Install a MetroCluster IP configuration

ONTAP MetroCluster

NetApp
September 06, 2022

Table of Contents

Install a MetroCluster IP configuration

Overview

To install your MetroCluster IP configuration, you must perform a number of procedures in the correct order.

- Prepare for the installation and understand all requirements.
- Cable the components
- Configure the software
- Configure ONTAP mediator (optional)
- Test the configuration

Prepare for the MetroCluster installation

Differences among the ONTAP MetroCluster configurations

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

In all configurations, each of the two MetroCluster sites are 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		Stretch configurations	
		Four- or eight-node	Two-node	Two-node bridge-attached	Two-node direct-attached
Number of controllers	Four or eight*	Four or eight	Two	Two	Two
Uses an FC switch storage fabric	No	Yes	Yes	No	No
Uses an IP switch storage fabric	Yes	No	No	No	No
Uses FC-to-SAS bridges	No	Yes	Yes	Yes	No
Uses direct-attached SAS storage	Yes (local attached only)	No	No	No	Yes

Supports ADP	Yes (beginning with ONTAP 9.4)	No	No	No	No
Supports local HA	Yes	Yes	No	No	No
Supports ONTAP AUSO	No	Yes	Yes	Yes	Yes
Supports unmirrored aggregates	Yes (beginning with ONTAP 9.8)	Yes	Yes	Yes	Yes
Supports array LUNs	No	Yes	Yes	Yes	Yes
Supports ONTAP Mediator	Yes (beginning with ONTAP 9.7)	No	No	No	No
Supports MetroCluster Tiebreaker	Yes (not in combination with ONTAP Mediator)	Yes	Yes	Yes	Yes
Supports All SAN Arrays	Yes	Yes	Yes	Yes	Yes

Important

Notice the following considerations for eight-node MetroCluster IP configurations:

- Eight-node configurations are supported beginning with ONTAP 9.9.1.
- Only NetApp-validated MetroCluster switches (ordered from NetApp) are supported.
- Configurations using IP-routed (layer 3) backend connections are not supported.
- Configurations using shared private layer 2 networks are not supported.
- Configurations using a Cisco 9336C-FX2 shared switch are not supported.

Support for All SAN Array systems in MetroCluster configurations

Some of the All SAN Arrays (ASAs) are supported in MetroCluster configurations. In the MetroCluster documentation, the information for AFF models applies to the corresponding ASA system. For example, all cabling and other information for the AFF A400 system also applies to the ASA AFF A400 system.

Supported platform configurations are listed in the NetApp Hardware Universe.

Differences between ONTAP Mediator and MetroCluster Tiebreaker

Beginning 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).

  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 MetroCluster Tiebreaker Software installation and configuration

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.

A MAUSO is only initiated in the following scenarios:

- 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.
- None of the nodes at the surviving site are in takeover state.

A MAUSO is only initiated if a site disaster occurs. A site disaster is a failure of **all** nodes at the same site; however, there are some exceptions.

A MAUSO is **not** initiated in the following shutdown scenarios:

- You initiate a shutdown. For example, when you:

- halt the nodes
- reboot the nodes

- A fan or component failure initiates a shutdown (environmental shutdown)

Considerations for MetroCluster IP configurations

You should understand how the controllers access the remote storage and how the MetroCluster IP addresses work.

Access to remote storage in MetroCluster IP configurations

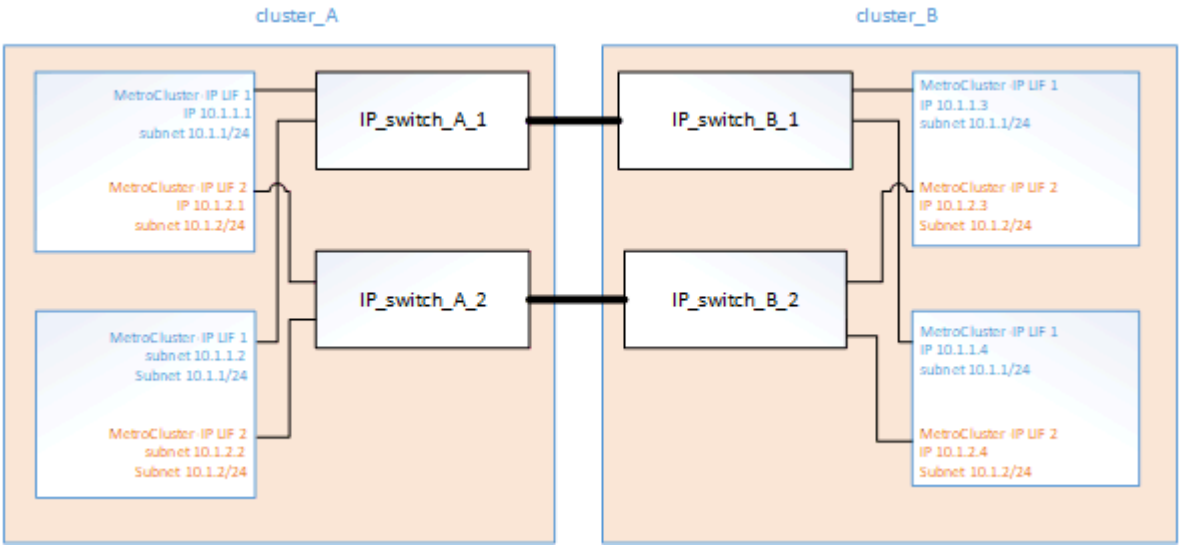
In MetroCluster IP configurations, the only way the local controllers can reach the remote storage pools is via the remote controllers. The IP switches are connected to the Ethernet ports on the controllers; they do not have direct connections to the disk shelves. If the remote controller is down, the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

MetroCluster IP addresses

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:

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
node_A_1	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
node_A_2	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
node_B_1	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
node_B_2	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.

  Beginning with ONTAP 9.9.1, if you are using a layer 3 configuration, you must also specify the `-gateway` parameter when creating MetroCluster IP interfaces. Refer to Considerations for layer 3 wide-area networks.

  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 in ONTAP 9.4 and later

Beginning with ONTAP 9.4, MetroCluster IP configurations support new installations with AFF systems using ADP (Advanced Drive Partitioning). In most configurations, partitioning and disk assignment are 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.

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.

Beginning with ONTAP 9.5, automatic assignment of disks must be enabled with the command `storage disk option modify -autoassign on`.

A maximum of 96 drives can be automatically partitioned during installation. You can add extra drives after the initial installation.

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

How to populate partially-full shelves

If your configuration is using shelves that are not fully populated (have empty drive bays) you must distribute the drives evenly throughout the shelf, depending on the disk assignment policy. The disk assignment policy depends on how many shelves are at each MetroCluster site.

If you are using a single shelf at each site (or just the internal shelf on an AFF A800 system), disks are assigned using a quarter-shelf policy. If the shelf is not fully populated, install the drives equally on all quarters.

The following table shows an example of how to place 24 disks in a 48 drive internal shelf. The ownership for the drives is also shown.

The 48 drive bays are divided into four quarters:	**Install six drives in the first six bays in each quarter…**
Quarter 1: Bays 0-11	Bays 0-5
Quarter 2: Bays 12-23	Bays 12-17
Quarter 3: Bays 24-35	Bays 24-29
Quarter 4: Bays 36-48	Bays 36-41

If you are using two shelves at each site, disks are assigned using a half-shelf policy. If the shelves are not fully populated, install the drives equally from either end of the shelf.

For example, if you are installing 12 drives in a 24-drive shelf, install drives in bays 0-5 and 18-23.

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.

Related information

Required MetroCluster IP components and naming conventions

Disk and aggregate management

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.

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.

NetApp Hardware Universe

ADP and disk assignment on AFF A320 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition

Minimum recommended drives (per site)	48 drives	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. Each of the two plexes in the root aggregate includes the following partitions** • Eight partitions for data • Two parity partitions • Two spare partitions
Minimum supported drives (per site)	24 drives	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

ADP and disk assignment on AFF A220 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition

Minimum recommended drives (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. 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 drives (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 ⓘ 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 AFF A250 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition

Minimum recommended drives (per site)	48 drives	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 drives (per site)	16 internal drives only	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: • Two partitions for data • Two parity partitions • No spare partitions

ADP and disk assignment on AFF A300 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition
Minimum recommended drives (per site)	48 drives	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 drives (per site)	24 drives	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

ADP and disk assignment on AFF A400 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition
Minimum recommended drives (per site)	96 drives	Drives are automatically assigned on a shelf-by-shelf basis.	Each of the two plexes in the root aggregate includes: • 20 partitions for data • Two parity partitions • Two spare partitions
Minimum supported drives (per site)	24 drives	The drives are divided into four equal groups (quarters). Each quarter-shelf is automatically assigned to a separate pool.	Each of the two plexes in the root aggregate includes: • Three partitions for data • Two parity partitions • One spare partition

ADP and disk assignment on AFF A700 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition
Minimum recommended drives (per site)	96 drives	Drives are automatically assigned on a shelf-by-shelf basis.	Each of the two plexes in the root aggregate includes: • 20 partitions for data • Two parity partitions • Two spare partitions

Minimum supported drives (per site)	24 drives	The drives are divided into four equal groups (quarters). Each quarter-shelf is automatically assigned to a separate pool.	Each of the two plexes in the root aggregate includes: • Three partitions for data • Two parity partitions • One spare partition

ADP and disk assignment on AFF A800 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root aggregate
Minimum recommended drives (per site)	Internal drives and 96 external drives	The internal partitions are divided into four equal groups (quarters). Each quarter is automatically assigned to a separate pool. The drives on the external shelves are automatically assigned on a shelf-by-shelf basis, with all of the drives on each shelf assigned to one of the four nodes in the MetroCluster configuration.	The root aggregate is created with 12 root partitions on the internal shelf. Each of the two plexes in the root aggregate includes: • Eight partitions for data • Two parity partitions • Two spare partitions
Minimum supported drives (per site)	24 internal drives only	The internal partitions are divided into four equal groups (quarters). Each quarter is automatically assigned to a separate pool.	The root aggregate is created with 12 root partitions on the internal shelf. Each of the two plexes in the root aggregate includes: • Three partitions for data • Two parity partitions • One spare partitions

ADP and disk assignment on AFF A900 systems

Guideline	Shelves per site	Drive assignment rules	ADP layout for root partition

Minimum recommended drives (per site)	96 drives	Drives are automatically assigned on a shelf-by-shelf basis.	Each of the two plexes in the root aggregate includes: • 20 partitions for data • Two parity partitions • Two spare partitions
Minimum supported drives (per site)	24 drives	The drives are divided into four equal groups (quarters). Each quarter-shelf is automatically assigned to a separate pool.	Each of the two plexes in the root aggregate includes: • Three partitions for data • Two parity partitions • One spare partition

Disk assignment on FAS2750 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition
Minimum recommended drives (per site)	24 internal drives and 24 external drives	The internal and external shelves are divided into two equal halves. Each half is automatically assigned to different pool	Not applicable
Minimum supported drives (per site) (active/passive HA configuration)	Internal drives only	Manual assignment required	Not applicable

Disk assignment on FAS8200 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition
Minimum recommended drives (per site)	48 drives	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 drives (per site) (active/passive HA configuration)	24 drives	Manual assignment required.	Not applicable

Disk assignment on FAS500f systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition
Minimum recommended drives (per site)	96 drives	Drives are automatically assigned on a shelf-by-shelf basis.	Not applicable
Minimum supported drives (per site)	24 drives	The drives are divided into four equal groups. Each quarter-shelf is automatically assigned to a separate pool.	Not applicable

Disk assignment on FAS9000 systems

Guideline	Drives per site	Drive assignment rules	ADP layout for root partition
Minimum recommended drives (per site)	96 drives	Drives are automatically assigned on a shelf-by-shelf basis.	Not applicable
Minimum supported drives (per site)	48 drives	The drives on the shelves are divided into two equal groups (halves). Each half-shelf is automatically assigned to a separate pool.	Minimum supported drives (per site) (active/passive HA configuration)

Disk assignment on FAS9500 systems

Guideline	Shelves per site	Drive assignment rules	ADP layout for root partition
Minimum recommended drives (per site)	96 drives	Drives are automatically assigned on a shelf-by-shelf basis.	Not applicable

Minimum supported drives (per site)	24 drives	The drives are divided into four equal groups (quarters). Each quarter-shelf is automatically assigned to a separate pool.	Minimum supported drives (per site) (active/passive HA configuration)

Cluster peering

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

Related information

Cluster and SVM peering express configuration

Prerequisites for cluster peering

Before you set up cluster peering, you should confirm that connectivity between 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 four-node cluster, the subnet used for intercluster communication must have four 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.

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 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 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 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. This 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 the 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. This 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 ISLs

You should know the ISL requirements for your configuration.

Basic MetroCluster ISL requirements

The following requirements must be met for the ISLs on all MetroCluster IP configurations:

- 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.

ISL requirements in shared layer 2 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 customer network must provide the same VLANs between the sites matching the MetroCluster VLANs as set in the RCF file.

  Layer 2 VLANs with IDs that match the MetroCluster VLAN IDs must span the shared network.

  - In ONTAP 9.7 and earlier, FAS2750 and AFF A220 systems require VLAN 10 and 20.
  - In ONTAP 9.8 and later, FAS2750, AFF A220, FAS500f, AFF A250, FAS8300, AFF A400, and FAS8700 systems use VLAN 10 and 20 by default. You can configure other VLANs during interface creation, and they must be withing the range 101-4096. For all the platforms mentioned previously, you can only specify the VLAN during interface creation. Once the MetroCluster interfaces are created, the VLAN ID cannot not be changed. For all other platforms not mentioned previously, you can use any VLAN and you can change the VLAN ID for those platforms at any time, but it requires that a new RCF file is created and applied.

> The RcfFileGenerator does not allow the creation of an RCF file using VLANs that are not supported by the platform.
>
> The RcfFileGenerator might restrict the use of certain VLAN IDs (for example, if they are intended for future use). Generally, reserved VLANs are up to and including 100.

- 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.

Cabling requirements when using shared ISLs

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.

> You must follow the Basic MetroCluster ISL requirements.

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
Cisco 3132Q-V	Breakout cables	16 x 10-Gbps ISLs
Cisco 3232C	Native ports	6 ISLs using 40 or 100-Gbps ports
Cisco 3232C	Breakout cables	16 x 10-Gbps ISLs
Cisco 9336C-FX2 (not connecting NS224 shelves)	Native ports	6 ISLs using 40 or 100-Gbps
Cisco 9336C-FX2 (not connecting NS224 shelves)	Breakout cables	16 ISLs using 10-Gbps
Cisco 9336C-FX2 (connecting NS224 shelves)	Native ports (2)	4 ISLs using 40 or 100-Gbps
Cisco 9336C-FX2 (connecting NS224 shelves)	Breakout cables (2)	16 ISLs using 10-Gbps

- Using 40 or 100-Gbps ISL ports on the BES-53248 switch requires an additional license.
- When you create the RCF files for a Cisco 9336C-FX2 (connecting NS224 shelves), you must choose to configure the ISL's in native **or** breakout mode.
- 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. Using a single 10 Gbps ISL per network for AFF A800 systems is strongly discouraged.

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.

If the configuration uses L2 VLANs, they must natively span the sites. VLAN overlay such as Virtual Extensible LAN (VXLAN) is not supported.

ISLs carrying MetroCluster traffic must be native links between the switches. Link sharing services such as

Multiprotocol Label Switching (MPLS) links are not supported.

Support for WAN ISLs on the Broadcom BES-53248 switch

- Minimum number of WAN ISLs per fabric: 1 (10 GbE, or 25 GbE, or 40 GbE, or 100 GbE)
- Maximum number of 10-GbE WAN ISLs per fabric: 4
- Maximum number of 25-GbE WAN ISLs per fabric: 4
- Maximum number of 40-GbE WAN ISLs per fabric: 2
- Maximum number of 100-GbE WAN ISLs per fabric: 2

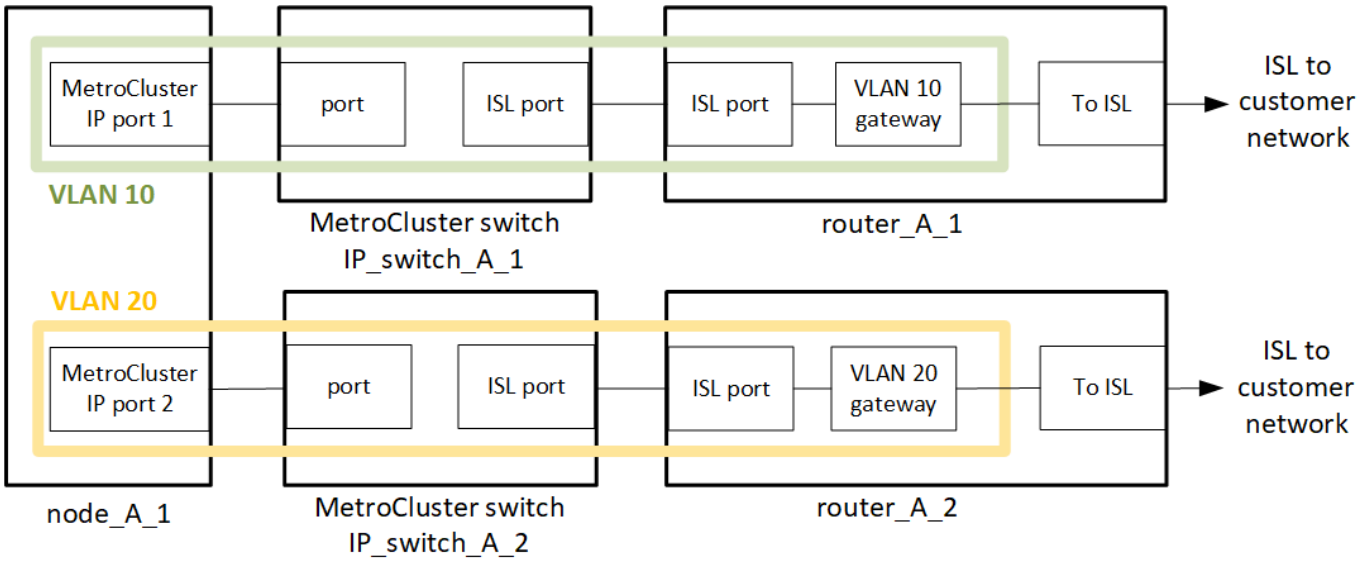
A 40-GbE or 100-GbE WAN ISL requires an RCF file version 1.40 or higher.

Extra licenses are required for additional ports.

Considerations for layer 3 wide-area networks

Beginning with ONTAP 9.9.1, MetroCluster IP configurations can be implemented with IP-routed (layer 3) backend connections.

The MetroCluster backend switches are connected to the routed IP network, either directly to routers (as shown in the following simplified example) or through other intervening switches.



NetApp supports only NetApp-validated switches. These switches are tested and sold by NetApp. They are listed in the NetApp Interoperability Matrix Tool (IMT) and in Cabling the IP switches.

The MetroCluster environment is configured and cabled as a standard MetroCluster IP configuration as described in Configure the MetroCluster hardware components. When you perform the installation and cabling procedure, you must perform the steps specific to the layer 3 configuration:

- The MetroCluster switches can be connected directly to the router or to one or more intervening switches. The VLAN must be extended to the gateway device.
- You use the `-gateway` parameter to configure the MetroCluster IP (MCC-IP) interface address with an IP gateway address.

When you configure routers and gateway IP addresses, ensure the following requirements are met:

- On each node, two interfaces cannot have the same gateway IP address.
- The corresponding interfaces on the HA pairs on each site must have the same gateway IP address.
- The corresponding interfaces on a node and its DR and AUX partners cannot have the same gateway IP address.
- The corresponding interfaces on a node and its DR and AUX partners must have the same VLAN ID.

The MetroCluster VLANs must extend from the edge MetroCluster switch to the gateway router so that MetroCluster traffic reaches the gateway (refer to the diagram shown above). The VLAN IDs for the MetroCluster VLANs must be the same at each site. However, the subnets can be different.

You use the RCF files that are created by the RcfFileGenerator tool. The network between the MetroCluster nodes and the gateway router must provide the same VLAN IDs as set in the RCF file.

IP-routed network requirements

The IP-routed network must meet the following requirements:

- Basic MetroCluster ISL requirements
- ISL requirements in shared layer 2 networks
- Required settings on intermediate switches
- Dynamic routing is not supported for the MetroCluster traffic.
- Only four-node MetroCluster configurations are supported (two nodes at each site).
- Two subnets are required on each MetroCluster site—one in each network.
- Auto-IP assignment is not supported.

Modifying address, netmask, and gateway in a MetroCluster IP

Starting from ONTAP 9.10.1, you can change the following properties of a MetroCluster IP interface: IP address and mask, and gateway. You can use any combination of parameters to update.

You might need to update these properties, for example, if a duplicate IP address is detected or if a gateway needs to change in the case of a layer 3 network due to router configuration changes.

You can only change one interface at a time. There will be traffic disruption on that interface until the other interfaces are updated and connections are reestablished.

Use the `metrocluster configuration-settings interface modify` command to change any MetroCluster IP interface property.

> These commands change the configuration on a particular node for a particular port. To restore complete network connectivity, similar commands are needed on other ports. Similarly, network switches also need to update their configuration. For example, if the gateway is updated, ideally it is changed on both nodes of an HA pair, since they are same. Plus the switch connected to those nodes also needs to update its gateway.

Use the `metrocluster configuration-settings interface show`, `metrocluster connection`

check and `metrocluster connection show` commands to verify that all connectivity is working in all interfaces.

Modify the IP address, netmask, and gateway

1. Update the IP address, netmask, and gateway for a single node and interface: `metrocluster configuration-settings interface modify`

The following command shows how to update the IP address, netmask and gateway:

```
cluster_A::* metrocluster configuration-settings interface modify -cluster
-name cluster_A -home-node node_A_1 -home-port e0a-10 -address
192.168.12.101 -gateway 192.168.12.1 -netmask 255.255.254.0
(metrocluster configuration-settings interface modify)

Warning: This operation will disconnect and reconnect iSCSI and RDMA
connections used for DR protection through port "e0a-10". Partner nodes
may need modifications for port "e0a-10" in order to completely establish
network connectivity.
Do you want to continue?" yes

[Job 28] Setting up iSCSI target configuration. (pass2:iscsi13:0:-1:0):
xpt_action_default: CCB type 0xe XPT_DEV_ADVINFO not supported
[Job 28] Establishing iSCSI initiator connections.
(pass6:iscsi14:0:-1:0): xpt_action_default: CCB type 0xe XPT_DEV_ADVINFO
not supported
(pass8:iscsi15:0:-1:0): xpt_action_default: CCB type 0xe XPT_DEV_ADVINFO
not supported
(pass9:iscsi16:0:-1:0): xpt_action_default: CCB type 0xe XPT_DEV_ADVINFO
not supported
[Job 28] Job succeeded: Interface Modify is successful.

cluster_A::*> metrocluster configuration-settings interface modify
-cluster-name cluster_A -home-node node_A_2 -home-port e0a-10 -address
192.168.12.201 -gateway 192.168.12.1 -netmask 255.255.254.0
(metrocluster configuration-settings interface modify)
Warning: This operation will disconnect and reconnect iSCSI and RDMA
connections used for DR protection through port "e0a-10". Partner nodes
may need modifications for port "e0a-10" in order to completely establish
network connectivity.
Do you want to continue?" yes
[Job 28] Job succeeded: Interface Modify is successful
```

2. Verify that all connectivity is working for all interfaces: `metrocluster configuration-settings interface show`

The following command shows how to verify that all connectivity is working for all interfaces:

```
cluster_A::*> metrocluster configuration-settings interface show
(metrocluster configuration-settings interface show)
DR              Config
Group Cluster Node    Network Address Netmask         Gateway
State
----- ------- ------- --------------- --------------- ---------------
---------
1     cluster_A node_A_2
                 Home Port: e0a-10
                      192.168.12.201  255.255.254.0   192.168.12.1
completed
                 Home Port: e0b-20
                      192.168.20.200  255.255.255.0   192.168.20.1
completed
                node_A_1
                 Home Port: e0a-10
                      192.168.12.101  255.255.254.0   192.168.12.1
completed
                 Home Port: e0b-20
                      192.168.20.101  255.255.255.0   192.168.20.1
completed
      cluster_B node_B_1
                 Home Port: e0a-10
                      192.168.11.151  255.255.255.0   192.168.11.1
completed
                 Home Port: e0b-20
                      192.168.21.150  255.255.255.0   192.168.21.1
completed
                node_B_2
                 Home Port: e0a-10
                      192.168.11.250  255.255.255.0   192.168.11.1
completed
                 Home Port: e0b-20
                      192.168.21.250  255.255.255.0   192.168.21.1
completed
8 entries were displayed.
```

3. Verify that all connections are working: `metrocluster configuration-settings connection show`

The following command shows how to verify that all connections are working:

```
cluster_A::*> metrocluster configuration-settings connection show
  (metrocluster configuration-settings connection show)
DR                    Source          Destination
Group Cluster Node    Network Address Network Address Partner Type Config
State
----- ------- ------- --------------- --------------- ------------
------------
1     cluster_A node_A_2
                  Home Port: e0a-10
                       192.168.10.200  192.168.10.101  HA Partner
completed
                  Home Port: e0a-10
                       192.168.10.200  192.168.11.250  DR Partner
completed
                  Home Port: e0a-10
                       192.168.10.200  192.168.11.151  DR Auxiliary
completed
                  Home Port: e0b-20
                       192.168.20.200  192.168.20.100  HA Partner
completed
                  Home Port: e0b-20
                       192.168.20.200  192.168.21.250  DR Partner
completed
                  Home Port: e0b-20
                       192.168.20.200  192.168.21.150  DR Auxiliary
completed
                 node_A_1
                  Home Port: e0a-10
                       192.168.10.101  192.168.10.200  HA Partner
completed
                  Home Port: e0a-10
                       192.168.10.101  192.168.11.151  DR Partner
completed
                  Home Port: e0a-10
                       192.168.10.101  192.168.11.250  DR Auxiliary
completed
                  Home Port: e0b-20
                       192.168.20.100  192.168.20.200  HA Partner
completed
                  Home Port: e0b-20
                       192.168.20.100  192.168.21.150  DR Partner
completed
                  Home Port: e0b-20
                       192.168.20.100  192.168.21.250  DR Auxiliary
completed
```

Considerations for sharing private layer 2 networks

Beginning 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.

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

ISL requirements

You must meet the requirements in:

- Basic MetroCluster ISL requirements
- ISL requirements in shared layer 2 networks

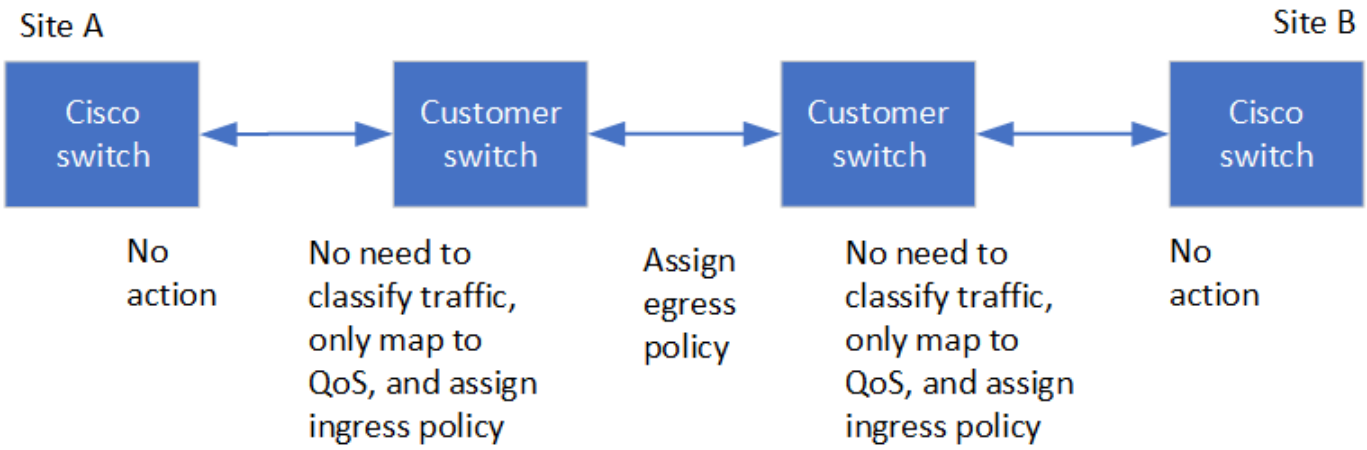
Required settings on intermediate switches

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:

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.

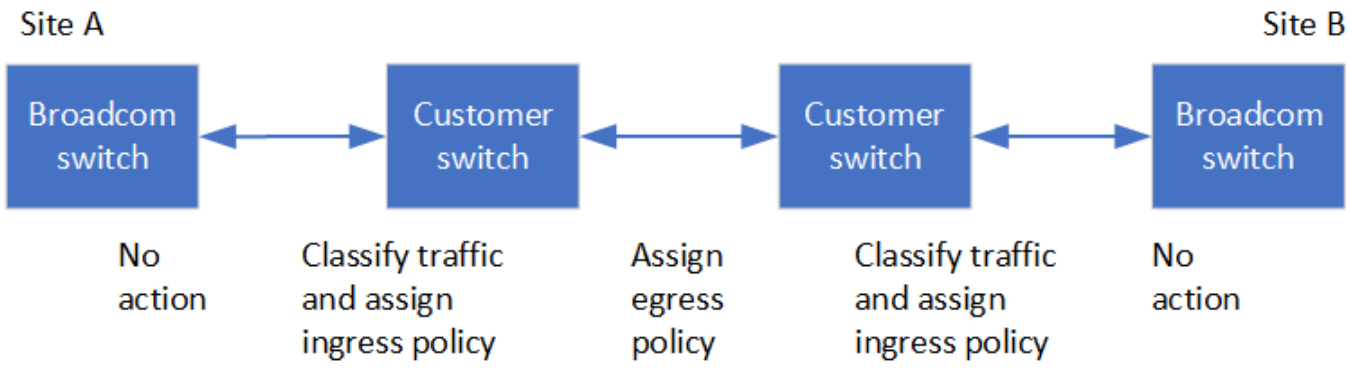
```
class-map type qos match-all rdma
   match cos 5
class-map type qos match-all storage
   match cos 4
```

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.

  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:

  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*.

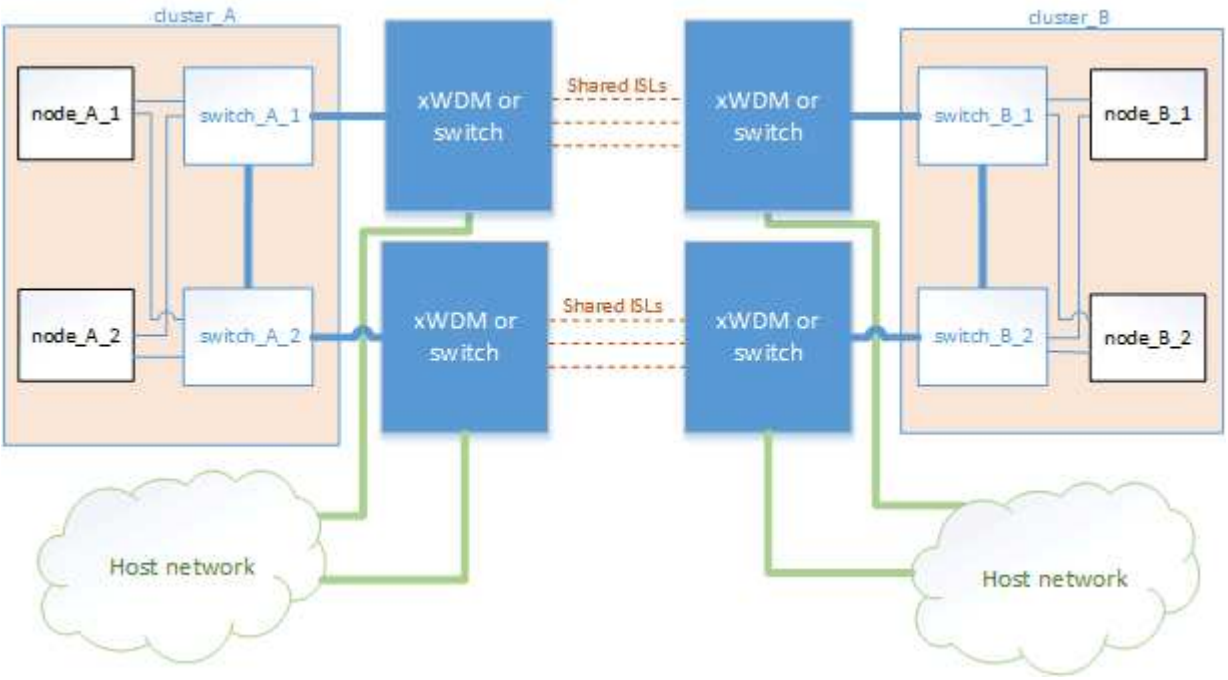
Examples of MetroCluster network topologies

Beginning 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.

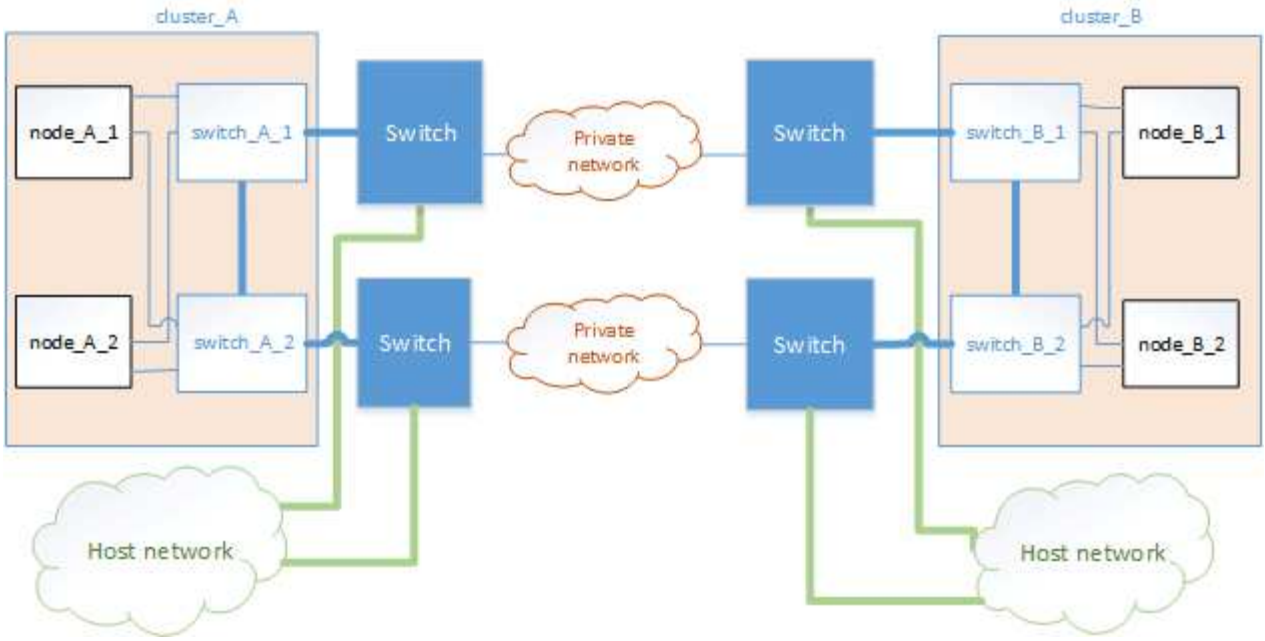
The ISL capacity must meet the minimum requirements. Depending on whether you use xWDM devices or switches a different combination of network configurations might apply.



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 NetApp. 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.

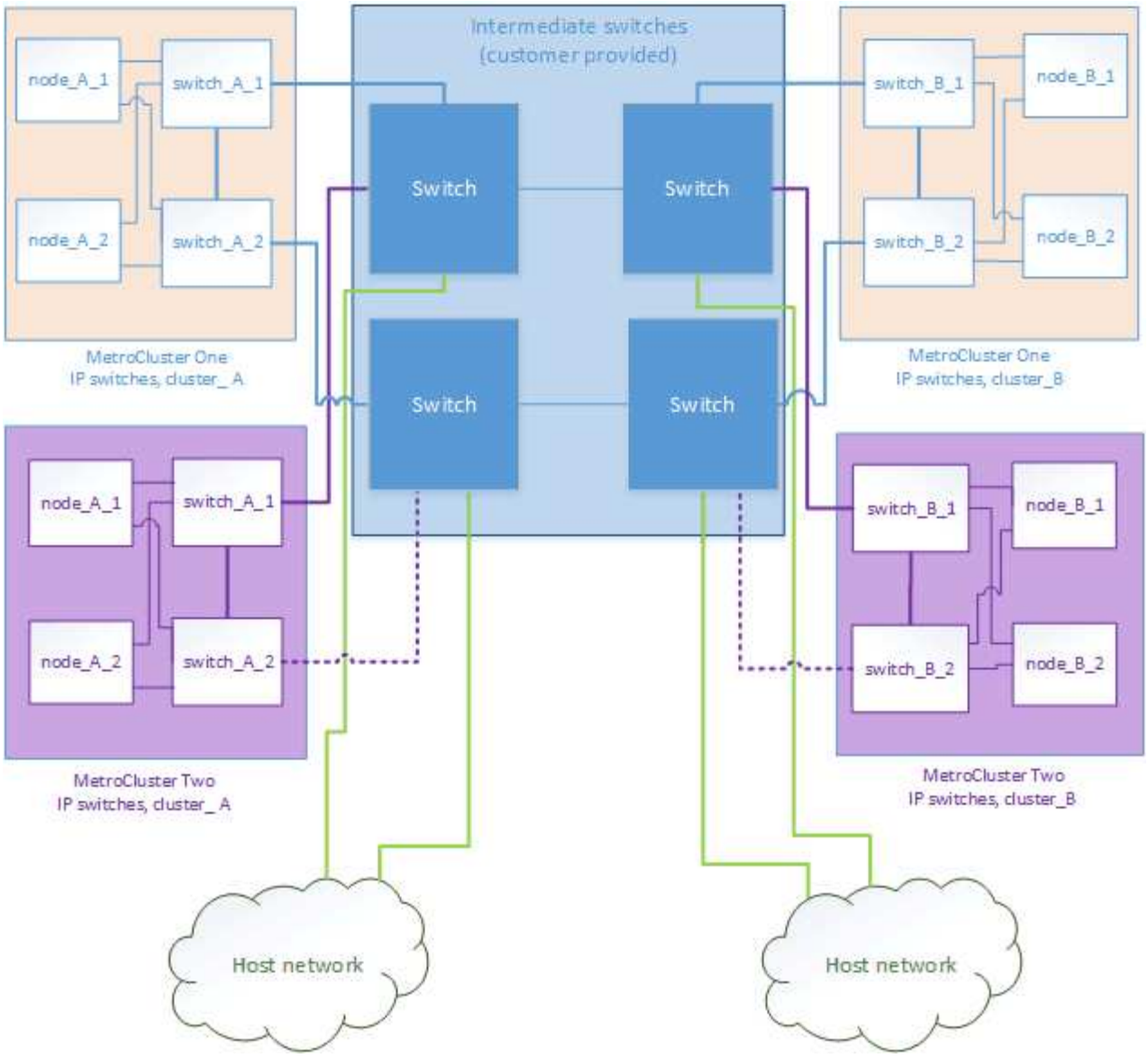
The VLAN and network xWDM or switch configuration must meet the minimum requirements.



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:



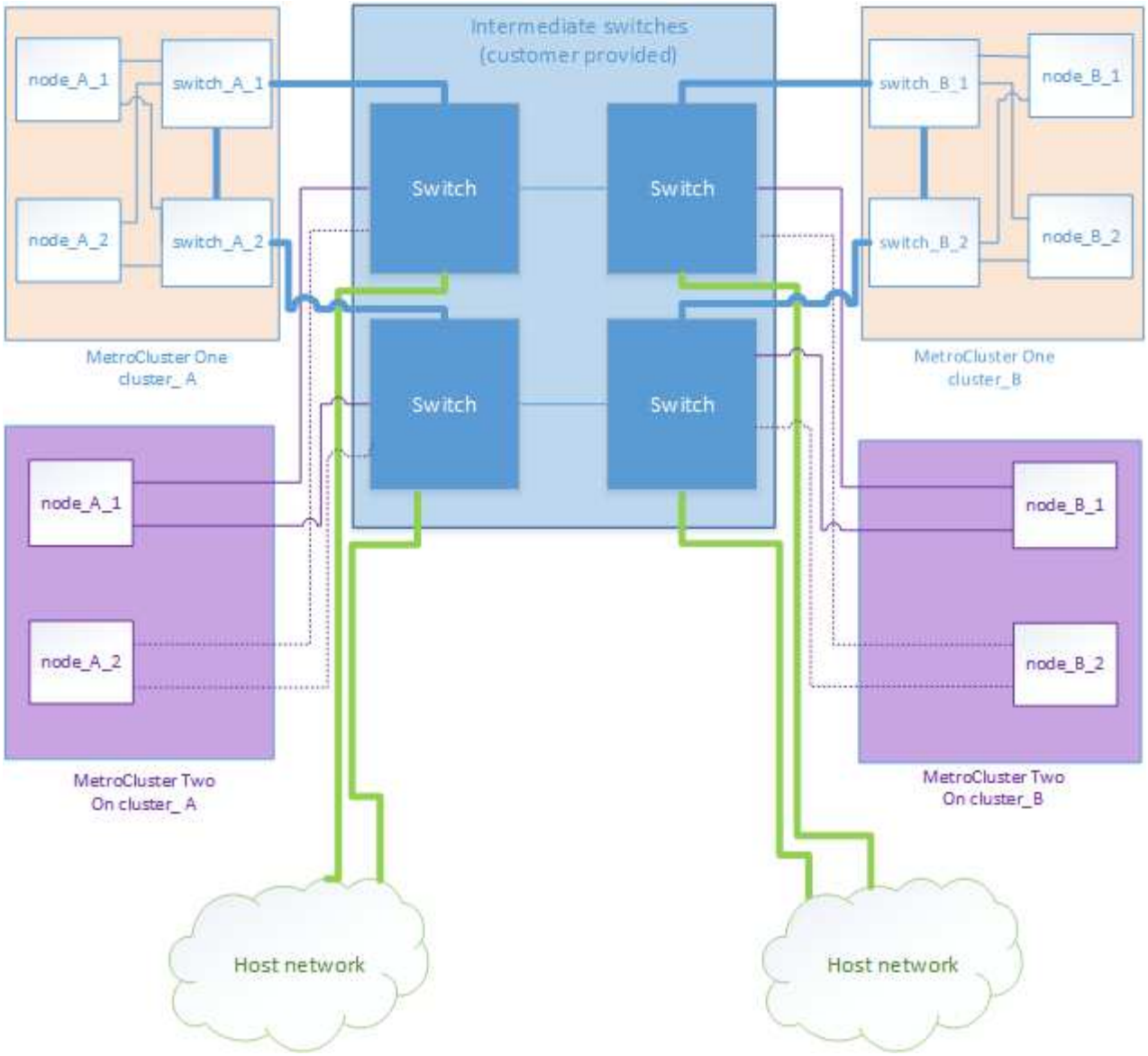
Two MetroCluster configurations with one connecting directly to the intermediate network

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

The intermediate switches must be compliant with NetApp specifications.

The example is simplified for illustration purposes only:



Considerations for using MetroCluster-compliant switches

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

General requirements

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

- The switches connecting to the MetroCluster nodes can carry non-MetroCluster traffic.
- Only platforms that provide dedicated ports for switchless cluster interconnects are supported. Platforms such as FAS2750 and AFF A220 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 FAS8200 and AFF A300 platforms, and at least 40 Gbps for all other platforms (40 Gbps or 100 Gbps).
- Four IP switches are required, two for each switch fabric.
- 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 following requirements:
  - Basic MetroCluster ISL requirements
  - ISL requirements in shared layer 2 networks
  - Required settings on intermediate switches
- In MetroCluster IP configurations using MetroCluster-compliant switches, 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 NetApp. The cables we provide must be supported by the switch vendor.

Limitations

Any configuration or feature that requires that the local cluster connections are connected to a switch is not supported. For example, the following configurations and procedures are not supported:

- Eight-node MetroCluster configurations
- Transitioning from MetroCluster FC to MetroCluster IP configurations
- Refreshing a four-node MetroCluster IP configuration

Platform-specific network speeds and switch port modes for MetroCluster-compliant switches

The following table provides platform-specific network speeds and switch port modes for MetroCluster compliant switches. You should configure the switch port mode as outlined in the table.

Missing values indicate that the platform is not supported.

Platform	Network Speed (Gbps)	Switch port mode

AFF A900	100	trunk mode
AFF A800	40 or 100	access mode
AFF A700	40	access mode
AFF A400	40 or 100	trunk mode
AFF A320	100	access mode
AFF A300	25	access mode
AFF A250	-	-
AFF A220	-	-
FAS9000	40	access mode
FAS9500	100	trunk mode
FAS8700	100	trunk mode
FAS8300	40 or 100	trunk mode
FAS8200	25	access mode
FAS2750	-	-
FAS500f	-	-

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
```

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

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 Hardware Universe tool provides some notes about the requirements that TDM/xWDM equipment must meet to work with a MetroCluster IP configuration.

NetApp Hardware Universe

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 IMT 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 NetApp 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.

The maximum supported difference in distance between fabric 1 and fabric 2 is 20 km.

Using unmirrored aggregates

If your configuration includes unmirrored aggregates, you must be aware of potential access issues after switchover operations.

Considerations for unmirrored aggregates when doing maintenance requiring power shutdown

If you are performing negotiated switchover for maintenance reasons requiring site-wide power shutdown, you should first manually take offline any unmirrored aggregates owned by the disaster site.

If you do not, nodes at the surviving site might go down due to multi-disk panics. This could occur if switched-over unmirrored aggregates go offline or are missing because of the loss of connectivity to storage at the disaster site due to the power shutdown or a loss of ISLs.

Considerations for unmirrored aggregates and hierarchical namespaces

If you are using hierarchical namespaces, you should configure the junction path so that all of the volumes in that path are either on mirrored aggregates only or on unmirrored aggregates only. Configuring a mix of unmirrored and mirrored aggregates in the junction path might prevent access to the unmirrored aggregates after the switchover operation.

Considerations for unmirrored aggregates and CRS metadata volume and data SVM root volumes

The configuration replication service (CRS) metadata volume and data SVM root volumes must be on a mirrored aggregate. You cannot move these volumes to unmirrored aggregate. If they are on unmirrored aggregate, negotiated switchover and switchback operations are vetoed. The metrocluster check command provides a warning if this is the case.

Considerations for unmirrored aggregates and SVMs

SVMs should be configured on mirrored aggregates only or on unmirrored aggregates only. Configuring a mix of unmirrored and mirrored aggregates can result in a switchover operation that exceeds 120 seconds and result in a data outage if the unmirrored aggregates do not come online.

Considerations for unmirrored aggregates and SAN

Prior to ONTAP 9.9.1, a LUN should not be located on an unmirrored aggregate. Configuring a LUN on an

unmirrored aggregate can result in a switchover operation that exceeds 120 seconds and a data outage.

Considerations for adding storage shelves for unmirrored aggregates

If you are adding shelves that will be used for unmirrored aggregates in a MetroCluster IP configuration, you must do the following:

1. Before starting the procedure to add the shelves, issue the following command:

   ```
   metrocluster modify -enable-unmirrored-aggr-deployment true
   ```

2. Verify that automatic disk assignment is off:

   ```
   disk option show
   ```

3. Follow the steps of the procedure to add the shelf.
4. Manually assign all disks from new shelf to the node that will own the unmirrored aggregate or aggregates.
5. Create the aggregates:

   ```
   storage aggregate create
   ```

6. After completing the procedure, issue the following command:

   ```
   metrocluster modify -enable-unmirrored-aggr-deployment false
   ```

7. Verify that automatic disk assignment is enabled:

   ```
   disk option show
   ```

Firewall usage at MetroCluster sites

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

Considerations for firewall usage at MetroCluster sites

If you are using a firewall at a MetroCluster site, you must ensure access for 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

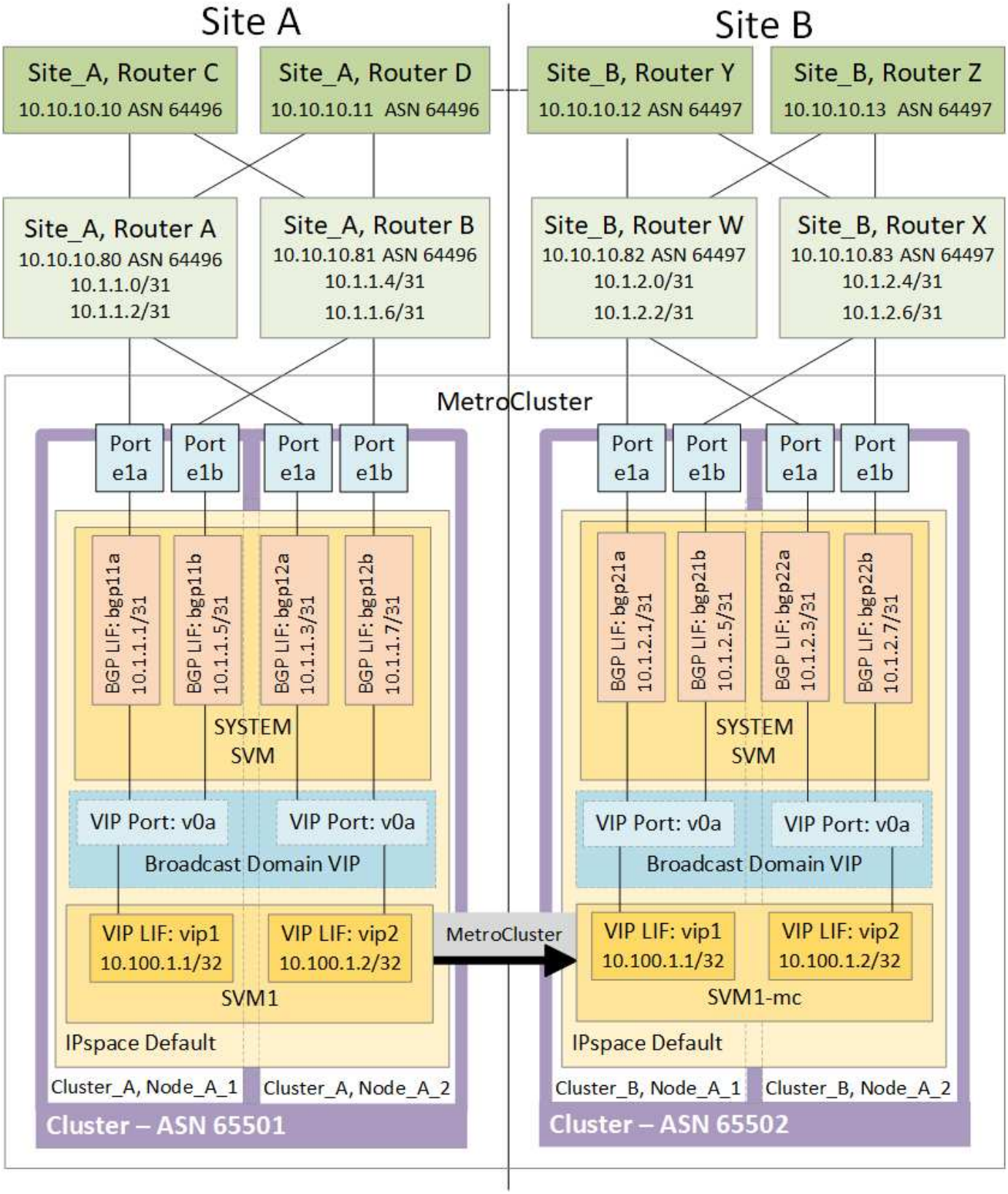
MetroCluster IP intercluster LIFs	65200 / TCP 10006 / TCP and UDP
Hardware assist	4444 / TCP

Considerations for using virtual IP and Border Gateway Protocol with a MetroCluster configuration

Beginning with ONTAP 9.5, ONTAP supports layer 3 connectivity using virtual IP (VIP) and Border Gateway Protocol (BGP). The combination VIP and BGP for redundancy in the front-end networking with the back-end MetroCluster redundancy provides a layer 3 disaster recovery solution.

Review the following guidelines and illustration when planning your layer 3 solution. For details on implementing VIP and BGP in ONTAP, refer to the following section:

Configuring virtual IP (VIP) LIFs



ONTAP limitations

ONTAP does not automatically verify that all nodes on both sites of the MetroCluster configuration are configured with BGP peering.

ONTAP does not perform route aggregation but announces all individual virtual LIF IPs as unique host routes

at all times.

ONTAP does not support true AnyCast — only a single node in the cluster presents a specific virtual LIF IP (but is accepted by all physical interfaces, regardless of whether they are BGP LIFs, provided the physical port is part of the correct IPspace). Different LIFs can migrate independently of each other to different hosting nodes.

Guidelines for using this Layer 3 solution with a MetroCluster configuration

You must configure your BGP and VIP correctly to provide the required redundancy.

Simpler deployment scenarios are preferred over more complex architectures (for example, a BGP peering router is reachable across an intermediate, non-BGP router). However, ONTAP does not enforce network design or topology restrictions.

VIP LIFs only cover the frontend/data network.

Depending on your version of ONTAP, you must configure BGP peering LIFs in the node SVM, not the system or data SVM. In 9.8, the BGP LIFs are visible in the cluster (system) SVM and the node SVMs are no longer present.

Each data SVM requires the configuration of all potential first hop gateway addresses (typically, the BGP router peering IP address), so that the return data path is available if a LIF migration or MetroCluster failover occurs.

BGP LIFs are node specific, similar to intercluster LIFs — each node has a unique configuration, which does not need to be replicated to DR site nodes.

configured, the existence of the v0a (v0b and so on.) continuously validates the connectivity, guaranteeing that a LIF migrate or failover succeeds (unlike L2, where a broken configuration is only visible after the outage).

A major architectural difference is that clients should no longer share the same IP subnet as the VIP of data SVMs. An L3 router with appropriate enterprise grade resiliency and redundancy features enabled (for example, VRRP/HSRP) should be on the path between storage and clients for the VIP to operate correctly.

The reliable update process of BGP allows for smoother LIF migrations because they are marginally faster and have a lower chance of interruption to some clients

You can configure BGP to detect some classes of network or switch misbehaviors faster than LACP, if configured accordingly.

External BGP (EBGP) uses different AS numbers between ONTAP node(s) and peering routers and is the preferred deployment to ease route aggregation and redistribution on the routers. Internal BGP (IBGP) and the use of route reflectors is not impossible but outside the scope of a straightforward VIP setup.

After deployment, you must check that the data SVM is accessible when the associated virtual LIF is migrated between all nodes on each site (including MetroCluster switchover) to verify the correct configuration of the static routes to the same data SVM.

VIP works for most IP-based protocols (NFS, SMB, iSCSI).

Configure the MetroCluster hardware components

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:

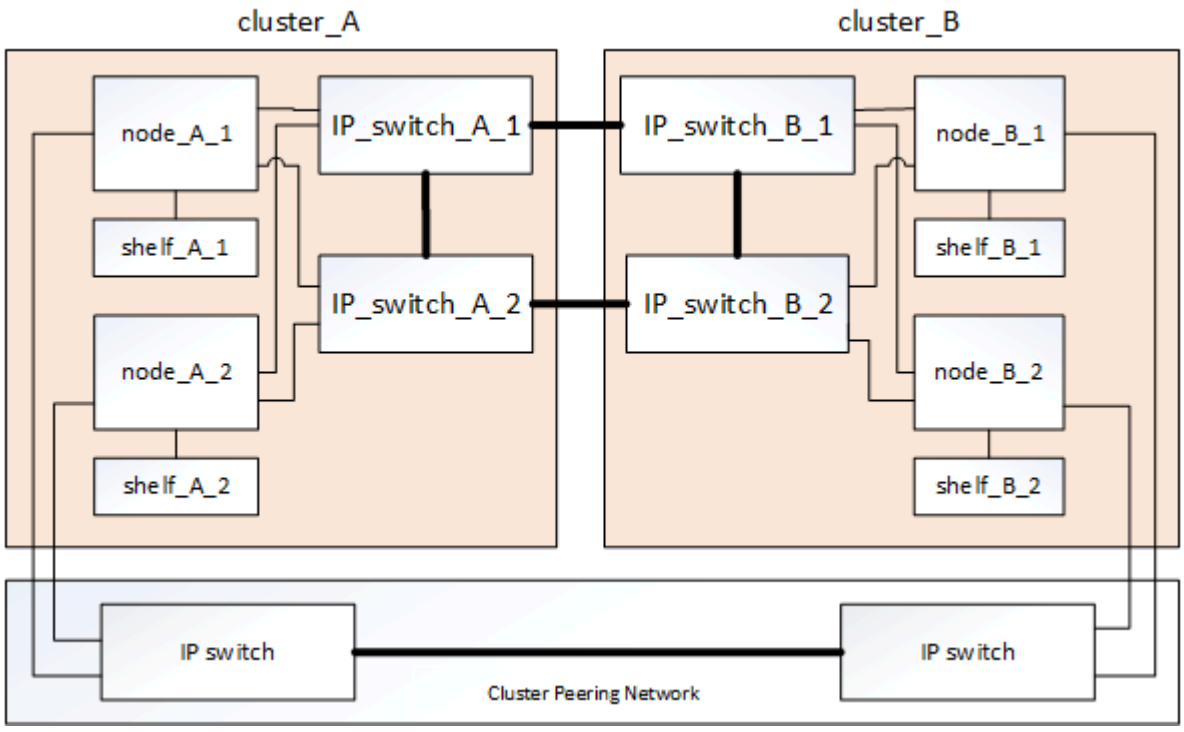
  - Standard cluster connectivity for intra-cluster communications.

    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:

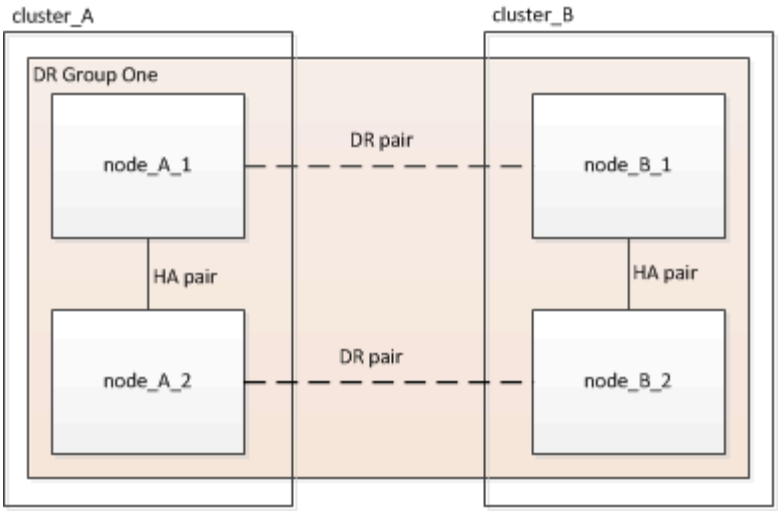
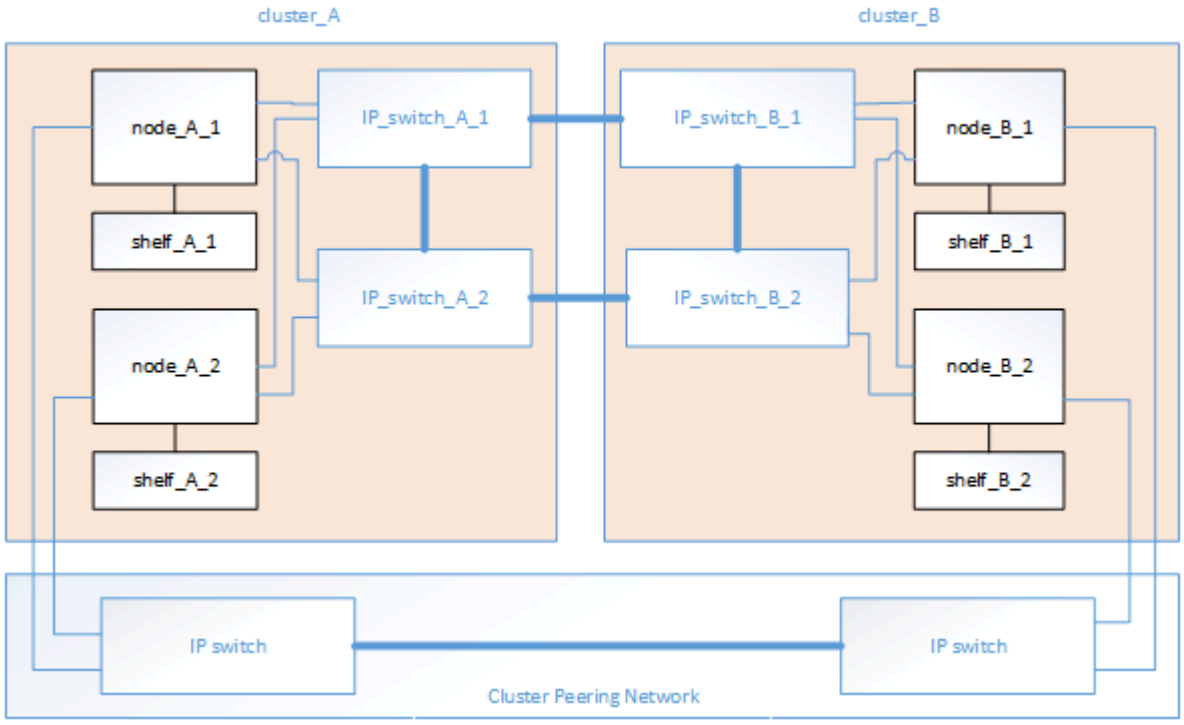


Illustration of the local HA pairs in a MetroCluster configuration

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.



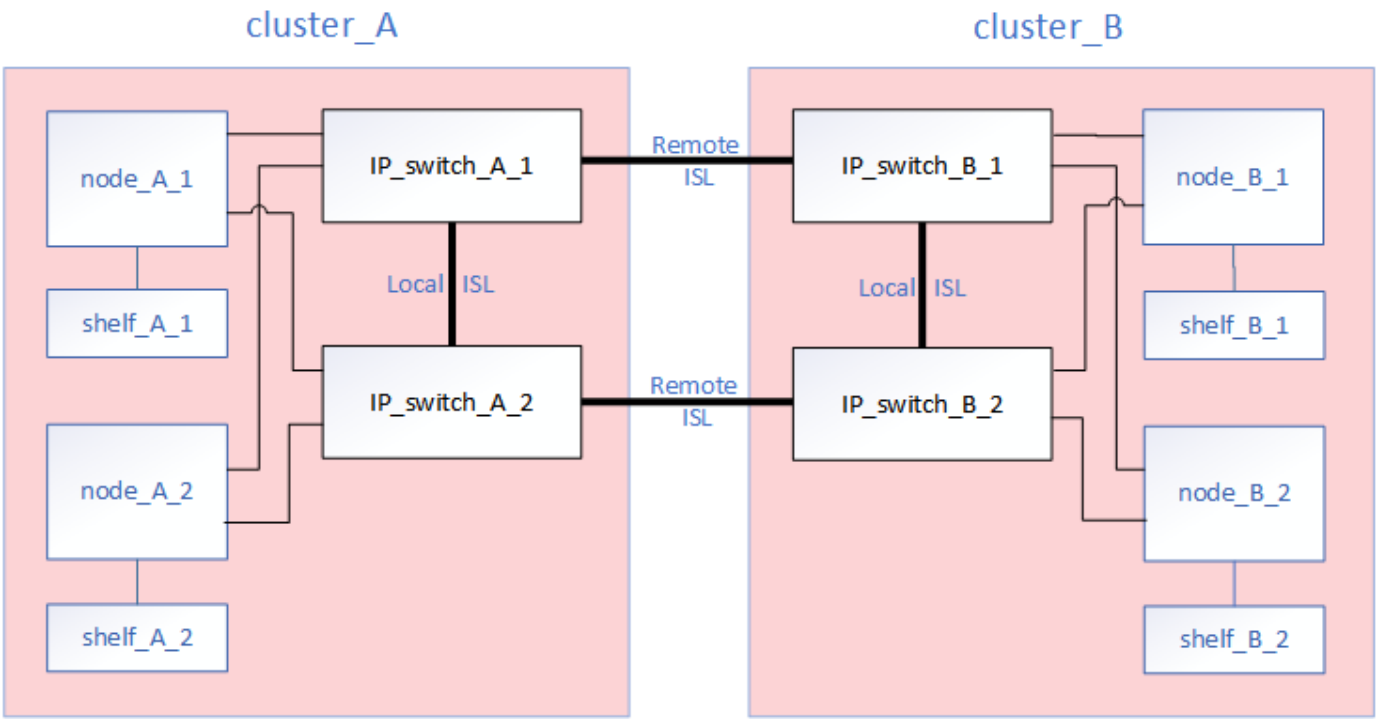
Related information

ONTAP concepts

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

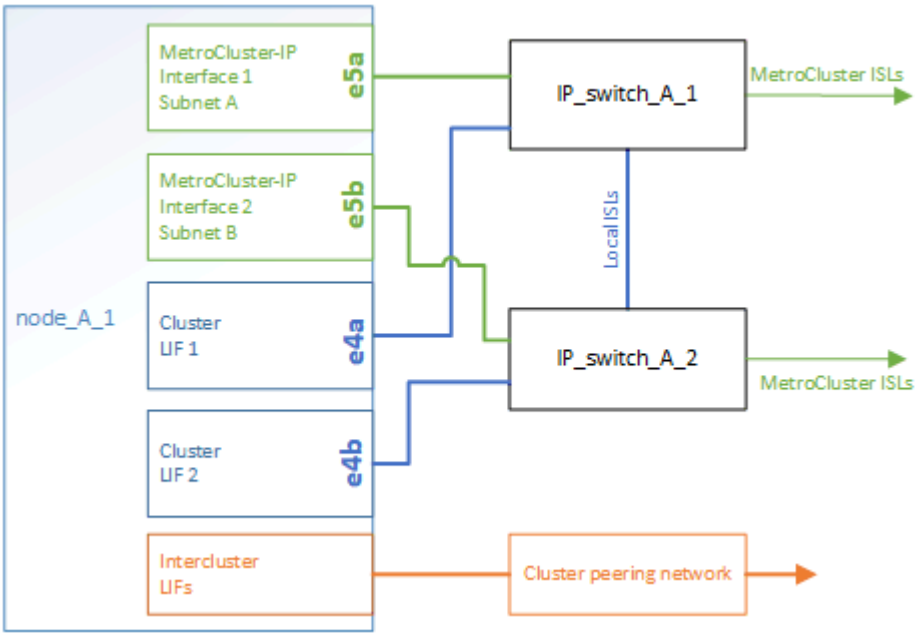
MetroCluster sites.



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

The following illustration shows these interfaces. The port usage shown is for an AFF A700 or FAS9000 system.



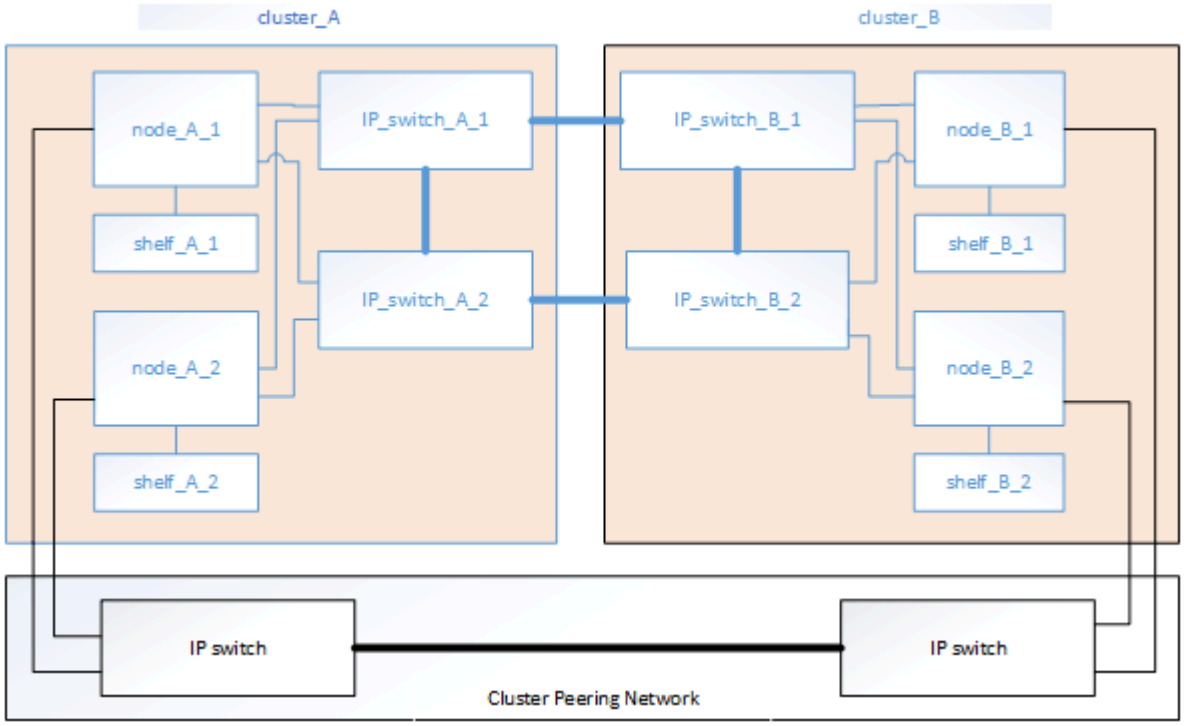
Related information

Considerations for MetroCluster IP configurations

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.



Related information

Cluster and SVM peering express configuration

Considerations for configuring cluster peering

Cabling the cluster peering connections

Peering the clusters

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.

NetApp Hardware Universe

When using AFF systems, all controller modules in the MetroCluster configuration must be configured as AFF 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 or eight 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:

- New MetroCluster IP installations on FAS systems are not supported in ONTAP 9.4.

  Existing MetroCluster IP configurations on FAS systems can be upgraded to ONTAP 9.4.

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

Controller models limited to four-node configurations

These models are limited to four in a MetroCluster configuration.

- AFF A220
- AFF A250
- FAS2750
- FAS500f

For example, the following configurations are not supported:

- An eight-node configuration consisting of eight AFF A250 controllers.
- An eight-node configuration consisting of four AFF 220 controllers and four FAS500f controllers.
- Two four-node MetroCluster IP configurations each consisting of AFF A250 controllers and sharing the same back-end switches.
- An eight-node configuration consisting of DR Group 1 with AFF A250 controllers and DR Group 2 with FAS9000 controllers.

You can configure two separate four-node MetroCluster IP configurations with the same back-end switches if the second MetroCluster does not include any of the above models.

Example names

The following example names are used in the documentation:

- 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
AFF A900 and FAS9500	X91146A	Slot 5, Slot 7	e5b, e7b
AFF A700 and FAS9000	X91146A-C	Slot 5	e5a, e5b
AFF A800	X1146A/onboard ports	Slot 1	e0b. e1b

AFF A400 and FAS8300	X1146A	Slot 1	e1a, e1b
AFF A300 and FAS8200	X1116A	Slot 1	e1a, e1b
AFF A220, and FAS2750	Onboard ports	Slot 0	e0a, e0b
AFF A250 and FAS500f	Onboard ports	Slot 0	e0c, e0d
AFF A320	Onboard ports	Slot 0	e0g, e0h

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.

A four-node MetroCluster IP configuration requires the minimum configuration at each site:

- Each node has at least one local pool and one remote pool at the site.
- At least seven drives in each pool.

  In a four-node MetroCluster configuration with a single mirrored data aggregate per node, the minimum configuration requires 24 disks at the site.

In a minimum supported configuration, each pool has the following drive layout:

- Three root drives
- Three data drives
- One spare drive

In a minimum supported configuration, at least one shelf is needed per site.

MetroCluster configurations support RAID-DP and RAID4.

Drive location considerations for partially populated shelves

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.

In a configuration with a partially populated shelf, the drives must be evenly distributed in the four quadrants of the shelf.

Drive location considerations for AFF A800 internal drives

For correct implementation of the ADP feature, the AFF A800 system disk slots must be divided into quarters and the disks must be located symmetrically in the quarters.

An AFF A800 system has 48 drive bays. The bays can be divided into quarters:

- Quarter one:

    - Bays 0 - 5
    - Bays 24 - 29
- Quarter two:
    - Bays 6 - 11
    - Bays 30 - 35
- Quarter three:
    - Bays 12 - 17
    - Bays 36 - 41
- Quarter four:
    - Bays 18 - 23
    - Bays 42 - 47

If this system is populated with 16 drives, they must be symmetrically distributed among the four quarters:

- Four drives in the first quarter: 0, 1, 2, 3
- Four drives in the second quarter: 6, 7, 8, 9
- Four drives in the third quarter: 12, 13, 14, 15
- Four drives in the fourth quarter: 18, 19, 20, 21

Mixing IOM12 and IOM 6 modules in a stack

Your version of ONTAP must support shelf mixing. Refer to the NetApp Interoperability Matrix Tool (IMT) to see if your version of ONTAP supports shelf mixing.

For further details on shelf mixing, see Hot-adding shelves with IOM12 modules to a stack of shelves with IOM6 modules

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.

Steps

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.

2. Properly ground yourself.
3. Install the controller modules in the rack or cabinet.

   AFF A220/FAS2700 Systems Installation and Setup Instructions

   AFF A250 Systems Installation and Setup Instructions

AFF A300 Systems Installation and Setup Instructions

AFF A320 systems: Installation and setup

AFF A400 Systems Installation and Setup Instructions

AFF A700 Systems Installation and Setup Instructions

AFF A800 Systems Installation and Setup Instructions

FAS500f Systems Installation and Setup Instructions

FAS8200 Systems Installation and Setup Instructions

FAS8300 and FAS8700 Systems Installation and Setup Instructions

FAS9000 Systems Installation and Setup Instructions

1. Install the IP switches in the rack or cabinet.
2. Install the disk shelves, power them on, and then set the shelf IDs.
   - You must power-cycle each disk shelf.
   - Unique shelf IDs are highly recommended for each SAS disk shelf within each MetroCluster DR group to aid troubleshooting.

> Do not cable disk shelves intended to contain unmirrored aggregates at this time. You must wait to deploy shelves intended for unmirrored aggregates until after the MetroCluster configuration is complete and only deploy them after using the `metrocluster modify -enable-unmirrored-aggr-deployment true` command.

Cable the MetroCluster IP switches

Using the port tables with the RcfFileGenerator tool or multiple MetroCluster configurations

You must understand how to use the information in the port tables to correctly generate your RCF files.

Before you begin

Review these considerations before using the tables:

- The following tables show the port usage for site A. The same cabling is used for site B.
- The switches cannot be configured with ports of different speeds (for example, a mix of 100 Gbps ports and 40 Gbps ports).
- Keep track of the MetroCluster port group (MetroCluster 1, MetroCluster 2, etc.). You will need this information when using the RcfFileGenerator tool as described later in this configuration procedure.
- The RcfFileGenerator for MetroCluster IP also provides a per-port cabling overview for each switch. Use this cabling overview to verify your cabling.

Cabling eight-node MetroCluster configurations

For MetroCluster configuration running ONTAP 9.8 and earlier, some procedures that are performed to transition an upgrade require the addition of a second four-node DR group to the configuration to create a temporary eight-node configuration. Beginning with ONTAP 9.9.1, permanent 8-node MetroCluster configurations are supported.

About this task

For such configurations, you use the same method as described above. Instead of a second MetroCluster, you are cabling an additional four-node DR group.

For example, your configuration includes the following:

- Cisco 3132Q-V switches
- MetroCluster 1: FAS2750 platforms
- MetroCluster 2: AFF A700 platforms (these platforms are being added as a second four-node DR group)

Steps

1. For MetroCluster 1, cable the Cisco 3132Q-V switches using the table for the FAS2750 platform and the rows for MetroCluster 1 interfaces.
2. For MetroCluster 2 (the second DR group), cable the Cisco 3132Q-V switches using the table for the AFF A700 platform and the rows for MetroCluster 2 interfaces.

Platform port assignments for Cisco 3132Q-V switches

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

Port usage for FAS2750 or AFF A220 systems and a Cisco 3132Q-V switch

Cabling an AFF A220 or FAS2750 to a Cisco 3132Q-V switch			
Port use	FAS2750. AFF A220		Switch Port
	IP_switch_x_1	IP_switch_x_2	
Unused	-		1
			2
			3
			4
			5
			6
ISL, Local Cluster native speed / 40G / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, Shared Cluster and MetroCluster interface	e0a	e0b	9/1
	disabled		9/2-4
	e0a	e0b	10/1
	disabled		10/2-4
MetroCluster 2, Shared Cluster and MetroCluster interface	e0a	e0b	11/1
	disabled		11/2-4
	e0a	e0b	12/1
	disabled		12/2-4
MetroCluster 3, Shared Cluster and MetroCluster interface	e0a	e0b	13/1
	disabled		13/2-4
	e0a	e0b	14/1
	disabled		14/2-4
ISL, MetroCluster native speed 40G	ISL, MetroCluster		15 - 20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25 - 32

Port usage for FAS9000, AFF A700 and a Cisco 3132Q-V switch

Cabling an AFF A700 or FAS9000 to a Cisco 3132Q-V switch			
Port use	FAS9000, AFF A700		Switch port
	IP_switch_x_1	IP_switch_x_2	Port
MetroCluster 1 Local Cluster interface	See Hardware Universe for available ports		1
			2
MetroCluster 2 Local Cluster interface			3
			4
MetroCluster 3 Local Cluster interface			5
			6
ISL, Local Cluster native speed / 40G / 100G	ISL, Local Cluster		7
			8
MetroCluster 1 MetroCluster interface	e5a	e5b	9
	e5a	e5b	10
MetroCluster 2 MetroCluster interface	e5a	e5b	11
	e5a	e5b	12
MetroCluster 3 MetroCluster interface	e5a	e5b	13
	e5a	e5b	14
ISL, MetroCluster native speed 40G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25 - 32

Port usage for AFF A800 and a Cisco 3132Q-V switch

Cabling an AFF A800 to a Cisco 3132Q-V switch			
Port use	AFF A800		Switch Port
	IP_switch_x_1	IP_switch_x_2	
MetroCluster 1 Local Cluster interface	See Hardware Universe for available ports		1
			2
MetroCluster 2 Local Cluster interface			3
			4
MetroCluster 3 Local Cluster interface			5
			6
ISL, Local Cluster native speed / 40G / 100G	ISL, Local Cluster		7
			8
MetroCluster 1 MetroCluster interface	e0b	e1b	9
	e0b	e1b	10
MetroCluster 2 MetroCluster interface	e0b	e1b	11
	e0b	e1b	12
MetroCluster 3 MetroCluster interface	e0b	e1b	13
	e0b	e1b	14
ISL, MetroCluster native speed 40G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25 - 32

Platform port assignments for Cisco 3232C or Cisco 9336C switches

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

Review these considerations before using the tables:

- The following tables show the port usage for site A. The same cabling is used for site B.
- The switches cannot be configured with ports of different speeds (for example, a mix of 100 Gbps ports and 40 Gbps ports).
- If you are configuring a single MetroCluster with the switches, use the **MetroCluster 1** port group.

  Keep track of the MetroCluster port group (MetroCluster 1, MetroCluster 2, or MetroCluster 3). You will need it when using the RcfFileGenerator tool as described later in this configuration procedure.

- The RcfFileGenerator for MetroCluster IP also provides a per-port cabling overview for each switch.

  Use this cabling overview to verify your cabling.

Cabling two MetroCluster configurations to the switches

When cabling more than one MetroCluster configuration to a Cisco 3132Q-V switch, then cable each MetroCluster according to the appropriate table. For example, if cabling a FAS2750 and an A700 to the same Cisco 3132Q-V switch. Then you cable the FAS2750 as per 'MetroCluster 1' in Table 1, and the A700 as per 'MetroCluster 2' or 'MetroCluster 3' in Table 2. You cannot physically cable both the FAS2750 and A700 as 'MetroCluster 1'.

Cabling a FAS2750 or AFF A220 system to a Cisco 3232C or Cisco 9336C switch

Cabling an AFF A220 or FAS2750 to a Cisco 3232C or Cisco 9336C switch			
Port use	FAS2750, AFF A220		Switch port
	IP_switch_x_1	IP_switch_x_2	
Unused	-		1 - 6
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, Shared Cluster and MetroCluster interface	e0a	e0b	9/1
	disabled		9/2-4
	e0a	e0b	10/1
	disabled		10/2-4
MetroCluster 2, Shared Cluster and MetroCluster interface	e0a	e0b	11/1
	disabled		11/2-4
	e0a	e0b	12/1
	disabled		12/2-4
MetroCluster 3, Shared Cluster and MetroCluster interface	e0a	e0b	13/1
	disabled		13/2-4
	e0a	e0b	14/1
	disabled		14/2-4
ISL, MetroCluster native speed 40G / 100G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25 - 32

Cabling a AFF A300 or FAS8200 to a Cisco 3232C or Cisco 9336C switch

Cabling a AFF A300 or FAS8200 to a Cisco 3232C or Cisco 9336C switch			
Port use	FAS8200, AFF A300		Switch port
	IP_switch_x_1	IP_switch_x_2	
MetroCluster 1 Local Cluster interface	See Hardware Universe for available ports		1/1
			1/2 - 4
			2/1
			2/2 - 4
MetroCluster 2 Local Cluster interface			3/1
			3/2 - 4
			4/1
			4/2 - 4
MetroCluster 3 Local Cluster interface			5/1
			5/2 - 4
			6/1
			6/2 - 4
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1 MetroCluster interface	e1a	e1b	9/1
	disabled		9/2-4
	e1a	e1b	10/1
	disabled		10/2-4
MetroCluster 2 MetroCluster interface	e1a	e1b	11/1
	disabled		11/2-4
	e1a	e1b	12/1
	disabled		12/2-4
MetroCluster 3 MetroCluster interface	e1a	e1b	13/1
	disabled		13/2-4
	e1a	e1b	14/1
	disabled		14/2-4
ISL, MetroCluster	ISL, MetroCluster		15 - 20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
MetroCluster 4 MetroCluster interface	e1a	e1b	25/1
	disabled		25/2-4
	e1a	e1b	26/1
	disabled		26/2-4
Unused	-		27 - 28
MetroCluster 4 Local Cluster interface	See Hardware Universe		29/1
	disabled		29/2-4
	See Hardware Universe		30/1
	disabled		30/2-4
Unused	-		31 - 32

Cabling a AFF A250 or FAS500f to a Cisco 3232C or Cisco 9336C switch

Cabling an AFF A250 or FAS500f to a Cisco 3232C or Cisco 9336C switch			
Port use	FAS500f, AFF A250		Switch port
	IP_switch_x_1	IP_switch_x_2	
Unused	-		1 - 6
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, Shared Cluster and MetroCluster interface	e0c	e0d	9/1
	disabled		9/2-4
	e0c	e0d	10/1
	disabled		10/2-4
MetroCluster 2, Shared Cluster and MetroCluster interface	e0c	e0d	11/1
	disabled		11/2-4
	e0c	e0d	12/1
	disabled		12/2-4
MetroCluster 3, Shared Cluster and MetroCluster interface	e0c	e0d	13/1
	disabled		13/2-4
	e0c	e0d	14/1
	disabled		14/2-4
ISL, MetroCluster native speed 40G / 100G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25 - 32

Cabling a AFF A320 to a Cisco 3232C or Cisco 9336C switch

Cabling a AFF A320 to a Cisco 3232C or Cisco 9336C switch			
Port use	AFF A320		Switch port
	IP_switch_x_1	IP_switch_x_2	
MetroCluster 1, Local Cluster interface	See Hardware Universe for available ports		1
			2
MetroCluster 2, Local Cluster interface			3
			4
MetroCluster 3, Local Cluster interface			5
			6
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, MetroCluster interface	e0g	e0h	9
	e0g	e0h	10
MetroCluster 2, MetroCluster interface	e0g	e0h	11
	e0g	e0h	12
MetroCluster 3, MetroCluster interface	e0g	e0h	13
	e0g	e0h	14
ISL, MetroCluster native speed 40G / 100G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25
			26
			27
			28
			29
			30
			31
			32

Cabling an AFF A400, FAS8300 or FAS8700 to a Cisco 3232C or Cisco 9336C switch

Cabling a AFF A400, FAS8300 or FAS8700 to a Cisco 3232C or Cisco 9336C switch			
Port use	FAS8300, FAS8700, AFF A400		Switch port
	IP_switch_x_1	IP_switch_x_2	
MetroCluster 1, Local Cluster interface	See Hardware Universe for available ports		1
			2
MetroCluster 2, Local Cluster interface			3
			4
MetroCluster 3, Local Cluster interface			5
			6
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, MetroCluster interface	e1a	e1b	9
	e1a	e1b	10
MetroCluster 2, MetroCluster interface	e1a	e1b	11
	e1a	e1b	12
MetroCluster 3, MetroCluster interface	e1a	e1b	13
	e1a	e1b	14
ISL, MetroCluster native speed 40G / 100G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25
			26
			27
			28
			29
			30
			31
			32

Cabling a AFF A700 or FAS9000 to a Cisco 3232C or Cisco 9336C switch

Cabling a AFF A700 or FAS9000 to a Cisco 3232C or Cisco 9336C switch			
Port use	FAS9000, AFF A700		Switch port
	IP_switch_x_1	IP_switch_x_2	
MetroCluster 1, Local Cluster interface	See Hardware Universe for available ports		1
			2
MetroCluster 2, Local Cluster interface			3
			4
MetroCluster 3, Local Cluster interface			5
			6
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, MetroCluster interface	e5a	e5b	9
	e5a	e5b	10
MetroCluster 2, MetroCluster interface	e5a	e5b	11
	e5a	e5b	12
MetroCluster 3, MetroCluster interface	e5a	e5b	13
	e5a	e5b	14
ISL, MetroCluster native speed 40G / 100G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25
			26
			27
			28
			29
			30
			31
			32

Cabling a AFF A800 to a Cisco 3232C or Cisco 9336C switch

Cabling an AFF A800 to a Cisco 3232C or Cisco 9336C switch			
Port use	AFF A800		Switch port
	IP_switch_x_1	IP_switch_x_2	
MetroCluster 1, Local Cluster interface	See Hardware Universe for available ports		1
			2
MetroCluster 2, Local Cluster interface			3
			4
MetroCluster 3, Local Cluster interface			5
			6
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, MetroCluster interface	e0b	e1b	9
	e0b	e1b	10
MetroCluster 2, MetroCluster interface	e0b	e1b	11
	e0b	e1b	12
MetroCluster 3, MetroCluster interface	e0b	e1b	13
	e0b	e1b	14
ISL, MetroCluster native speed 40G / 100G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused	-		25
			26
			27
			28
			29
			30
			31
			32

Cabling an AFF A900 or FAS9500 to a Cisco 3232C or Cisco 9336C switch

Cabling a FAS9500 or AFF A900 to a Cisco 3232C or Cisco 9336C-FX2 switch			
Port use	FAS9500 / A900		Switch port
	IP_switch_x_1	IP_switch_x_2	
MetroCluster 1, Local Cluster interface	See Hardware Universe for available ports		1
			2
MetroCluster 2, Local Cluster interface			3
			4
MetroCluster 3, Local Cluster interface Ports for Transition (10/40/100Gbps)			5
			6
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, MetroCluster interface	e5a	e7a	9
	e5a	e7a	10
MetroCluster 2, MetroCluster interface	e5a	e7a	11
	e5a	e7a	12
MetroCluster 3, MetroCluster interface	e5a	e7a	13
	e5a	e7a	14
ISL, MetroCluster native speed 40G / 100G	ISL, MetroCluster		15
			16
			17
			18
			19
			20
ISL, MetroCluster breakout mode 10G	ISL, MetroCluster		21/1-4
			22/1-4
			23/1-4
			24/1-4
Unused			25
			26
			27
			28
			29
			30
			31
			32
9336C-FX2 only: Ports disabled	9336C-FX2 only: Ports disabled		33
			34
			35
			36

Cabling an AFF A320, AFF A400, AFF A700 or AFF A800 to a Cisco 9336C-FX2 shared switch

Cabling an AFF A320, A400, A700, and A800 to a Cisco 9336C-FX2 shared switch			
MetroCluster 1, Local Cluster Interface	See Hardware Universe for available ports		1
			2
MetroCluster 2, Local Cluster Interface			3
			4
Storage shelf 1 (9)	NSM-A, e0a	NSM-A, e0b	5
	NSM-B, e0a	NSM-B, e0b	6
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		7
			8
MetroCluster 1, MetroCluster interface	Port 'A'	Port 'B'	9
	Port 'A'	Port 'B'	10
MetroCluster 2, MetroCluster interface	Port 'A'	Port 'B'	11
	Port 'A'	Port 'B'	12
ISL, MetroCluster, native speed 40G / 100G breakout mode 10G	ISL, MetroCluster	ISL, MetroCluster	13
			14
			15
			16
MetroCluster 1, Storage Interface	See Hardware Universe for available ports		17
			18
MetroCluster 2, Storage Interface			19
			20
Storage shelf 2 (8)	NSM-A, e0a	NSM-A, e0b	21
	NSM-B, e0a	NSM-B, e0b	22
Storage shelf 3 (7)	NSM-A, e0a	NSM-A, e0b	23
	NSM-B, e0a	NSM-B, e0b	24
Storage shelf 4 (6)	NSM-A, e0a	NSM-A, e0b	25
	NSM-B, e0a	NSM-B, e0b	26
Storage shelf 5 (5)	NSM-A, e0a	NSM-A, e0b	27
	NSM-B, e0a	NSM-B, e0b	28
Storage shelf 6 (4)	NSM-A, e0a	NSM-A, e0b	29
	NSM-B, e0a	NSM-B, e0b	30
Storage shelf 7 (3)	NSM-A, e0a	NSM-A, e0b	31
	NSM-B, e0a	NSM-B, e0b	32
Storage shelf 8 (2)	NSM-A, e0a	NSM-A, e0b	33
	NSM-B, e0a	NSM-B, e0b	34
Storage shelf 9 (1)	NSM-A, e0a	NSM-A, e0b	35
	NSM-B, e0a	NSM-B, e0b	36

MetroCluster interfaces per platform		
Platform	Port 'A'	Port 'B'
AFF A320	e0g	e0h
AFF A400	e1a	e1b
AFF A700	e5a	e5b
AFF A800	e0b	e1b

Platform port assignments for Broadcom supported BES-53248 IP switches

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

The switches cannot be used with remote ISL ports of different speeds (for example, a 25 Gbps port connected to a 10 Gbps ISL port).

Notes for the tables below:

1. For some platforms, you can use ports 49 - 54 for MetroCluster ISLs or MetroCluster interface connections.

   These ports require an additional license.

2. Only a single four-node MetroCluster using A320 systems can be connected to the switch.

   Features that require a switched cluster are not supported in this configuration, including MetroCluster FC to IP transition and tech refresh procedures.

3. AFF A320 systems configured with Broadcom BES-53248 switches might not support all features.

   Any configuration or feature that requires that the local cluster connections are connected to a switch is not supported. For example, the following configurations and procedures are not supported:

   - Eight-node MetroCluster configurations
   - Transitioning from MetroCluster FC to MetroCluster IP configurations
   - Refreshing a four-node MetroCluster IP configuration (ONTAP 9.8 and later)

4. If you connect two MetroCluster configurations and both use the same controller type, then you must use MetroCluster port groups 3 and 4. If the controllers are different, then you must use either MetroCluster port groups 3 and 4 for one type and MetroCluster port groups 1 and 2 for the other.
   - For example, if you connect:
     - Two MetroCluster configurations consisting of FAS2750/AFF A220 only, or FAS500f/AFF A250 only, you must select MetroCluster port groups 3 and 4.
     - Two MetroCluster configurations where one MetroCluster is type FAS2750/AFF A220 and the other is FAS500f/AFF A250, you must select port groups 3 and 4 for one, and port groups 1 and 2 for the other. In the RcfFileGenerator for MetroCluster IP, drop-down fields 1 and 2 only populate with the supported platform after you select platforms in drop-down fields 3 and 4. Refer to Using the port tables with the RcfFileGenerator tool or multiple MetroCluster configurations for more information on how to use the port tables.

Switch port usage for AFF A220 or FAS2750 systems

Cabling a AFF A220 or FAS2750 to a Broadcom BES-53248 switch			
Port use	FAS2750, A220		Switch port
	IP_switch_x_1	IP_switch_x_2	
Unused	-		1-6
MetroCluster 3, Shared Cluster and MetroCluster interface	e0a	e0b	9
	e0a	e0b	10
MetroCluster 4, Shared Cluster and MetroCluster interface	e0a	e0b	11
	e0a	e0b	12
ISL, MetroCLuster native speed 10G / 25G	ISL, MetroCluster		13
			14
			15
			16
Unused	-		17 - 52
ISL, MetroCluster, native speed 40G / 100G (see note 1)	ISL, MetroCluster		53
			54
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		55
			56

Switch port usage for AFF A250 or FAS500f systems

Cabling a AFF A250 or FAS500f to a Broadcom BES-53248 switch			
Port use	FAS500f, A250		Switch port
	IP_switch_x_1	IP_switch_x_2	
Unused	-		1-6
MetroCluster 3, Shared Cluster and MetroCluster interface	e0c	e0d	9
	e0c	e0d	10
MetroCluster 4, Shared Cluster and MetroCluster interface	e0c	e0d	11
	e0c	e0d	12
ISL, MetroCLuster native speed 10G / 25G	ISL, MetroCluster		13
			14
			15
			16
Unused	-		17 - 52
ISL, MetroCluster, native speed 40G / 100G (see note 1)	ISL, MetroCluster		53
			54
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		55
			56

Switch port usage for combined use of AFF A250 or FAS500f and AFF A220 or FAS2750 systems

Cabling a AFF A220 or FAS2750 and a AFF A250 or FAS500f to a Broadcom BES-53248 switch					
Port use	FAS2750, AFF A220		FAS500f, AFF A250		Switch port
	IP_switch_x_1	IP_switch_x_2	IP_switch_x_1	IP_switch_x_2	
Unused	-		-		1-4
MetroCluster 1, Shared Cluster and	e0a	e0b	e0c	e0d	5
MetroCluster interface (see note 4)	e0a	e0b	e0c	e0d	6
MetroCluster 2, Shared Cluster and	e0a	e0b	e0c	e0d	7
MetroCluster interface (see note 4)	e0a	e0b	e0c	e0d	8
MetroCluster 3, Shared Cluster and	e0a	e0b	e0c	e0d	9
MetroCluster interface (see note 4)	e0a	e0b	e0c	e0d	10
MetroCluster 4, Shared Cluster and	e0a	e0b	e0c	e0d	11
MetroCluster interface (see note 4)	e0a	e0b	e0c	e0d	12
ISL, MetroCLuster native speed 10G / 25G	ISL, MetroCluster		ISL, MetroCluster		13
					14
					15
					16
Unused	-		-		17 - 52
ISL, MetroCluster, native speed 40G / 100G (see note 1)	ISL, MetroCluster		ISL, MetroCluster		53
					54
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		ISL, Local Cluster		55
					56

Switch port usage for AFF A300 or FAS8200 systems

Cabling a AFF A300 or FAS8200 to a Broadcom BES-53248 switch			
Port use	FAS8200, AFF A300		Switch port
	IP_switch_x_1	IP_switch_x_2	
MetroCluster 1, Local Cluster interface	See Hardware Universe for available ports		1
			2
MetroCluster 2, Local Cluster interface			3
			4
MetroCluster 1, MetroCluster interface	e1a	e1b	5
	e1a	e1b	6
MetroCluster 2, MetroCluster interface	e1a	e1b	7
	e1a	e1b	8
Unused	-		9
			10
			11
			12
ISL, MetroCLuster native speed 10G / 25G	ISL, MetroCluster		13
			14
			15
			16
Unused	-		17 - 52
ISL, MetroCluster, native speed 40G / 100G (see note 1)	ISL, MetroCluster		53
			54
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		55
			56

Cabling a AFF A320 to a Broadcom BES-53248 switch			
Port use	AFF A320		Switch port
	IP_switch_x_1	IP_switch_x_2	
Ports not used	Ports not used		1 - 12
ISL, MetroCLuster native speed 10G / 25G	ISL, MetroCluster		13
			14
			15
			16
Ports not licensed (17 - 52)			..
ISL, MetroCluster, native speed 40G / 100G (see note 1)	ISL, MetroCluster		53
			54
MetroCluster 1, MetroCluster interface (see note 2)	e0g	e0h	55
	e0g	e0h	56

Switch port usage for AFF A400, FAS8300 or FAS8700 systems

Cabling a FAS8300, A400 or FAS8700 to a Broadcom BES-53248 switch			
Port use	FAS8300,FAS8700, A400		Switch port
	IP_switch_x_1	IP_switch_x_2	
Unused	-		1 - 12
ISL, MetroCluster native speed 10G / 25G	ISL, MetroCluster		13
			14
			15
			16
Unused	-		17 - 48
MetroCluster 5, Local Cluster interface (see note 1)	See Hardware Universe for available ports		49
			50
MetroCluster 5, MetroCluster interface (see note 1)	e1a	e1b	51
	e1a	e1b	52
ISL, MetroCluster, native speed 40G / 100G (see note 1)	ISL, MetroCluster		53
			54
ISL, Local Cluster native speed / 100G	ISL, Local Cluster		55
			56

Cabling the controller peering, data, and management ports

You must cable the controller module ports used for cluster peering, management and data connectivity.

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.

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

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

   Use the installation instructions for your platform at the AFF and FAS System Documentation.

Configure the MetroCluster IP switches

Configuring Broadcom IP switches

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

Your configuration requires additional licenses (6 x 100-Gb port license) in the following scenarios:

- You use ports 53 and 54 as a 40-Gbps or 100-Gbps MetroCluster ISL.
- You use a platform that connects the local cluster and MetroCluster interfaces to ports 49 - 52.

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.

Steps

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

```
(IP_switch_A_1)> enable
(IP_switch_A_1) #
```

2. Erase the startup configuration and remove the banner
   a. Erase the startup configuration:

      `erase startup-config`

```
(IP_switch_A_1) #erase startup-config

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

(IP_switch_A_1) #
```

      This command does not erase the banner.

   b. Remove the banner:

      `no set clibanner`

```
(IP_switch_A_1) #configure
(IP_switch_A_1)(Config) # no set clibanner
(IP_switch_A_1)(Config) #
```

3. Reboot the switch: **`(IP_switch_A_1) #reload`**

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

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

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:

```
(Routing)>
```

5. Change to the elevated command prompt:

   `enable`

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

6. Set the service port protocol to `none`:

   `serviceport protocol none`

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

(Routing) #
```

7. Assign the IP address to the service port:

   `serviceport ip` *`ip-address netmask gateway`*

   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
```

8. Verify that the service port is correctly configured:

   `show serviceport`

   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) #
```

9. If desired, configure the SSH server.

   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 (optional)

   ```
   (Routing) #configure
   (Routing) (Config)#crypto key generate dsa
   ```

   c. If you are using the FIPS compliant version of EFOS, generate the ECDSA keys. The following example creates the keys with a length of 256. Valid values are 256, 384 or 521.

   ```
   (Routing) #configure
   (Routing) (Config)#crypto key generate ecdsa 256
   ```

   d. Enable the SSH server.

   If necessary, exit the configuration context.

```
(Routing) (Config)#end
(Routing) #ip ssh server enable
```

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

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

    `configure`

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

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

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

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

    ```
    (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)#
    ```

12. Configure the switch name:

    `hostname IP_switch_A_1`

    The switch prompt will display the new name:

    ```
    (Routing) # hostname IP_switch_A_1

    (IP_switch_A_1) #
    ```

13. Save the configuration:

    `write memory`

    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) #
```

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.

Note the following:

- When upgrading from EFOS 3.4.x.x to EFOS 3.7.x.x or later, the switch must be running EFOS 3.4.4.6 (or later 3.4.x.x release). If you are running a release prior to that, then upgrade the switch to EFOS 3.4.4.6 (or later 3.4.x.x release) first, then upgrade the switch to EFOS 3.7.x.x or later.
- The configuration for EFOS 3.4.x.x and 3.7.x.x or later are different. Changing the EFOS version from 3.4.x.x to 3.7.x.x or later, or vice versa, requires the switch to be reset to factory defaults and the RCF files for the corresponding EFOS version to be (re)applied. This procedure requires access through the serial console port.
- Beginning with EFOS version 3.7.x.x or later, a non-FIPS compliant and a FIPS compliant version is available. Different steps apply when moving to from a non-FIPS compliant to a FIPS compliant version or vice versa. Changing EFOS from a non-FIPS compliant to a FIPS compliant version or vice versa will reset the switch to factory defaults. This procedure requires access through the serial console port.

Steps

1. Check if your version of EFOS is FIPS compliant or non-FIPS compliant by using the `show fips status` command. In the following examples, `IP_switch_A_1` is using FIPS compliant EFOS and `IP_switch_A_2` is using non-FIPS compliant EFOS.

   Example 1

```
IP_switch_A_1 #show fips status

System running in FIPS mode

IP_switch_A_1 #
```

Example 2

```
IP_switch_A_2 #show fips status
                      ^
% Invalid input detected at `^` marker.

IP_switch_A_2 #
```

2. Use the following table to determine which method you must follow:

Procedure	Current EFOS version	New EFOS version	High level steps
Steps to upgrade EFOS between two (non) FIPS compliant versions	3.4.x.x	3.4.x.x	Install the new EFOS image using method 1) The configuration and license information is retained
	3.4.4.6 (or later 3.4.x.x)	3.7.x.x or later non-FIPS compliant	Upgrade EFOS using method 1. Reset the switch to factory defaults and apply the RCF file for EFOS 3.7.x.x or later
	3.7.x.x or later non-FIPS compliant	3.4.4.6 (or later 3.4.x.x)	Downgrade EFOS using method 1. Reset the switch to factory defaults and apply the RCF file for EFOS 3.4.x.x
		3.7.x.x or later non-FIPS compliant	Install the new EFOS image using method 1. The configuration and license information is retained
	3.7.x.x or later FIPS compliant	3.7.x.x or later FIPS compliant	Install the new EFOS image using method 1. The configuration and license information is retained

Steps to upgrade to/from a FIPS compliant EFOS version	Non-FIPS compliant	FIPS compliant	Installation of the EFOS image using method 2. The switch configuration and license information will be lost.
	FIPS compliant	Non-FIPS compliant	

- Method 1: Steps to upgrade EFOS with downloading the software image to the backup boot partition
- Method 2: Steps to upgrade EFOS using the ONIE OS installation

Steps to upgrade EFOS with downloading the software image to the backup boot partition

You can perform the following steps only if both EFOS versions are non-FIPS compliant or both EFOS versions are FIPS compliant.

> Do not use these steps if one version is FIPS compliant and the other version is non-FIPS compliant.

Steps

1. Copy the switch software to the switch: `copy sftp://user@50.50.50.50/switchsoftware/efos-3.4.4.6.stk backup`

   In this example, the efos-3.4.4.6.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-
3.4.4.6.stk backup
Remote Password:*************

Mode..................................... SFTP
Set Server IP.................................. 50.50.50.50
Path..................................... /switchsoftware/
Filename..................................... efos-3.4.4.6.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) #
```

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

   `boot system backup`

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

(IP_switch_A_1) #
```

3. Verify that the new boot image will be active on the next boot:

   `show bootvar`

```
(IP_switch_A_1) #show bootvar

Image Descriptions

 active :
 backup :


 Images currently available on Flash

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

    1       3.4.4.2    3.4.4.6      3.4.4.2        3.4.4.6

(IP_switch_A_1) #
```

4. Save the configuration:

   `write memory`

```
(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) #
```

5. Reboot the switch:

   `reload`

```
(IP_switch_A_1) #reload

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

6. Wait for the switch to reboot.

In rare scenarios the switch may fail to boot. Follow the Steps to upgrade EFOS using the ONIE OS installation to install the new image.

7. If you change the switch from EFOS 3.4.x.x to EFOS 3.7.x.x or vice versa then follow the following two procedures to apply the correct configuration (RCF):
   a. Resetting the Broadcom IP switch to factory defaults
   b. Downloading and installing the Broadcom RCF files
8. Repeat these steps on the remaining three IP switches in the MetroCluster IP configuration.

Steps to upgrade EFOS using the ONIE OS installation

You can perform the following steps if one EFOS version is FIPS compliant and the other EFOS version is non-FIPS compliant. These steps can be used to install the non-FIPS or FIPS compliant EFOS 3.7.x.x image from ONIE if the switch fails to boot.

Steps

1. Boot the switch into ONIE installation mode.

   During boot, select ONIE when the following screen appears:

```
+----------------------------------------------------------------------+
|EFOS                                                                  |
|*ONIE                                                                 |
|                                                                      |
|                                                                      |
|                                                                      |
|                                                                      |
|                                                                      |
|                                                                      |
|                                                                      |
|                                                                      |
|                                                                      |
|                                                                      |
+----------------------------------------------------------------------+
```

   After selecting "ONIE", the switch will then load and present you with the following choices:

```
 +------------------------------------------------------------------------+
 |*ONIE: Install OS                                                       |
 | ONIE: Rescue                                                           |
 | ONIE: Uninstall OS                                                     |
 | ONIE: Update ONIE                                                      |
 | ONIE: Embed ONIE                                                       |
 | DIAG: Diagnostic Mode                                                  |
 | DIAG: Burn-In Mode                                                     |
 |                                                                        |
 |                                                                        |
 |                                                                        |
 |                                                                        |
 |                                                                        |
 +------------------------------------------------------------------------+
```

The switch now will boot into ONIE installation mode.

2. Stop the ONIE discovery and configure the ethernet interface

   Once the following message appears press <enter> to invoke the ONIE console:

```
 Please press Enter to activate this console. Info: eth0:  Checking
link... up.
 ONIE:/ #
```

   The ONIE discovery will continue and messages will be printed to the console.

```
Stop the ONIE discovery
ONIE:/ # onie-discovery-stop
discover: installer mode detected.
Stopping: discover... done.
ONIE:/ #
```

3. Configure the ethernet interface and add the route using `ifconfig eth0 <ipAddress> netmask <netmask> up` and `route add default gw <gatewayAddress>`

```
ONIE:/ # ifconfig eth0 10.10.10.10 netmask 255.255.255.0 up
ONIE:/ # route add default gw 10.10.10.1
```

4. Verify that the server hosting the ONIE installation file is reachable:

```
ONIE:/ # ping 50.50.50.50
PING 50.50.50.50 (50.50.50.50): 56 data bytes
64 bytes from 50.50.50.50: seq=0 ttl=255 time=0.429 ms
64 bytes from 50.50.50.50: seq=1 ttl=255 time=0.595 ms
64 bytes from 50.50.50.50: seq=2 ttl=255 time=0.369 ms
^C
--- 50.50.50.50 ping statistics ---
3 packets transmitted, 3 packets received, 0% packet loss
round-trip min/avg/max = 0.369/0.464/0.595 ms
ONIE:/ #
```

5. Install the new switch software

```
ONIE:/ # onie-nos-install http:// 50.50.50.50/Software/onie-installer-
x86_64
discover: installer mode detected.
Stopping: discover... done.
Info: Fetching http:// 50.50.50.50/Software/onie-installer-3.7.0.4 ...
Connecting to 50.50.50.50 (50.50.50.50:80)
installer            100% |*******************************| 48841k
0:00:00 ETA
ONIE: Executing installer: http:// 50.50.50.50/Software/onie-installer-
3.7.0.4
Verifying image checksum ... OK.
Preparing image archive ... OK.
```

The software will install and then reboot the switch. Let the switch reboot normally into the new EFOS version.

6. Verify that the new switch software is installed

`show bootvar`

```
(Routing) #show bootvar
Image Descriptions
active :
backup :
Images currently available on Flash
----    ----------- -------- --------------- ------------
unit    active     backup   current-active  next-active
----    ----------- -------- --------------- ------------
1   3.7.0.4     3.7.0.4  3.7.0.4         3.7.0.4
(Routing) #
```

7. Complete the installation

   The switch will reboot with no configuration applied and reset to factory defaults. Follow the two procedures to configure the switch basic settings and apply the RCF file as outlined in the following two documents:

   a. Configure the switch basic settings. Follow step 4 and later: Resetting the Broadcom IP switch to factory defaults
   b. Create and apply the RCF file as outlined in Downloading and installing the Broadcom RCF files

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	v1.32_Switch-A1.txt
IP_switch_A_2	v1.32_Switch-A2.txt
IP_switch_B_1	v1.32_Switch-B1.txt
IP_switch_B_2	v1.32_Switch-B2.txt

The RCF files for EFOS version 3.4.4.6 or later 3.4.x.x. release and EFOS version 3.7.0.4 are different. You need to make sure that you have created the correct RCF files for the EFOS version that the switch is running.

EFOS version	RCF file version
3.4.x.x	v1.3x, v1.4x
3.7.x.x	v2.x

Steps

1. Generate the Broadcom RCF files for MetroCluster IP.
   a. Download the RcfFileGenerator for MetroCluster IP
   b. Generate the RCF file for your configuration using the RcfFileGenerator for MetroCluster IP.

      Modifications to the RCF files after download are not supported.

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/BES-53248_v1.32_Switch-A1.txt nvram:script BES-53248_v1.32_Switch-A1.scr`

   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
```

```
(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
```

```
(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
"*****************************************************************
************

* NetApp 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
```

```
(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`

```
(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.

3. Reload the switch:

`reload`

```
IP_switch_A_1# reload
```

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

Configure Cisco IP switches

Configuring Cisco IP switches

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

About this task

Several of the procedures in this section are independent procedures and you only need to execute those you are directed to or are relevant to your task.

Resetting the Cisco IP switch to factory defaults

Before installing any RCF file, you must erase the Cisco switch configuration and perform basic configuration. This procedure is required when you want to reinstall the same RCF

file after a previous installation failed, or if you want to install a new version of an RCF file.

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.

Steps

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.

      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#
```

2. Save the configuration:

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

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

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

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.

About this task

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

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

You must use the supported switch software version.

NetApp Hardware Universe

Steps

1. Download the supported NX-OS software file.

   Cisco Software Download

2. Copy the switch software to the switch:

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

   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)...
   ```

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

   ```
   dir bootflash:
   ```

   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#
```

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.

   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#
```

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.
```

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#
```

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.

About this task

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

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

You must use the supported switch software version.

NetApp Hardware Universe

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_v1.80_Switch-A1.txt
IP_switch_A_2	NX3232_v1.80_Switch-A2.txt
IP_switch_B_1	NX3232_v1.80_Switch-B1.txt
IP_switch_B_2	NX3232_v1.80_Switch-B2.txt

Steps

1. Download the MetroCluster IP RCF files.

   Modifications to the RCF files after download are not supported.

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
      ```

      In this example, the NX3232_v1.80_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_v1.80_Switch-A1.txt
bootflash: vrf management
root@10.10.99.99's password: password
sftp> progress
Progress meter enabled
sftp> get   /tftpboot/NX3232_v1.80_Switch-A1.txt
/bootflash/NX3232_v1.80_Switch-A1.txt
Fetching /tftpboot/NX3232_v1.80_Switch-A1.txt to
/bootflash/NX3232_v1.80_Switch-A1.txt
/tftpboot/NX3232_v1.80_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.

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

   `dir bootflash:`

   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_v1.80_Switch-A1.txt
                  .
                  .
                  .

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

4. Configure the TCAM regions on Cisco 3132Q-V and Cisco 3232C switches.

   Skip this step if you do not have Cisco 3132Q-V or Cisco 3232C switches.

   a. On Cisco 3132Q-V switch, set the following TCAM regions:

```
conf t
hardware access-list tcam region span 0
hardware access-list tcam region racl 256
hardware access-list tcam region e-racl 256
hardware access-list tcam region qos 256
```

b. On Cisco 3232C switch, set the following TCAM regions:

```
conf t
hardware access-list tcam region span 0
hardware access-list tcam region racl-lite 0
hardware access-list tcam region racl 256
hardware access-list tcam region e-racl 256
hardware access-list tcam region qos 256
```

c. After setting the TCAM regions, save the configuration and reload the switch:

```
copy running-config startup-config
reload
```

5. Copy the matching RCF file from the local bootflash to the running configuration on each switch:

   `copy bootflash:switch-specific-RCF.txt running-config`

6. 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_v1.80_Switch-A1.txt running-config
IP_switch-A-1# copy running-config startup-config
```

7. Reload the switch:

   `reload`

```
IP_switch_A_1# reload
```

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

Setting Forward Error Correction for systems using 25-Gbps connectivity

If your 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.

About this task

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

Platform port assignments for Cisco 3232C or Cisco 9336C switches

This task only applies to platforms using 25-Gbps connectivity:

- AFF A300
- FAS 8200
- FAS 500f
- AFF A250

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

Steps

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`

   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
```

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

Configure MACsec encryption on Cisco 9336C switches

If desired, you can configure MACsec encryption on the WAN ISL ports that run between the sites. You must configure MACsec after applying the correct RCF file.

MACsec encryption can only be applied to the WAN ISL ports.

Configure MACsec encryption on Cisco 9336C switches

You must only configure MACsec encryption on the WAN ISL ports that run between the sites. You must configure MACsec after applying the correct RCF file.

Licensing requirements for MACsec

MACsec requires a security license. For a complete explanation of the Cisco NX-OS licensing scheme and how to obtain and apply for licenses, see the Cisco NX-OS Licensing Guide

Enable Cisco MACsec Encryption WAN ISLs in MetroCluster IP configurations

You can enable MACsec encryption for Cisco 9336C switches on the WAN ISLs in a MetroCluster IP configuration.

Steps

1. Enter global configuration mode:

   `configure terminal`

   ```
   IP_switch_A_1# configure terminal
   IP_switch_A_1(config)#
   ```

2. Enable MACsec and MKA on the device:

   `feature macsec`

   ```
   IP_switch_A_1(config)# feature macsec
   ```

3. Copy the running configuration to the startup configuration:

   `copy running-config startup-config`

   ```
   IP_switch_A_1(config)# copy running-config startup-config
   ```

Configure a MACsec key chain and keys

You can create a MACsec key chain or keys on your configuration.

Key Lifetime and Hitless Key Rollover

A MACsec keychain can have multiple pre-shared keys (PSKs), each configured with a key ID and an optional lifetime. A key lifetime specifies at which time the key activates and expires. In the absence of a lifetime configuration, the default lifetime is unlimited. When a lifetime is configured, MKA rolls over to the next configured pre-shared key in the keychain after the lifetime is expired. The time zone of the key can be local or UTC. The default time zone is UTC. A key can roll over to a second key within the same keychain if you configure the second key (in the keychain) and configure a lifetime for the first key. When the lifetime of the first key expires, it automatically rolls over to the next key in the list. If the same key is configured on both sides of

the link at the same time, then the key rollover is hitless (that is, the key rolls over without traffic interruption).

Steps

1. Enter the global configuration mode:

   `configure terminal`

   ```
   IP_switch_A_1# configure terminal
   IP_switch_A_1(config)#
   ```

2. To hide the encrypted key octet string, replace the string with a wildcard character in the output of the `show running-config` and `show startup-config` commands:

   ```
   IP_switch_A_1(config)# key-chain macsec-psk no-show
   ```

   The octet string is also hidden when you save the configuration to a file.

   By default, PSK keys are displayed in encrypted format and can easily be decrypted. This command applies only to MACsec key chains.

3. Create a MACsec key chain to hold a set of MACsec keys and enter MACsec key chain configuration mode:

   `key chain name macsec`

   ```
   IP_switch_A_1(config)# key chain 1 macsec
   IP_switch_A_1(config-macseckeychain)#
   ```

4. Create a MACsec key and enter MACsec key configuration mode:

   `key key-id`

   The range is from 1 to 32 hex digit key-string, and the maximum size is 64 characters.

   ```
   IP_switch_A_1 switch(config-macseckeychain)# key 1000
   IP_switch_A_1 (config-macseckeychain-macseckey)#
   ```

5. Configure the octet string for the key:

   `key-octet-string octet-string cryptographic-algorithm AES_128_CMAC | AES_256_CMAC`

```
IP_switch_A_1(config-macseckeychain-macseckey)# key-octet-string
abcdef0123456789abcdef0123456789abcdef0123456789abcdef0123456789
cryptographic-algorithm AES_256_CMAC
```

The octet-string argument can contain up to 64 hexadecimal characters. The octet key is encoded internally, so the key in clear text does not appear in the output of the `show running-config macsec` command.

6. Configure a send lifetime for the key (in seconds):

   `send-lifetime start-time duration duration`

   ```
   IP_switch_A_1(config-macseckeychain-macseckey)# send-lifetime 00:00:00
   Oct 04 2020 duration 100000
   ```

   By default, the device treats the start time as UTC. The start-time argument is the time of day and date that the key becomes active. The duration argument is the length of the lifetime in seconds. The maximum length is 2147483646 seconds (approximately 68 years).

7. Copy the running configuration to the startup configuration:

   `copy running-config startup-config`

   ```
   IP_switch_A_1(config)# copy running-config startup-config
   ```

8. Displays the keychain configuration:

   `show key chain name`

   ```
   IP_switch_A_1(config-macseckeychain-macseckey)# show key chain 1
   ```

Configure a MACsec policy

Steps

1. Enter global configuration mode:

   `configure terminal`

   ```
   IP_switch_A_1# configure terminal
   IP_switch_A_1(config)#
   ```

2. Create a MACsec policy:

`macsec policy name`

```
IP_switch_A_1(config)# macsec policy abc
IP_switch_A_1(config-macsec-policy)#
```

3. Configure one of the following ciphers, GCM-AES-128, GCM-AES-256, GCM-AES-XPN-128, or GCM-AES-XPN-256:

   `cipher-suite name`

   ```
   IP_switch_A_1(config-macsec-policy)# cipher-suite GCM-AES-256
   ```

4. Configure the key server priority to break the tie between peers during a key exchange:

   `key-server-priority number`

   ```
   switch(config-macsec-policy)# key-server-priority 0
   ```

5. Configure the security policy to define the handling of data and control packets:

   `security-policy security policy`

   Choose a security policy from the following options:

   - must-secure — packets not carrying MACsec headers are dropped
   - should-secure — packets not carrying MACsec headers are permitted (this is the default value)

   ```
   IP_switch_A_1(config-macsec-policy)# security-policy should-secure
   ```

6. Configure the replay protection window so the secured interface does not accept a packet that is less than the configured window size: `window-size number`

   The replay protection window size represents the maximum out-of-sequence frames that MACsec accepts and are not discarded. The range is from 0 to 596000000.

   ```
   IP_switch_A_1(config-macsec-policy)# window-size 512
   ```

7. Configure the time in seconds to force an SAK rekey:

   `sak-expiry-time time`

   You can use this command to change the session key to a predictable time interval. The default is 0.

```
IP_switch_A_1(config-macsec-policy)# sak-expiry-time 100
```

8. Configure one of the following confidentiality offsets in the layer 2 frame where encryption begins:

   `conf-offsetconfidentiality offset`

   Choose from the following options:

   - CONF-OFFSET-0.
   - CONF-OFFSET-30.
   - CONF-OFFSET-50.

   ```
   IP_switch_A_1(config-macsec-policy)# conf-offset CONF-OFFSET-0
   ```

   This command might be necessary for intermediate switches to use packet headers (dmac, smac, etype) like MPLS tags.

9. Copy the running configuration to the startup configuration:

   `copy running-config startup-config`

   ```
   IP_switch_A_1(config)# copy running-config startup-config
   ```

10. Display the MACsec policy configuration:

    `show macsec policy`

    ```
    IP_switch_A_1(config-macsec-policy)# show macsec policy
    ```

Enable Cisco MACsec encryption on the interfaces

1. Enter global configuration mode:

   `configure terminal`

   ```
   IP_switch_A_1# configure terminal
   IP_switch_A_1(config)#
   ```

2. Select the interface that you configured with MACsec encryption.

   You can specify the interface type and identity. For an Ethernet port, use ethernet slot/port.

```
IP_switch_A_1(config)# interface ethernet 1/15
switch(config-if)#
```

3. Add the keychain and policy to be configured on the interface to add the MACsec configuration:

   `macsec keychain keychain-name policy policy-name`

```
IP_switch_A_1(config-if)# macsec keychain 1 policy abc
```

4. Repeat steps 1 and 2 on all interfaces where MACsec encryption is to be configured.
5. Copy the running configuration to the startup configuration:

   `copy running-config startup-config`

```
IP_switch_A_1(config)# copy running-config startup-config
```

Disable Cisco MACsec Encryption WAN ISLs in MetroCluster IP configurations

You might need to disable MACsec encryption for Cisco 9336C switches on the WAN ISLs in a MetroCluster IP configuration.

Steps

1. Enter global configuration mode:

   `configure terminal`

```
IP_switch_A_1# configure terminal
IP_switch_A_1(config)#
```

2. Disable the MACsec configuration on the device:

   `macsec shutdown`

```
IP_switch_A_1(config)# macsec shutdown
```

   Selecting the “no” option restores the MACsec feature.

3. Select the interface that you already configured with MACsec.

   You can specify the interface type and identity. For an Ethernet port, use ethernet slot/port.

```
IP_switch_A_1(config)# interface ethernet 1/15
switch(config-if)#
```

4. Remove the keychain and policy configured on the interface to remove the MACsec configuration:

   `no macsec keychain keychain-name policy policy-name`

```
IP_switch_A_1(config-if)# no macsec keychain 1 policy abc
```

5. Repeat steps 3 and 4 on all interfaces where MACsec is configured.
6. Copy the running configuration to the startup configuration:

   `copy running-config startup-config`

```
IP_switch_A_1(config)# copy running-config startup-config
```

Verifying the MACsec configuration

Steps

1. Repeat **all** of the previous procedures on the second switch within the configuration to establish a MACsec session.
2. Run the following commands to verify that both switches are successfully encrypted:
   a. Run: `show macsec mka summary`
   b. Run: `show macsec mka session`
   c. Run: `show macsec mka statistics`

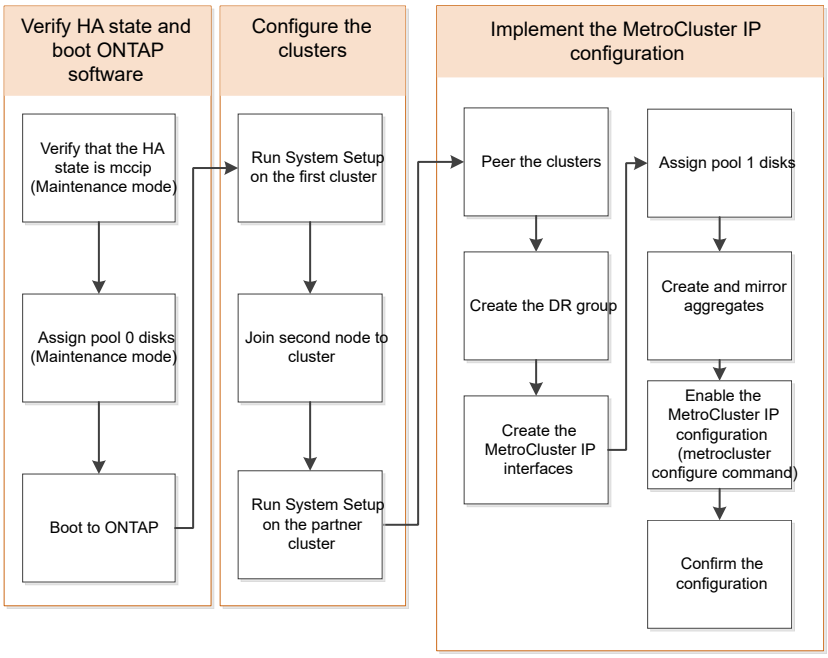
   You can verify the MACsec configuration using the following commands:

Command	Displays information about…
`show macsec mka session interface typeslot/port number`	The MACsec MKA session for a specific interface or for all interfaces
`show key chain name`	The key chain configuration
`show macsec mka summary`	The MACsec MKA configuration
`show macsec policy policy-name`	The configuration for a specific MACsec policy or for all MACsec policies

Configure the MetroCluster software in ONTAP

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.



Handling eight-node Configurations

An eight-node configuration will consist of two DR groups. Configure the first DR group by using the tasks in this section.

Then perform the tasks in Expanding a four-node MetroCluster IP configuration to an eight-node configuration

Gathering required information

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

You can use these links to download csv files and fill in the tables with your site-specific information.

MetroCluster IP setup worksheet, site_A

MetroCluster IP setup worksheet, site_B

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

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.

Steps

1. Display the HA state of the controller module and chassis:

   ```
   ha-config show
   ```

   The controller module and chassis should show the value “mccip”.

2. If the displayed system state of the controller is not “mccip”, set the HA state for the controller:

```
ha-config modify controller mccip
```

3. If the displayed system state of the chassis is not “mccip”, set the HA state for the chassis:

```
ha-config modify chassis mccip
```

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

Restoring system defaults on a controller module

Reset and restore defaults on the controller modules.

1. At the LOADER prompt, return environmental variables to their default setting: `set-defaults`
2. Boot the node to the boot menu: `boot_ontap menu`

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

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.

     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.
```

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 (ONTAP 9.4 and later)

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

This procedure applies to configurations running ONTAP 9.4 or later.

To determine if your system requires manual disk assignment, you should review Considerations for automatic drive assignment and ADP systems in ONTAP 9.4 and later.

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)

Steps

1. Display the boot menu:

   `boot_ontap menu`

2. Select option 9a.

   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 MetroCluster 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
```

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

5. Exit Maintenance mode:

```
halt
```

6. Display the boot menu:

```
boot_ontap menu
```

7. Select option "4" from the boot menu and let the system boot.
8. Repeat these steps on the other nodes in the MetroCluster IP configuration.
9. Proceed to Setting up ONTAP.

Manually assigning drives for pool 0 (ONTAP 9.3)

If you have at least two disk shelves for each node, you use ONTAP's auto-assignment functionality to automatically assign the local (pool 0) disks.

About this task

While the node is in Maintenance mode, you must first assign a single disk on the appropriate shelves to pool 0. ONTAP then automatically assigns the rest of the disks on the shelf to the same pool. This task is not required on systems received from the factory, which have pool 0 to contain the pre-configured root aggregate.

This procedure applies to configurations running ONTAP 9.3.

This procedure is not required if you received your MetroCluster configuration from the factory. Nodes from the factory are configured with pool 0 disks and root aggregates.

This procedure can be used only if you have at least two disk shelves for each node, which allows shelf-level autoassignment of disks. If you cannot use shelf-level autoassignment, you must manually assign your local disks so that each node has a local pool of disks (pool 0).

These steps must be performed in Maintenance mode.

Examples in this section assume the following disk shelves:

- node_A_1 owns disks on:
  - site_A-shelf_1 (local)
  - site_B-shelf_2 (remote)
- node_A_2 is connected to:
  - site_A-shelf_3 (local)
  - site_B-shelf_4 (remote)

- node_B_1 is connected to:
  - site_B-shelf_1 (local)
  - site_A-shelf_2 (remote)
- node_B_2 is connected to:
  - site_B-shelf_3 (local)
  - site_A-shelf_4 (remote)

Steps

1. Manually assign a single disk for root aggregate on each node:

   `disk assign` *`disk-id`* `-p 0 -s` *`local-node-sysid`*

   The manual assignment of these disks allows the ONTAP autoassignment feature to assign the rest of the disks on each shelf.

   a. On node_A_1, manually assign one disk from local site_A-shelf_1 to pool 0.
   b. On node_A_2, manually assign one disk from local site_A-shelf_3 to pool 0.
   c. On node_B_1, manually assign one disk from local site_B-shelf_1 to pool 0.
   d. On node_B_2, manually assign one disk from local site_B-shelf_3 to pool 0.
2. Boot each node at site A, using option 4 on the boot menu:

   You should complete this step on a node before proceeding to the next node.

   a. Exit Maintenance mode:

      `halt`

   b. Display the boot menu:

      `boot_ontap menu`

   c. Select option 4 from the boot menu and proceed.
3. Boot each node at site B, using option 4 on the boot menu:

   You should complete this step on a node before proceeding to the next node.

   a. Exit Maintenance mode:

      `halt`

   b. Display the boot menu:

      `boot_ontap menu`

   c. Select option 4 from the boot menu and proceed.

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.

If you need to netboot the new controllers, see Netbooting the new controller modules.

About this task

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

Steps

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.

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.

      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
NetApp 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.

   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.

   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.

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: lab.netapp.com
Enter the name server IP addresses: 172.19.2.30
DNS lookup for the admin Vserver will use the lab.netapp.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.

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 ONTAP System Manager, pointing your web browser to the cluster management IP address (https://172.17.12.153).

Cluster management using System Manager (Versions 9.0 to 9.6)

ONTAP System Manager (Version 9.7 and later)

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

The expected configuration state of HA and storage failover is as follows:

- HA mode is configured but storage failover is not enabled.
- HA takeover capability is disabled.
- HA interfaces are offline.
- HA mode, storage failover, and interfaces are configured later in the process.

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.

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.
```

6. Repeat these steps on the partner cluster.

What to do next

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.

This task does not apply to an AFF A800 system with an internal shelf and no external shelves.

Considerations for automatic drive assignment and ADP systems in ONTAP 9.4 and later

Steps

1. Disable automatic drive assignment:

   ```
   storage disk option modify -node node_name -autoassign off
   ```

2. 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 in ONTAP 9.4 and later

Steps

1. Verify that pool 0 drives are assigned automatically:

   ```
   disk show
   ```

   The following example shows the "cluster_A" output for an AFF A800 system with no external shelves.

   One quarter (8 drives) were automatically assigned to "node_A_1" and one quarter were automatically assigned to "node_A_2". The remaining drives will be remote (pool 1) drives for "node_B_1" and "node_B_2".

   ```
   cluster_A::*> disk show
                    Usable     Disk      Container           Container
   ```

```
Disk             Size       Shelf Bay Type    Type        Name
Owner
---------------- ---------- ----- --- ------- ----------- ---------
--------
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_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
node_A_2:0n.24   -          0     24  SSD-NVM unassigned  -         -
node_A_2:0n.25   -          0     25  SSD-NVM unassigned  -         -
node_A_2:0n.26   -          0     26  SSD-NVM unassigned  -         -
node_A_2:0n.27   -          0     27  SSD-NVM unassigned  -         -
node_A_2:0n.28   -          0     28  SSD-NVM unassigned  -         -
node_A_2:0n.29   -          0     29  SSD-NVM unassigned  -         -
node_A_2:0n.30   -          0     30  SSD-NVM unassigned  -         -
node_A_2:0n.31   -          0     31  SSD-NVM unassigned  -         -
node_A_2:0n.36   -          0     36  SSD-NVM unassigned  -         -
node_A_2:0n.37   -          0     37  SSD-NVM unassigned  -         -
```

```
node_A_2:0n.38   -          0     38  SSD-NVM unassigned  -         -
node_A_2:0n.39   -          0     39  SSD-NVM unassigned  -         -
node_A_2:0n.40   -          0     40  SSD-NVM unassigned  -         -
node_A_2:0n.41   -          0     41  SSD-NVM unassigned  -         -
node_A_2:0n.42   -          0     42  SSD-NVM unassigned  -         -
node_A_2:0n.43   -          0     43  SSD-NVM unassigned  -         -
32 entries were displayed.
```

The following example shows the "cluster_B" output:

```
cluster_B::> disk show
                 Usable     Disk              Container   Container
Disk             Size       Shelf Bay Type    Type        Name
Owner
---------------- ---------- ----- --- ------- ----------- ---------
--------

Info: This cluster has partitioned disks. To get a complete list of
spare disk
capacity use "storage aggregate show-spare-disks".
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_2:0n.0    1.75TB     0     0   SSD-NVM shared
aggr0_node_B_1_0 node_B_2
node_B_2:0n.1    1.75TB     0     1   SSD-NVM shared
aggr0_node_B_1_0 node_B_2
node_B_2:0n.2    1.75TB     0     2   SSD-NVM shared
aggr0_node_B_1_0 node_B_2
node_B_2:0n.3    1.75TB     0     3   SSD-NVM shared
aggr0_node_B_1_0 node_B_2
node_B_2:0n.4    1.75TB     0     4   SSD-NVM shared
```

```
aggr0_node_B_1_0 node_B_2
node_B_2:0n.5    1.75TB     0     5   SSD-NVM shared
aggr0_node_B_1_0 node_B_2
node_B_2:0n.6    1.75TB     0     6   SSD-NVM shared
aggr0_node_B_1_0 node_B_2
node_B_2:0n.7    1.75TB     0     7   SSD-NVM shared      -
node_B_2
node_B_2:0n.24   -          0     24  SSD-NVM unassigned  -         -
node_B_2:0n.25   -          0     25  SSD-NVM unassigned  -         -
node_B_2:0n.26   -          0     26  SSD-NVM unassigned  -         -
node_B_2:0n.27   -          0     27  SSD-NVM unassigned  -         -
node_B_2:0n.28   -          0     28  SSD-NVM unassigned  -         -
node_B_2:0n.29   -          0     29  SSD-NVM unassigned  -         -
node_B_2:0n.30   -          0     30  SSD-NVM unassigned  -         -
node_B_2:0n.31   -          0     31  SSD-NVM unassigned  -         -
node_B_2:0n.36   -          0     36  SSD-NVM unassigned  -         -
node_B_2:0n.37   -          0     37  SSD-NVM unassigned  -         -
node_B_2:0n.38   -          0     38  SSD-NVM unassigned  -         -
node_B_2:0n.39   -          0     39  SSD-NVM unassigned  -         -
node_B_2:0n.40   -          0     40  SSD-NVM unassigned  -         -
node_B_2:0n.41   -          0     41  SSD-NVM unassigned  -         -
node_B_2:0n.42   -          0     42  SSD-NVM unassigned  -         -
node_B_2:0n.43   -          0     43  SSD-NVM unassigned  -         -
32 entries were displayed.

cluster_B::>
```

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.

Related information

Cluster and SVM peering express configuration

Considerations when using dedicated ports

Considerations when sharing data ports

Configuring intercluster LIFs for cluster peering

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.

Steps

1. List the ports in the cluster:

   `network port show`

   For complete command syntax, see the man page.

   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
```

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.

   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
```

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`*

   The following example assigns ports "e0e" and" e0f" to failover group "intercluster01" on system "SVMcluster01":

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

4. Verify that the failover group was created:

   `network interface failover-groups show`

   For complete command syntax, see the man page.

```
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
```

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

ONTAP version	Command
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*
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 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
```

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
```

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 "SVMe0e" 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
```

Related information

Considerations when using dedicated ports

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.

Steps

1. List the ports in the cluster:

   `network port show`

   For complete command syntax, see the man page.

   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
```

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.

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
```

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.

```
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
```

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.

The following example shows that 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
```

Related information

Considerations when sharing data ports

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.

About this task

- You must have created intercluster LIFs on every node in the clusters that are being peered.
- The clusters must be running ONTAP 9.3 or later.

Steps

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.

   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.
```

2. On the 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.

   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.

3. Verify that the cluster peer relationship was created:

   `cluster peer show -instance`

```
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
```

4. Check the connectivity and status of the nodes in the peer relationship:

   `cluster peer health show`

```
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
```

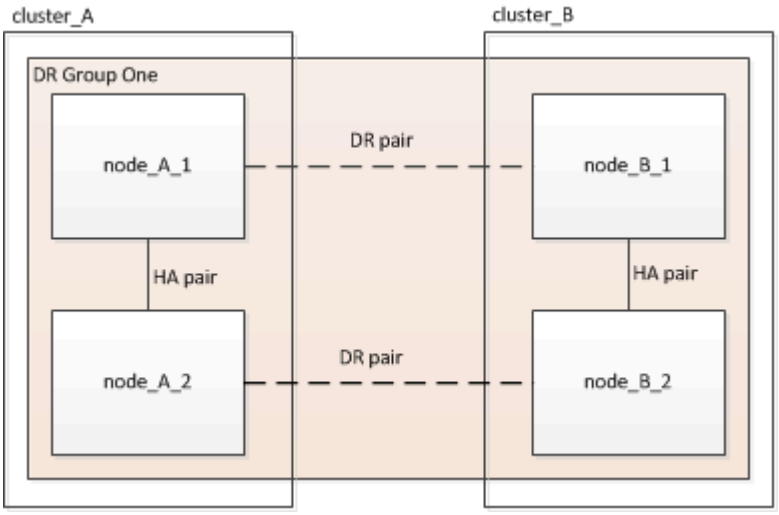
Creating the DR group

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.

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



Steps

1. Verify that the nodes are ready for creation of the DR group by entering the following command on each node:

   ```
   metrocluster configuration-settings show-status
   ```

   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.
   ```

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

About this task

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

- You must create two interfaces for each node. The interfaces must be associated with the VLANs defined in the MetroCluster RCF file.
- You must create all MetroCluster IP interface "A" ports in the same VLAN and all MetroCluster IP interface "B" ports in the other VLAN. Refer to Considerations for MetroCluster IP configuration.

- Certain platforms use a VLAN for the MetroCluster IP interface. By default, each of the two ports use a different VLAN: 10 and 20. You can also specify a different (non-default) VLAN higher than 100 (between 101 and 4095) using the `-vlan-id parameter` in the `metrocluster configuration-settings interface create` command.
- Beginning with ONTAP 9.9.1, if you are using a layer 3 configuration, you must also specify the `-gateway` parameter when creating MetroCluster IP interfaces. Refer to Considerations for layer 3 wide-area networks.

The following platform models can be added to the existing MetroCluster configuration if the VLANs used are 10/20 or greater than 100. If any other VLANs are used, then these platforms cannot be added to the existing configuration as the MetroCluster interface cannot be configured. If you are using any other platform, the VLAN configuration is not relevant as this is not required in ONTAP.

AFF platforms	FAS platforms

• AFF A220 • AFF A250 • AFF A400	• FAS2750 • FAS500f • FAS8300 • FAS8700

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

The physical ports used by the MetroCluster IP interfaces depends on the platform model, as shown in the following table.

Platform model	MetroCluster IP port	Note
AFF A900 and FAS9500	e5b	
	e7b	
AFF A800	e0b	
	e1b	

Platform model	MetroCluster IP port	Note
AFF A700 and FAS9000	e5a	
	e5b	
AFF A400	e1a	
	e1b	
AFF A320	e0g	
	e0h	
AFF A300 and FAS8200	e1a	
	e1b	
AFF A220 and FAS2750	e0a	On these systems, these physical ports are also used as cluster interfaces.
	e0b	
AFF A250 and FAS500f	e0c	
	e0d	
FAS8300 and FAS8700	e1a	
	e1b	

The port usage in the following examples is for an AFF A700 or a FAS9000 system.

Steps

1. Confirm that each node has disk automatic assignment enabled:

   ```
   storage disk option show
   ```

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

   The Auto Assign column indicates whether disk automatic assignment 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.
```

2. Verify you can create MetroCluster IP interfaces on the nodes:

   `metrocluster configuration-settings show-status`

   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.
```

3. Create the interfaces on node_A_1.

   - The port usage in the following examples is for an AFF A700 or a FAS9000 system (e5a and e5b). You must configure the interfaces on the correct ports for your platform model, as given above.
   - Beginning with ONTAP 9.9.1, if you are using a layer 3 configuration, you must also specify the `-gateway` parameter when creating MetroCluster IP interfaces. Refer to Considerations for layer 3 wide-area networks.
   - On platform models that support VLANs for the MetroCluster IP interface, you can include the `-vlan-id` parameter if you don't want to use the default VLAN IDs.

   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*

      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 e5a -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`*

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 e5b -address
10.1.2.1 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

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

4. Create the interfaces on node_A_2.

- The port usage in the following examples is for an AFF A700 or a FAS9000 system (e5a and e5b). You must configure the interfaces on the correct ports for your platform model, as given above.
- Beginning with ONTAP 9.9.1, if you are using a layer 3 configuration, you must also specify the `-gateway` parameter when creating MetroCluster IP interfaces. Refer to Considerations for layer 3 wide-area networks.
- On platform models that support VLANs for the MetroCluster IP interface, you can include the `-vlan-id` parameter if you don't want to use the default VLAN IDs.

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`*

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 e5a -address
10.1.1.2 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

On platform models that support VLANs for the MetroCluster IP interface, you can include the `-vlan -id` parameter if you don't want to use the default VLAN IDs. The following example shows the command for an AFF A220 system with a VLAN ID of 120:

```
cluster_A::> metrocluster configuration-settings interface create
-cluster-name cluster_A -home-node node_A_2 -home-port e0a -address
10.1.1.2 -netmask 255.255.255.0 -vlan-id 120
[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`*

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 e5b -address
10.1.2.2 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

On platform models that support VLANs for the MetroCluster IP interface, you can include the `-vlan -id` parameter if you don't want to use the default VLAN IDs. The following example shows the command for an AFF A220 system with a VLAN ID of 220:

```
cluster_A::> metrocluster configuration-settings interface create
-cluster-name cluster_A -home-node node_A_2 -home-port e0b -address
10.1.2.2 -netmask 255.255.255.0 -vlan-id 220
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

5. Create the interfaces on "node_B_1".

- The port usage in the following examples is for an AFF A700 or a FAS9000 system (e5a and e5b). You must configure the interfaces on the correct ports for your platform model, as given above.
- Beginning with ONTAP 9.9.1, if you are using a layer 3 configuration, you must also specify the `-gateway` parameter when creating MetroCluster IP interfaces. Refer to Considerations for layer 3 wide-area networks.
- On platform models that support VLANs for the MetroCluster IP interface, you can include the `-vlan-id` parameter if you don't want to use the default VLAN IDs.

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`*

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_B -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_B::>
```

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`*

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_B -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_B::>
```

6. Create the interfaces on "node_B_2".

- The port usage in the following examples is for an AFF A700 or a FAS9000 system (e5a and e5b). You must configure the interfaces on the correct ports for your platform model, as given above.
- Beginning with ONTAP 9.9.1, if you are using a layer 3 configuration, you must also specify the `-gateway` parameter when creating MetroCluster IP interfaces. Refer to Considerations for layer 3 wide-area networks.
- On platform models that support VLANs for the MetroCluster IP interface, you can include the `-vlan-id` parameter if you don't want to use the default VLAN IDs.

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 -address ip-address -netmask netmask`

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`

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::>
```

7. Verify that the interfaces have been configured:

`metrocluster configuration-settings interface show`

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: e5a
                       10.1.1.1     255.255.255.0   -         completed
                  Home Port: e5b
                       10.1.2.1     255.255.255.0   -         completed
                  node_A_2
                  Home Port: e5a
                       10.1.1.2     255.255.255.0   -         completed
                  Home Port: e5b
                       10.1.2.2     255.255.255.0   -         completed
      cluster_B  node_B_1
                  Home Port: e5a
                       10.1.1.3     255.255.255.0   -         completed
                  Home Port: e5b
                       10.1.2.3     255.255.255.0   -         completed
                  node_B_2
                  Home Port: e5a
                       10.1.1.4     255.255.255.0   -         completed
                  Home Port: e5b
                       10.1.2.4     255.255.255.0   -         completed
8 entries were displayed.
cluster_A::>
```

8. Verify that the nodes are ready to connect the MetroCluster interfaces:

   `metrocluster configuration-settings show-status`

   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.
```

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

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

   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::>
   ```

10. Verify that the connections have been established:

    `metrocluster configuration-settings show-status`

    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.
    ```

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 MetroCluster IP initiators on each cluster that should appear in the output.

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

       The following example shows the eight MetroCluster IP 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
```

12. Verify that the nodes are ready for final implementation of the MetroCluster configuration:

```
metrocluster node show
```

```
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.

Configuration type	Procedure
The systems meet the requirements for automatic drive assignment or, if running ONTAP 9.3, were received from the factory.	Verifying disk assignment for pool 1 disks
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)
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)
The systems were not received from the factory and are running ONTAP 9.3Systems received from the factory are pre-configured with assigned drives.	Manually assigning disks for pool 1 (ONTAP 9.3)

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.

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

Considerations for automatic drive assignment and ADP systems in ONTAP 9.4 and later

Steps

1. Verify pool 1 disks are auto-assigned:

   `disk show`

   The following output shows the output for an AFF A800 system with no external shelves.

   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 (pool 1) disks for "node_B_1" and "node_B_2".

```
cluster_B::> disk show -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 in ONTAP 9.4 and later.

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 in bays 0-11 on site_B-shelf_2 (remote)
- node_A_2 is assigned drives in bays 12-23 on site_B-shelf_2 (remote)

Steps

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
      ```

```
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`

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

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.

Manually assigning disks for pool 1 (ONTAP 9.3)

If you have at least two disk shelves for each node, you use ONTAP's auto-assignment functionality to automatically assign the remote (pool1) disks.

Before you begin

You must first assign a disk on the shelf to pool 1. ONTAP then automatically assigns the rest of the disks on the shelf to the same pool.

About this task

This procedure applies to configurations running ONTAP 9.3.

This procedure can be used only if you have at least two disk shelves for each node, which allows shelf-level auto-assignment of disks.

If you cannot use shelf-level auto-assignment, you must manually assign your remote disks so that each node has a remote pool of disks (pool 1).

The ONTAP automatic disk assignment feature assigns the disks on a shelf-by-shelf basis. For example:

- All the disks on site_B-shelf_2 are auto-assigned to pool1 of node_A_1
- All the disks on site_B-shelf_4 are auto-assigned to pool1 of node_A_2
- All the disks on site_A-shelf_2 are auto-assigned to pool1 of node_B_1
- All the disks on site_A-shelf_4 are auto-assigned to pool1 of node_B_2

You must "seed" the auto-assignment by specifying a single disk on each shelf.

Steps

1. From each node in the MetroCluster IP configuration, assign a remote disk to pool 1.

a. Display the list of unassigned disks:

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

```
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. Select a remote disk (0m) and assign ownership of the disk to pool 1 of the first node (for example, "node_A_1"):

```
disk assign -disk disk-id -pool 1 -owner owner-node-name
```

The `disk-id` must identify a disk on a remote shelf of `owner-node-name`.

The ONTAP disk auto-assignment feature assigns all disks on the remote shelf that contains the specified disk.

c. After waiting at least 60 seconds for disk auto-assignment to take place, verify that the remote disks on the shelf were auto-assigned to pool 1:

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

> The iSCSI connection used to access the remote disks appears as device 0m.

The following output shows that the disks on shelf 23 have now been assigned and no longer appear:

```
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.i1.2L72       -    21  23 SSD     unassigned  -         -
node_A_2:0m.i1.2L74       -    21   1 SSD     unassigned  -         -
node_A_2:0m.i1.2L83       -    21  22 SSD     unassigned  -         -
node_A_2:0m.i1.2L90       -    21   7 SSD     unassigned  -         -
node_A_2:0m.i1.3L52       -    21   6 SSD     unassigned  -         -
node_A_2:0m.i1.3L59       -    21  13 SSD     unassigned  -         -
node_A_2:0m.i1.3L66       -    21  17 SSD     unassigned  -         -
node_A_2:0m.i1.3L73       -    21  12 SSD     unassigned  -         -
node_A_2:0m.i1.3L80       -    21   5 SSD     unassigned  -         -
node_A_2:0m.i1.3L81       -    21   2 SSD     unassigned  -         -
node_A_2:0m.i1.3L82       -    21  16 SSD     unassigned  -         -
node_A_2:0m.i1.3L91       -    21   3 SSD     unassigned  -         -
node_A_2:0m.i2.0L49       -    21  15 SSD     unassigned  -         -
node_A_2:0m.i2.0L50       -    21   4 SSD     unassigned  -         -
node_A_2:0m.i2.1L57       -    21  18 SSD     unassigned  -         -
node_A_2:0m.i2.1L58       -    21  11 SSD     unassigned  -         -
node_A_2:0m.i2.1L59       -    21  21 SSD     unassigned  -         -
node_A_2:0m.i2.1L65       -    21  20 SSD     unassigned  -         -
node_A_2:0m.i2.1L72       -    21   9 SSD     unassigned  -         -
node_A_2:0m.i2.1L80       -    21   0 SSD     unassigned  -         -
node_A_2:0m.i2.1L88       -    21   8 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 disks 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

About this task

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

Considerations for automatic drive assignment and ADP systems in ONTAP 9.4 and later

Steps

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`

> 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.

Steps

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.

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

Related information

Logical storage management

Creating a mirrored data aggregate on each node

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

About this task

- 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.
- 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

Steps

1. Display a list of available spares:

   ```
   storage disk show -spare -owner node_name
   ```

2. Create the aggregate:

   ```
   storage aggregate create -mirror true
   ```

   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

     > 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.

     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
     ```

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.

About this task

- There should be at least two non-root mirrored data aggregates on each cluster.

  You can verify this with the `storage aggregate show` command.

  > If you want to use a single mirrored data aggregate, then see Step 1 for instructions.

- The ha-config state of the controllers and chassis must be "mccip".

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.

> You must **not** configure Onboard Key Manager (OKM) or external key management before you run the command `metrocluster configure`.

Steps

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	a. 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 (*>).
	
	b. Configure the MetroCluster with the `-allow-with-one-aggregate true` parameter:
	
	`metrocluster configure -allow-with-one-aggregate true` *node-name*
	
	c. Return to the admin privilege level:
	
	`set -privilege admin`

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.
```

2. Verify the networking status on site A:

   `network port show`

   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.
```

3. Verify the MetroCluster configuration from both sites in the MetroCluster configuration.
   a. Verify the configuration from site A:

      `metrocluster show`

      ```
      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`

```
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
```

4. To avoid possible issues with nonvolatile memory mirroring, reboot each of the four nodes:

   `node reboot -node` *`node-name`* `-inhibit-takeover true`

5. Issue the `metrocluster show` command on both clusters to again verify the configuration.

Configuring the second DR group in an eight-node configuration

Repeat the previous tasks to configure the nodes in the second DR group.

Creating unmirrored data aggregates

You can optionally create unmirrored data aggregates for data that does not require the redundant mirroring provided by MetroCluster configurations.

About this task

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

> In MetroCluster IP configurations, remote unmirrored aggregates are not accessible after a switchover

> The unmirrored aggregates must be local to the node owning them.

- Drives and array LUNs 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.
- Aggregate names should conform to the naming scheme you determined when you planned your MetroCluster configuration.
- *Disks and aggregates management* contains more information about mirroring aggregates.

Steps

1. Enable unmirrored aggregate deployment:

   `metrocluster modify -enable-unmirrored-aggr-deployment true`

2. Verify that disk autoassignment is disabled:

```
disk option show
```

3. Install and cable the disk shelves that will contain the unmirrored aggregates.

   You can use the procedures in the Installation and Setup documentation for your platform and disk shelves.

   AFF and FAS Documentation Center

4. Manually assign all disks on the new shelf to the appropriate node:

   ```
   disk assign -disk disk-id -owner owner-node-name
   ```

5. Create the aggregate:

   ```
   storage aggregate create
   ```

   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 verify that the aggregate is created on a specific node, you should use the -node parameter or specify drives that are owned by that node.

   You must also ensure that you are only including drives on the unmirrored shelf to the aggregate.

   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 or array LUNs that are to be added to the aggregate
   - Number of drives to include
   - 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 or array LUNs 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.

     The following command creates a unmirrored aggregate with 10 disks:

     ```
     controller_A_1::> storage aggregate create aggr1_controller_A_1
     -diskcount 10 -node controller_A_1
     [Job 15] Job is queued: Create aggr1_controller_A_1.
     [Job 15] The job is starting.
     [Job 15] Job succeeded: DONE
     ```

6. Verify the RAID group and drives of your new aggregate:

   ```
   storage aggregate show-status -aggregate aggregate-name
   ```

7. Disable unmirrored aggregate deployment:

   ```
   metrocluster modify -enable-unmirrored-aggr-deployment false
   ```

8. Verify that disk autoassignment is enabled:

   ```
   disk option show
   ```

Related information

Disk and aggregate management

Checking the MetroCluster configuration

You can check that the components and relationships in the MetroCluster configuration are working correctly.

About this task

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.

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.

Steps

1. Check the configuration:

   ```
   metrocluster check run
   ```

   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.
```

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.

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.
```

Related information

Disk and aggregate management

Network and LIF management

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, beginning 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.

If you must install a new instance of the ONTAP Mediator, follow the directions to install and configure the software.

Configuring the ONTAP Mediator service for unplanned automatic switchover

If you must install a new instance of the Tiebreaker software, follow the directions to install and configure the software.

About this task

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

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:

     ```
     metrocluster configuration-settings mediator add -mediator-address ip-address-of-mediator-host
     ```

   - If you are using the Tiebreaker software, refer to the Tiebreaker documentation.

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 Cluster Management with the CLI contains additional information under the section *Managing configuration backups*.

Configure the ONTAP Mediator service for unplanned automatic switchover

Prepare to install the ONTAP Mediator service

Your environment must meet certain requirements.

The following requirements apply to one disaster recovery group (DR group). Learn more about DR groups.

- If you plan on updating your Linux version, do so before you install the most current ONTAP Mediator service.
- The ONTAP Mediator service and MetroCluster Tiebreaker software should not both be used with the same MetroCluster configuration.
- The ONTAP Mediator must be installed on a LINUX host at a separate location from the MetroCluster sites.

  The connectivity between the ONTAP Mediator and each site must be two separate failure domains.
- The ONTAP Mediator service can support up to five MetroCluster configurations simultaneously.
- Automatic unplanned switchover is supported in ONTAP 9.7 and later.

Network requirements for using Mediator in a MetroCluster configuration

To install the ONTAP Mediator service in a MetroCluster configuration, you must make sure that the configuration meets several network requirements.

- Latency

  Maximum latency of less than 75ms (RTT).

  Jitter must be no more than 5ms.
- MTU

  The MTU size must be at least 1400.
- Packet loss

  For both Internet Control Message Protocol (ICMP) and TCP traffic, packet loss must be less than 0.01%.
- Bandwidth

  The link between the Mediator service and one DR group must have at least 20Mbps of bandwidth.
- Independent connectivity

  Independent connectivity between each site and the ONTAP Mediator is required. A failure in one site must not interrupt the IP connectivity between the other two unaffected sites.

Host requirements for the ONTAP Mediator in a MetroCluster configuration

You must ensure that the configuration meets several host requirements.

- ONTAP Mediator must be installed at an external site that is physically separated from the two ONTAP clusters.
- ONTAP Mediator supports a maximum number of five MetroCluster configurations.
- ONTAP Mediator does not require more than the host operating system’s minimum requirements for CPU and memory (RAM).
- In addition to the host operating system’s minimum requirements, at least 30GB of additional usable disk space must be available.
  - Each DR group requires up to 200MB of disk space.

Firewall requirements for ONTAP Mediator

ONTAP Mediator uses a number of ports to communicate with specific services.

If you are 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.

The following table lists the ports that you must allow in your firewall:

Port/services	Source	Destination	Purpose
31784/tcp	ONTAP cluster management interfaces	ONTAP Mediator web server	REST API (HTTPS)
3260/tcp	ONTAP cluster	ONTAP Mediator iSCSI targets	iSCSI data connection for mailboxes

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 from an immediately prior version to the current version.
- All Mediator versions are supported on MetroCluster IP configurations running ONTAP 9.7 or later.

Install or upgrade the ONTAP Mediator service

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.

Configure 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.

  Install or upgrade the ONTAP Mediator service

- 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.

Steps

1. Add the ONTAP Mediator service to ONTAP:

   `metrocluster configuration-settings mediator add -mediator-address` *`ip-address-of-mediator-host`*

   You will be prompted for the username and password for the Mediator admin user account.

2. Verify that the automatic switchover feature in enabled:

   `metrocluster show`

3. Verify that the Mediator is now running.
   a. Show the Mediator virtual disks:

      `storage disk show -container-type mediator`

      ```
       cluster_A::> storage disk show -container-type mediator
                               Usable           Disk    Container
      Container
       Disk                    Size Shelf Bay Type    Type        Name
      Owner
       ---------------- ---------- ----- --- ------- ----------- ---------
      --------
       NET-1.5                    -     -   - VMDISK  mediator    -
      node_A_2
       NET-1.6                    -     -   - VMDISK  mediator    -
      node_B_1
       NET-1.7                    -     -   - VMDISK  mediator    -
      node_B_2
       NET-1.8                    -     -   - VMDISK  mediator    -
      node_A_1
      ```

   b. Set the privilege mode to advanced:

      `set advanced`

      ```
       cluster_A::> set advanced
      ```

c. Display the initiators labelled as mediator:

```
storage iscsi-initiator show -label mediator
```

```
 cluster_A::*> storage iscsi-initiator show -label mediator
   (storage iscsi-initiator show)
  +
Status
 Node Type Label    Target Portal      Target Name
Admin/Op
 ---- ---- -------- ------------------
-------------------------------- --------
 node_A_1
      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
 node_A_2
      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
```

Unconfigure the ONTAP Mediator service from the MetroCluster IP configuration

You can unconfigure the ONTAP Mediator service from the MetroCluster IP configuration.

Before you begin

You must have successfully installed and configured the ONTAP Mediator on a network location that can be reached by both MetroCluster sites.

About this task

You must perform this task on both clusters in the MetroCluster IP configuration.

Steps

1. Unconfigure the ONTAP Mediator service by using the following command:

```
metrocluster configuration-settings mediator remove
```

a. Check if there are any broken disks by using the following command:

```
disk show -broken
```

Example

```
There are no entries matching your query.
```

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.

Steps

1. Remove the current ONTAP Mediator from the MetroCluster configuration:

   `metrocluster configuration-settings mediator remove`

2. Establish the new ONTAP Mediator connection to the MetroCluster configuration:

   `metrocluster configuration-settings mediator add -mediator-address` *`ip-address-of-mediator-host`*

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) and Solaris Fibre Channel protocols) 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 is issued from will switch over the partner site.

For MetroCluster IP configurations:

- For ONTAP 9.4 and earlier:
  - Mirrored aggregates will become degraded after the negotiated switchover.
- For ONTAP 9.5 and later:
  - Mirrored aggregates will remain in normal state if the remote storage is accessible.

- Mirrored aggregates will become degraded after the negotiated switchover if access to the remote storage is lost.

- For ONTAP 9.8 and later:
  - Unmirrored aggregates that are located at the disaster site will become unavailable if access to the remote storage is lost. This might lead to a controller outage.

Steps

1. Confirm that all nodes are in the configured state and normal mode:

   `metrocluster node show`

```
cluster_A::>  metrocluster node show

Cluster                     Configuration State    Mode
--------------------------- ---------------------
-------------------------
 Local: cluster_A               configured             normal
Remote: cluster_B               configured             normal
```

2. Begin the switchover operation:

   `metrocluster switchover`

```
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".
```

3. Confirm that the local cluster is in the configured state and switchover mode:

   `metrocluster node show`

```
cluster_A::>  metrocluster node show

Cluster                     Configuration State    Mode
--------------------------- ---------------------
-------------------------
Local: cluster_A                configured             switchover
Remote: cluster_B               not-reachable          -
              configured             normal
```

4. Confirm that the switchover operation was successful:

   `metrocluster operation show`

```
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: -
```

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 and Solaris FC 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.

Steps

1. If the system is running ONTAP 9.4 or earlier, heal the data aggregate:

   `metrocluster heal aggregates`

   The following example shows the successful completion of the command:

   ```
   cluster_A::> metrocluster heal aggregates
   [Job 936] Job succeeded: Heal Aggregates is successful.
   ```

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. 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.
```

3. Verify that healing is completed:

   `metrocluster node show`

   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
```

   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.

4. Verify that all aggregates are mirrored:

   `storage aggregate show`

   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   -
```

5. Check the status of switchback recovery:

   `metrocluster node show`

```
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.
```

6. Perform the switchback:

```
metrocluster switchback
```

```
cluster_A::> metrocluster switchback
[Job 938] Job succeeded: Switchback is successful.Verify switchback
```

7. Confirm status of the nodes:

```
metrocluster node show
```

```
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.
```

8. Confirm status of the MetroCluster operation:

```
metrocluster operation show
```

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.

Steps

1. Turn off the power of the PDUs on the left-hand side of the rack containing the MetroCluster components.
2. Monitor the result on the console:

   ```
   system environment sensors show -state fault

   storage shelf show -errors
   ```

   ```
   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
   ```

3. Turn the power back on to the left-hand PDUs.
4. Make sure that ONTAP clears the error condition.

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.

Steps

1. Check the storage failover status:

   ```
   storage failover show
   ```

   ```
   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.
   ```

2. Check the aggregate status:

   ```
   storage aggregate show
   ```

```
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
```

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*
```

```
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%
```

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.

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.
```

5. Physically power off the shelf that you selected.
6. Check the aggregate status again:

   ```
   storage aggregate show

   storage aggregate show -r -node node_A_2 !*root
   ```

   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.
```

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*
```

```
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%
```

8. Physically power on the shelf.

   Resynchronization starts automatically.

9. Verify that resynchronization has started:

   `storage aggregate show`

   The affected aggregate should have a RAID status of "resyncing", 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
```

10. Monitor the aggregate to confirm that resynchronization is complete:

    `storage aggregate show`

    The affected aggregate should have a RAID status of "normal", 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
```

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.

You cannot reverse the MetroCluster unconfiguration. This process should only be done with the assistance of technical support. Please contact NetApp Technical Support and reference the appropriate guide for your configuration from the How to remove nodes from a MetroCluster configuration - Resolution Guide.

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.

The ONTAP configuration of the two clusters, including networking, should be identical, because the MetroCluster feature relies on the ability of a cluster to seamlessly serve data for its partner in the event of a switchover.

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.

MetroCluster operations in ONTAP System Manager

Depending on your ONTAP version, some MetroCluster-specific operations can be performed using ONTAP System Manager.

- Switchover and switchback in MetroCluster IP configurations (beginning with ONTAP 9.7).
- Provision and grow of mirrored aggregates in the MetroCluster IP configurations (beginning with ONTAP 9.8).

  Unmirrored aggregates are not supported in System Manager.

FlexCache support in a MetroCluster configuration

Beginning 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.

Related information

FlexCache volumes management

FabricPool support in MetroCluster configurations

Beginning with ONTAP 9.7, MetroCluster configurations support FabricPool storage tiers.

For general information on using FabricPools, see Disk and aggregate management.

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

> 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 Disk and aggregate management.

Steps

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 Disk and aggregate management

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

Beginning with ONTAP 9.6 MetroCluster configurations support FlexGroup volumes.

Job schedules in a MetroCluster configuration

In ONTAP 9.3 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).

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 8.3 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.

Related information

Network and LIF management

IPspace object replication and subnet configuration requirements

Requirements for LIF creation in a MetroCluster configuration

LIF replication and placement requirements and issues

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.

Related information

Requirements for LIF creation in a MetroCluster configuration

LIF replication and placement requirements and issues

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
- NFS and SAN configurations: You must use different storage virtual machines (SVMs) for both the unmirrored and mirrored aggregates

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.

Related information

IPspace object replication and subnet configuration requirements

LIF replication and placement requirements and 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.
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.

> 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.

Related information

IPspace object replication and subnet configuration requirements

Requirements for LIF creation in a MetroCluster configuration

Volume creation on a root aggregate

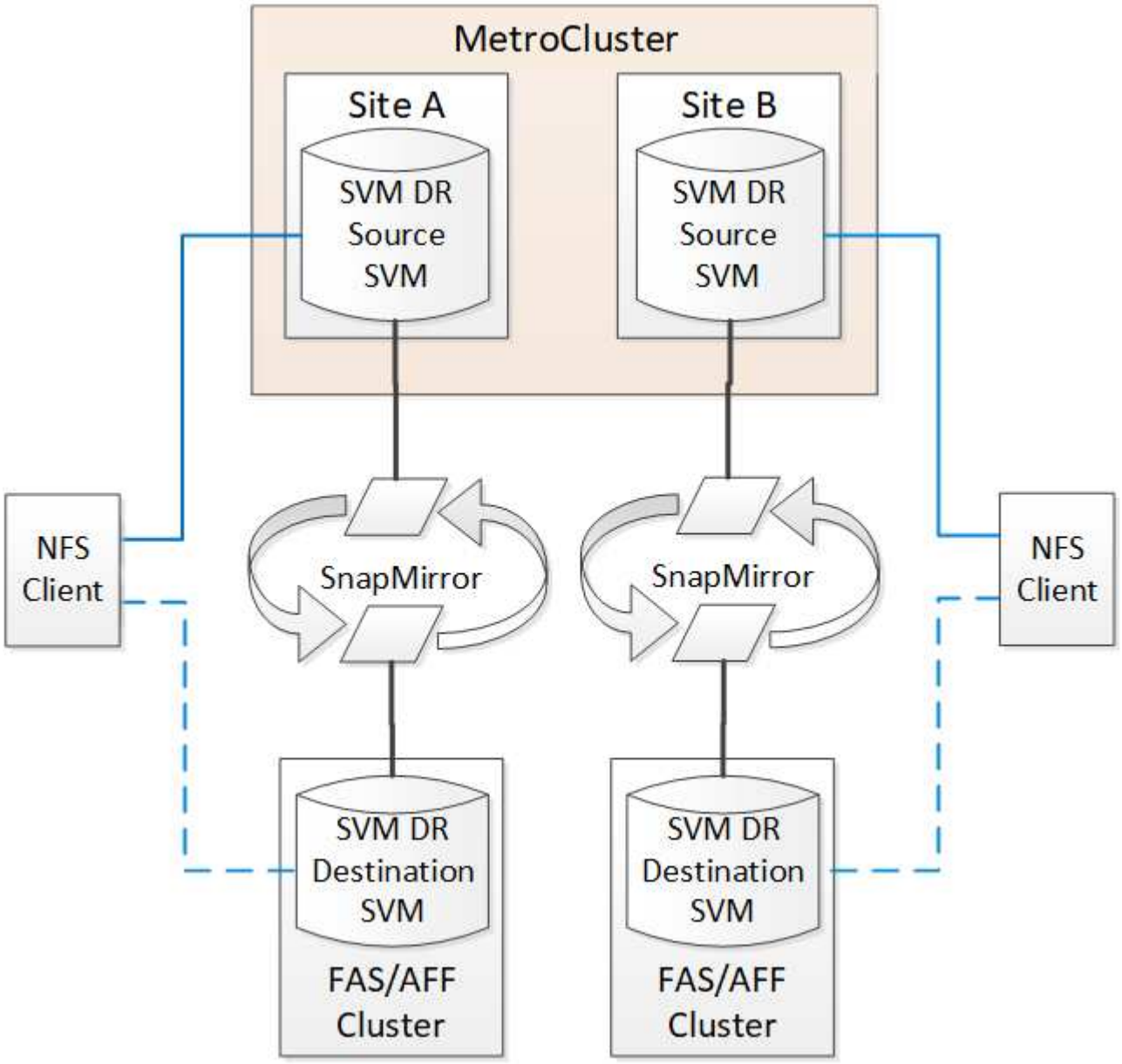
The system does not allow the creation of new volumes on the root aggregate (an aggregate with an HA policy of CFO) of a node in a MetroCluster configuration.

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

Beginning 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.

Beginning with ONTAP 9.11.1, both sites within a MetroCluster configuration can be the source for an SVM DR relationship with a FAS or AFF destination cluster as shown in the following image.



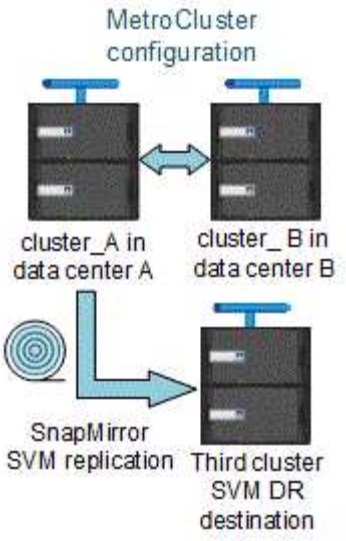
You should be aware of the following requirements and limitations of using SVMs with SnapMirror disaster recovery:

- 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.

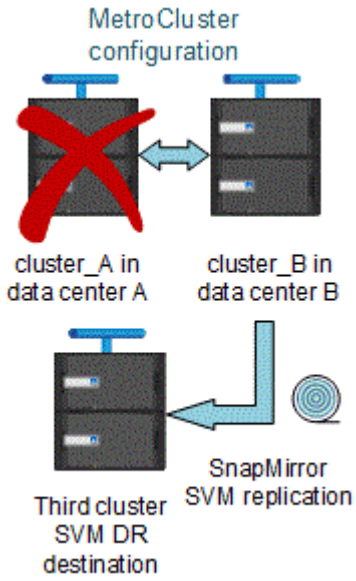
- When a MetroCluster configuration is in a steady state, the MetroCluster sync-destination SVM cannot be the source of an SVM disaster recovery relationship, since the volumes are not online.

  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.

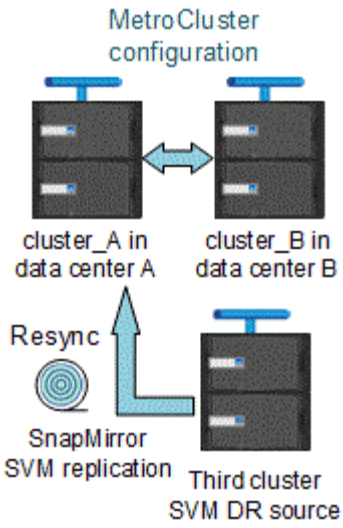
  However, after the switchover or switchback process completes, the next SVM DR scheduled updates will succeed.

See “Replicating the SVM configuration” in Data protection for details on configuring an SVM DR relationship.

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.

> In ONTAP versions earlier than 9.0, the NVFAIL flag is used for each switchover. In ONTAP 9.0 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
   ```

Where to find additional information

You can learn more about MetroCluster configuration.

MetroCluster and miscellaneous information

Information	Subject
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
Stretch MetroCluster installation and configuration	• Stretch MetroCluster architecture • Cabling the configuration • Configuring the FC-to-SAS bridges • Configuring the MetroCluster in ONTAP
MetroCluster management	• Understanding the MetroCluster configuration • Switchover, healing, and switchback
Disaster Recovery	• Disaster recovery • Forced switchover • Recovery from a multi-controller or storage failure

MetroCluster Maintenance	• Guidelines for maintenance in a MetroCluster FC configuration • Hardware replacement or upgrade and firmware upgrade procedures for FC-to-SAS bridges and FC switches • Hot-adding a disk shelf in a fabric-attached or stretch MetroCluster FC configuration • Hot-removing a disk shelf in a fabric-attached or stretch MetroCluster FC configuration • Replacing hardware at a disaster site in a fabric-attached or stretch MetroCluster FC configuration • Expanding a two-node fabric-attached or stretch MetroCluster FC configuration to a four-node MetroCluster configuration. • Expanding a four-node fabric-attached or stretch MetroCluster FC configuration to an eight-node MetroCluster FC configuration.
MetroCluster Upgrade and Expansion	• Upgrading or refreshing a MetroCluster configuration • Expanding a MetroCluster configuration by adding additional nodes
MetroCluster Transition	• Transitioning from a MetroCluster FC configuration to a MetroCluster IP configuration
MetroCluster Upgrade, Transition, and Expansion	• Monitoring the MetroCluster configuration with the MetroCluster Tiebreaker software
AFF and FAS Documentation **Note:** The standard storage shelf maintenance procedures can be used with MetroCluster IP configurations.	• Hot-adding a disk shelf • Hot-removing a disk shelf
Copy-based transition	• Transitioning data from 7-Mode storage systems to clustered storage systems
ONTAP concepts	• How mirrored aggregates work