the destination SVM.

Volume creation on a root aggregate

The system does not allow the creation of new volumes on the root aggregate (an aggregate with an HA policy of `CFO`) of a node in a MetroCluster configuration.

Because of this restriction, root aggregates cannot be added to an SVM using the `vserver add-aggregates` command.

SVM disaster recovery in a MetroCluster configuration

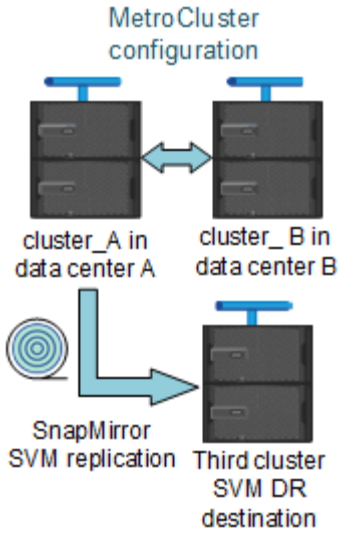
Starting with ONTAP 9.5, active storage virtual machines (SVMs) in a MetroCluster configuration can be used as sources with the SnapMirror SVM disaster recovery feature. The destination SVM must be on the third cluster outside of the MetroCluster configuration.

You should be aware of the following requirements and limitations of using SVMs with SnapMirror disaster recovery:

- Only an active SVM within a MetroCluster configuration can be the source of an SVM disaster recovery relationship.

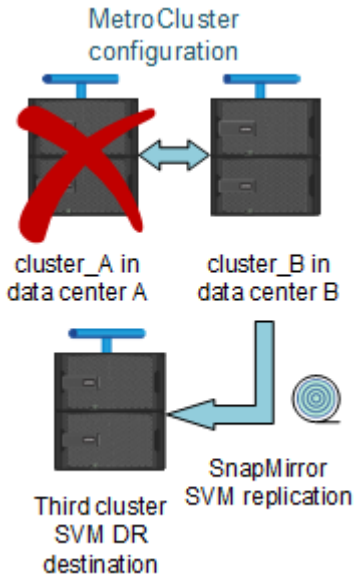
  A source can be a sync-source SVM before switchover or a sync-destination SVM after switchover.
- When a MetroCluster configuration is in a steady state, the MetroCluster sync-destination SVM cannot be the source of an SVM disaster recovery relationship, since the volumes are not online.

  The following image shows the SVM disaster recovery behavior in a steady state:



- When the sync-source SVM is the source of an SVM DR relationship, the source SVM DR relationship information is replicated to the MetroCluster partner.

  This enables the SVM DR updates to continue after a switchover as shown in the following image:



- During the switchover and switchback processes, replication to the SVM DR destination might fail.

  However, after the switchover or switchback process completes, the next SVM DR scheduled updates will succeed.

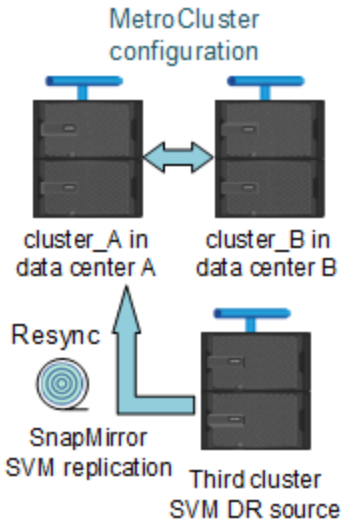
See the section “Replicating the SVM configuration” in the *Data Protection Power Guide* for details on configuring an SVM DR relationship.

Data Protection Power Guide

SVM resynchronization at a disaster recovery site

During resynchronization, the storage virtual machines (SVMs) disaster recovery (DR) source on the MetroCluster configuration is restored from the destination SVM on the non-MetroCluster site.

During resynchronization, the source SVM (cluster_A) temporarily acts as a destination SVM as shown in the following image:



If an unplanned switchover occurs during resynchronization

Unplanned switchovers that occur during the resynchronization will halt the resynchronization transfer. If an unplanned switchover occurs, the following conditions are true:

- The destination SVM on the MetroCluster site (which was a source SVM prior to resynchronization) remains as a destination SVM. The SVM at the partner cluster will continue to retain its subtype and remain inactive.
- The SnapMirror relationship must be re-created manually with the sync-destination SVM as the destination.
- The SnapMirror relationship does not appear in the SnapMirror show output after a switchover at the survivor site unless a SnapMirror create operation is executed.

Performing switchback after an unplanned switchover during resynchronization

To successfully perform the switchback process, the resynchronization relationship must be broken and deleted. Switchback is not permitted if there are any SnapMirror DR destination SVMs in the MetroCluster configuration or if the cluster has an SVM of subtype “dp-destination”.

Output for the storage aggregate plex show command is indeterminate after a MetroCluster switchover

When you run the `storage aggregate plex show` command after a MetroCluster switchover, the status of plex0 of the switched over root aggregate is indeterminate and is displayed as `failed`. During this time, the switched over root is not updated. The actual status of this plex can only be determined after the MetroCluster healing phase.

Modifying volumes to set the NVFAIL flag in case of switchover

You can modify a volume so that the NVFAIL flag is set on the volume in the event of a MetroCluster switchover. The NVFAIL flag causes the volume to be fenced off from any modification. This is required for volumes that need to be handled as if committed writes to the volume were lost after the switchover.

About this task

Note: In ONTAP 9.4 and later versions, the unplanned switchover (USO) is used.

Step 1. Enable MetroCluster configuration to trigger NVFAIL on switchover by setting the vol -dr-force-nvfail parameter to on :
`vol modify -vserver` *vserver-name* `-volume` *volume-name* `-dr-force-nvfail on`

Monitoring and protecting the file system consistency using NVFAIL

The -nvfail parameter of the `volume modify` command enables ONTAP to detect nonvolatile RAM (NVRAM) inconsistencies when the system is booting or after a switchover operation. It also warns you and protects the system against data access and modification until the volume can be manually recovered.

If ONTAP detects any problems, database or file system instances stop responding or shut down. ONTAP then sends error messages to the console to alert you to check the state of the database or file system. You can enable NVFAIL to warn database administrators of NVRAM inconsistencies among clustered nodes that can compromise database validity.

After the NVRAM data loss during failover or boot recovery, NFS clients cannot access data from any of the nodes until the NVFAIL state is cleared. CIFS clients are unaffected.

How NVFAIL impacts access to NFS volumes or LUNs

The NVFAIL state is set when ONTAP detects NVRAM errors when booting, when a MetroCluster switchover operation occurs, or during an HA takeover operation if the NVFAIL option is set on the volume. If no errors are detected at startup, the file service is started normally. However, if NVRAM errors are detected or NVFAIL processing is enforced on a disaster switchover, ONTAP stops database instances from responding.

When you enable the NVFAIL option, one of the processes described in the following table takes place during bootup:

If...	Then...
ONTAP detects no NVRAM errors	File service starts normally.
ONTAP detects NVRAM errors	• ONTAP returns a stale file handle (`ESTALE`) error to NFS clients trying to access the database, causing the application to stop responding, crash, or shut down. ONTAP then sends an error message to the system console and log file. • When the application restarts, files are available to CIFS clients even if you have not verified that they are valid. For NFS clients, files remain inaccessible until you reset the in-nvfailed-state option on the affected volume.

If...	Then...
If one of the following parameters is used: • dr-force-nvfail volume option is set • force-nvfail-all switchover command option is set.	You can unset the dr-force-nvfail option after the switchover, if the administrator is not expecting to force NVFAIL processing for possible future disaster switchover operations. For NFS clients, files remain inaccessible until you reset the in-nvfailed-state option on the affected volume. **Note:** Using the force-nvfail-all option causes the dr-force-nvfail option to be set on all of the DR volumes processed during the disaster switchover.
ONTAP detects NVRAM errors on a volume that contains LUNs	LUNs in that volume are brought offline. The in-nvfailed-state option on the volume must be cleared, and the `NVFAIL` attribute on the LUNs must be cleared by bringing each LUN in the affected volume online. You can perform the steps to check the integrity of the LUNs and recover the LUN from a Snapshot copy or back up as necessary. After all of the LUNs in the volume are recovered, the in-nvfailed-state option on the affected volume is cleared.

Commands for monitoring data loss events

If you enable the NVFAIL option, you receive notification when a system crash caused by NVRAM inconsistencies or a MetroCluster switchover occurs.

By default, the NVFAIL parameter is not enabled.

If you want to...	Use this command...
Create a new volume with NVFAIL enabled	`volume create` -nvfail on
Enable NVFAIL on an existing volume	`volume modify` **Note:** You set the -nvfail option to on to enable NVFAIL on the created volume.
Display whether NVFAIL is currently enabled for a specified volume	`volume show` **Note:** You set the -fields parameter to nvfail to display the NVFAIL attribute for a specified volume.

See the man page for each command for more information.

Accessing volumes in NVFAIL state after a switchover

After a switchover, you must clear the NVFAIL state by resetting the -in-nvfailed-state parameter of the `volume modify` command to remove the restriction of clients to access data.

Before you begin

The database or file system must not be running or trying to access the affected volume.

About this task

Setting -in-nvfailed-state parameter requires advanced-level privilege.

Step 1. Recover the volume by using the `volume modify` command with the -in-nvfailed-state parameter set to false .

After you finish

For instructions about examining database file validity, see the documentation for your specific database software.

If your database uses LUNs, review the steps to make the LUNs accessible to the host after an NVRAM failure.

Recovering LUNs in NVFAIL states after switchover

After a switchover, the host no longer has access to data on the LUNs that are in NVFAIL states. You must perform a number of actions before the database has access to the LUNs.

Before you begin

The database must not be running.

Step 1. Clear the NVFAIL state on the affect volume that hosts the LUNs by resetting the -in-nvfailed-state parameter of the `volume modify` command.

Step 2. Bring the affected LUNs online.

Step 3. Examine the LUNs for any data inconsistencies and resolve them.

This might involve host-based recovery or recovery done on the storage controller using SnapRestore.

Step 4. Bring the database application online after recovering the LUNs.

Chapter 9. Where to find additional information

You can learn more about MetroCluster IP configuration and operation from the Lenovo documentation library.

MetroCluster and miscellaneous guides

Guide	Content
ONTAP 9 Documentation Center	• All MetroCluster guides
Fabric-attached MetroCluster installation and configuration	• Fabric-attached MetroCluster architecture • Cabling the configuration • Configuring the FC-to-SAS bridges • Configuring the FC switches • Configuring the MetroCluster in ONTAP
MetroCluster management and disaster recovery	• Understanding the MetroCluster configuration • Switchover, healing and switchback • Disaster recovery
MetroCluster Tiebreaker Software Installation and Configuration Guide	• Monitoring the MetroCluster configuration with the MetroCluster Tiebreaker software
AFA and Hybrid Documentation Center **Note:** The standard storage shelf maintenance procedures can be used with MetroCluster IP configurations.	• Hot-adding a disk shelf • Hot-removing a disk shelf
ONTAP concepts	• How mirrored aggregates work

Appendix A. Contacting Support

You can contact Support to obtain help for your issue.

You can receive hardware service through a Lenovo Authorized Service Provider. To locate a service provider authorized by Lenovo to provide warranty service, go to https://datacentersupport.lenovo.com/serviceprovider and use filter searching for different countries. For Lenovo support telephone numbers, see https://datacentersupport.lenovo.com/supportphonelist for your region support details.

Appendix B. Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area.

Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document is not an offer and does not provide a license under any patents or patent applications. You can send inquiries in writing to the following:

Lenovo (United States), Inc.
8001 Development Drive
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Trademarks