ONTAP® 9

MetroCluster® IP Installation and Configuration Guide

June 2018 | 215-12877_D0
doccomments@netapp.com

NetApp®

Contents

Deciding whether to use the MetroCluster IP Installation and Configuration Guide

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

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

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

General information about ONTAP and MetroCluster configurations is also available.

ONTAP 9 Documentation Center

Preparing for the MetroCluster installation

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

Differences between the ONTAP MetroCluster configurations

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

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

Feature	IP configurations	Fabric-attached configurations		Stretch configurations	
		Four- or eight-node	Two-node	Two-node bridge-attached	Two-node direct-attached
Number of controllers	Four	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 (starting in ONTAP 9.4)	No	No	No	No
Supports local HA	Yes	Yes	No	No	No
Supports automatic switchover	No	Yes	Yes	Yes	Yes
Supports unmirrored aggregates	No	Yes	Yes	Yes	Yes
Supports array LUNs	No	Yes	Yes	Yes	Yes

Access to remote storage in MetroCluster IP configurations

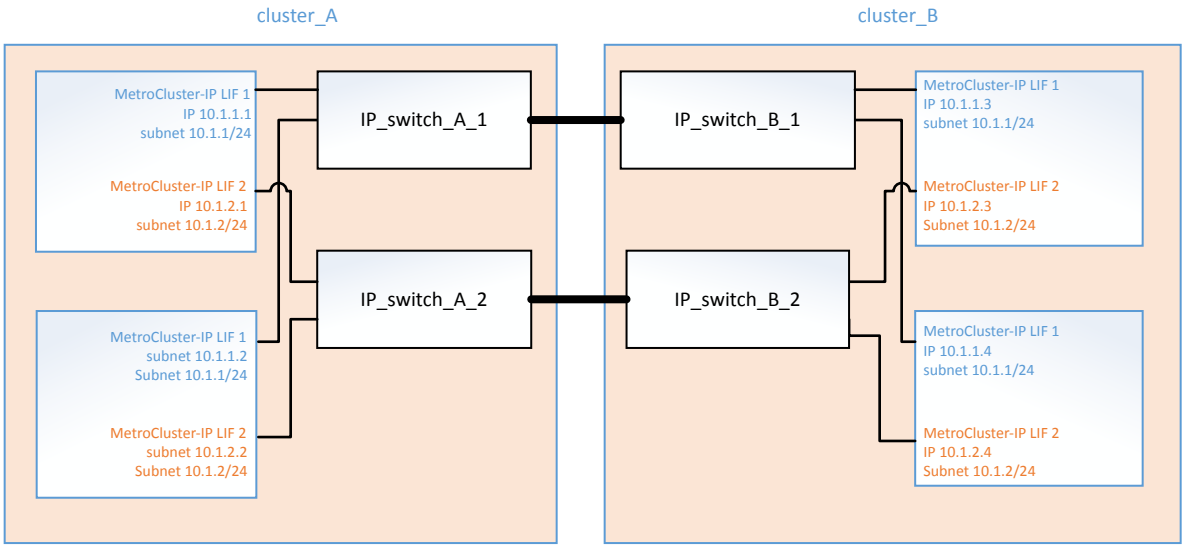
In MetroCluster IP configurations, the only way the local controllers can reach the remote storage pools is via the remote controllers. The IP switches are connected to the Ethernet ports on the controllers; they do not have direct connections to the disk shelves. If the remote controller is down, the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

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

Considerations for MetroCluster IP configuration

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

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



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

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

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

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

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

Characteristics of MetroCluster IP interfaces

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

- They are created by the `metrocluster configuration-settings interface create` command as part the initial MetroCluster configuration.
  They are not created or modified by the `network interface` commands.
- They do not appear in the output of the `network interface show` command.
- They do not fail over, but remain associated with the port on which they were created.
- MetroCluster IP configurations use a 40/100-Gbps Ethernet adapter for the IP interfaces to the IP switches used for the MetroCluster IP fabric.
  - In MetroCluster IP configurations on AFF A700 and FAS9000 systems, the X91146A-C 40/100-Gbps Ethernet adapter is required in slot 5 of each controller module.
  - In MetroCluster IP configurations on AFF A800 systems, the X1146A 40/100-Gbps Ethernet adapter is required in slot 0 and slot 1.

Considerations for ADP systems in ONTAP 9.4

Starting with ONTAP 9.4, MetroCluster IP configurations support new installations with AFF systems using ADP (Advanced Drive Partitioning).

ONTAP 9.4 includes 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.

Supported configurations for automatic drive assignment

The following table describes the supported configurations for automatic drive assignment and partitioning.

Platform	Drive shelf arrangement	Assignment rules
AFF A700 systems	Four external shelves	Drives are automatically assigned on a shelf-by-shelf basis.
AFF A800 systems	Internal drives only	The internal partitions are divided into four equal groups (quarters). Each quarter is automatically assigned to a separate pool.
	Internal drives and four external shelves	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 configurations.

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 manual assignment of a shelf works

Automatic drive assignment does not occur on ADP 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 shelves per site, but the total number of 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. If the configuration has two storage shelves at each site, you would assign half the drives on each shelf to a different pool.

Related concepts

Required MetroCluster IP components and naming conventions on page 20

Related information

Disk and aggregate management

Considerations for configuring cluster peering

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

Related information

Cluster and SVM peering express configuration

Prerequisites for cluster peering

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

Connectivity requirements

Intercluster LIFs must have *pair-wise full-mesh connectivity:* Every intercluster LIF on the local cluster must be able to communicate with every intercluster LIF on the remote cluster.

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

- The subnet must have enough IP addresses available to allocate to one intercluster LIF per node.
  For example, in a six-node cluster, the subnet used for intercluster communication must have six available IP addresses.

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

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

Port requirements

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

- All ports that are used to communicate with a given remote cluster must be in the same IPspace.
  You can use multiple IPspaces to peer with multiple clusters. Pair-wise full-mesh connectivity is required only within an IPspace.
- The broadcast domain that is used for intercluster communication must include at least two ports per node so that intercluster communication can fail over from one port to another port.
  Ports added to a broadcast domain can be physical network ports, VLANs, or interface groups (ifgrps).
- All ports must be cabled.
- All ports must be in a healthy state.
- The MTU settings of the ports must be consistent.

Firewall requirements

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

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

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.

Cluster requirements

Clusters must meet the following requirements:

- The time on the clusters in a cluster peering relationship must be synchronized within 300 seconds (5 minutes).
  Cluster peers can be in different time zones.

Considerations when using dedicated ports

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

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

- If the amount of available WAN bandwidth is similar to that of the LAN ports and the replication interval is such that replication occurs while regular client activity exists, then you should dedicate Ethernet ports for intercluster replication to avoid contention between replication and the data protocols.
- If the network utilization generated by the data protocols (CIFS, NFS, and iSCSI) is such that the network utilization is above 50 percent, then you should dedicate ports for replication to allow for nondegraded performance if a node failover occurs.
- When physical 10 GbE or faster ports are used for data and replication, you can create VLAN ports for replication and dedicate the logical ports for intercluster replication.
  The bandwidth of the port is shared between all VLANs and the base port.
- Consider the data change rate and replication interval and whether the amount of data that must be replicated on each interval requires enough bandwidth that it might cause contention with data protocols if sharing data ports.

Considerations when sharing data ports

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

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

- For a high-speed network, such as a 40-Gigabit Ethernet (40-GbE) network, a sufficient amount of local LAN bandwidth might be available to perform replication on the same 40-GbE ports that are used for data access.
  In many cases, the available WAN bandwidth is far less than 10 GbE LAN bandwidth.
- All nodes in the cluster might have to replicate data and share the available WAN bandwidth, making data port sharing more acceptable.
- Sharing ports for data and replication eliminates the extra port counts required to dedicate ports for replication.
- The maximum transmission unit (MTU) size of the replication network will be the same size as that used on the data network.

- Consider the data change rate and replication interval and whether the amount of data that must be replicated on each interval requires enough bandwidth that it might cause contention with data protocols if sharing data ports.
- When data ports for intercluster replication are shared, the intercluster LIFs can be migrated to any other intercluster-capable port on the same node to control the specific data port that is used for replication.

Preconfigured settings for new MetroCluster systems from the factory

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

Hardware racking and cabling

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

Software configuration of the MetroCluster configuration

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

Hardware setup checklist

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

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

Step	Completed at factory	Completed by you
Cable the components within the cabinet.	Yes Cables 5 meters and longer are removed for shipping and placed in the accessories box.	No
Connect the cables between cabinets, if applicable.	No	Yes Cables are in the accessories box.
Connect management cables to the customer's network.	No	Yes Connect them directly or through the CN1601 management switches, if present. **Attention:** To avoid address conflicts, do not connect management ports to the customer's network until after you change the default IP addresses to the customer's values.
Connect console ports to the customer's terminal server, if applicable.	No	Yes
Connect the customer's data cables to the cluster.	No	Yes
Connect the cabinets to power and power on the components.	No	Yes Power them on in the following order: **1.** PDUs **2.** Disk shelves **3.** Nodes
Assign IP addresses to the management ports of the cluster switches and to the management ports of the management switches, if present.	No	Yes Connect to the serial console port of each switch and log in with user name “admin” with no password. Suggested management addresses are 10.10.10.81, 10.10.10.82, 10.10.10.83, and 10.10.10.84.
Verify cabling by running the Config Advisor tool.	No	Yes

Using the Interoperability Matrix Tool to find MetroCluster information

When setting up the MetroCluster configuration, you can use the Interoperability Tool to ensure you are using supported software and hardware versions.

NetApp Interoperability Matrix Tool

After opening the Interoperability Matrix, you can use the Storage Solution field to select your MetroCluster solution.

You use the **Component Explorer** to select the components and ONTAP version to refine your search.

You can click **Show Results** to display the list of supported configurations that match the criteria.

Configuring the MetroCluster hardware components

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

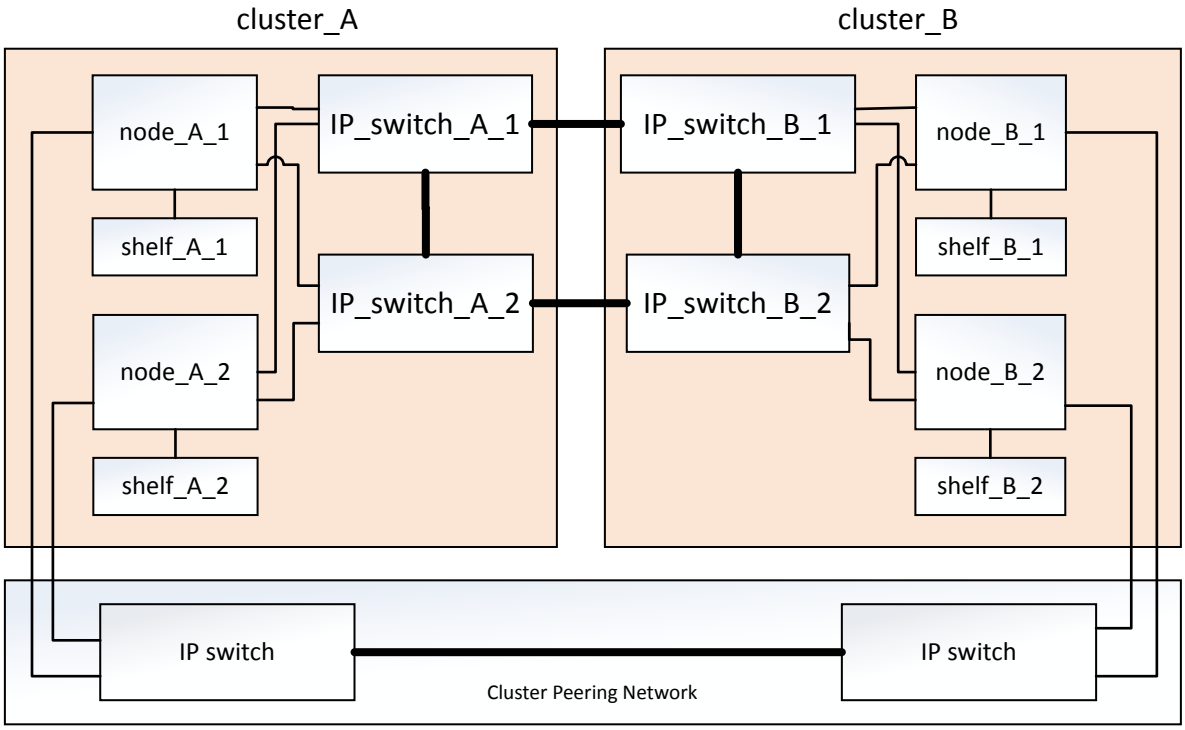
Parts of a MetroCluster IP configuration

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

Key hardware elements

A MetroCluster IP configuration includes the following key hardware elements:

- Storage controllers
  The storage controllers are configured as two two-node clusters.
- IP network
  This back-end IP network provides connectivity for two distinct uses:
  - Standard cluster connectivity for intra-cluster communications.
    This is the same cluster switch functionality used in non-MetroCluster switched ONTAP clusters.
  - MetroCluster back-end connectivity for replication of storage data and non-volatile cache.
- Cluster peering network
  The cluster peering network provides connectivity for mirroring of the cluster configuration, which includes storage virtual machine (SVM) configuration. The configuration of all of the SVMs on one cluster is mirrored to the partner cluster.



Disaster Recovery (DR) groups

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

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

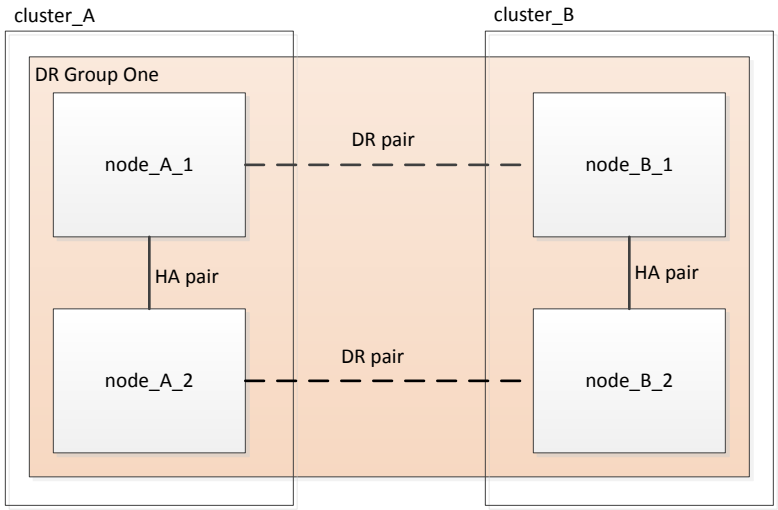
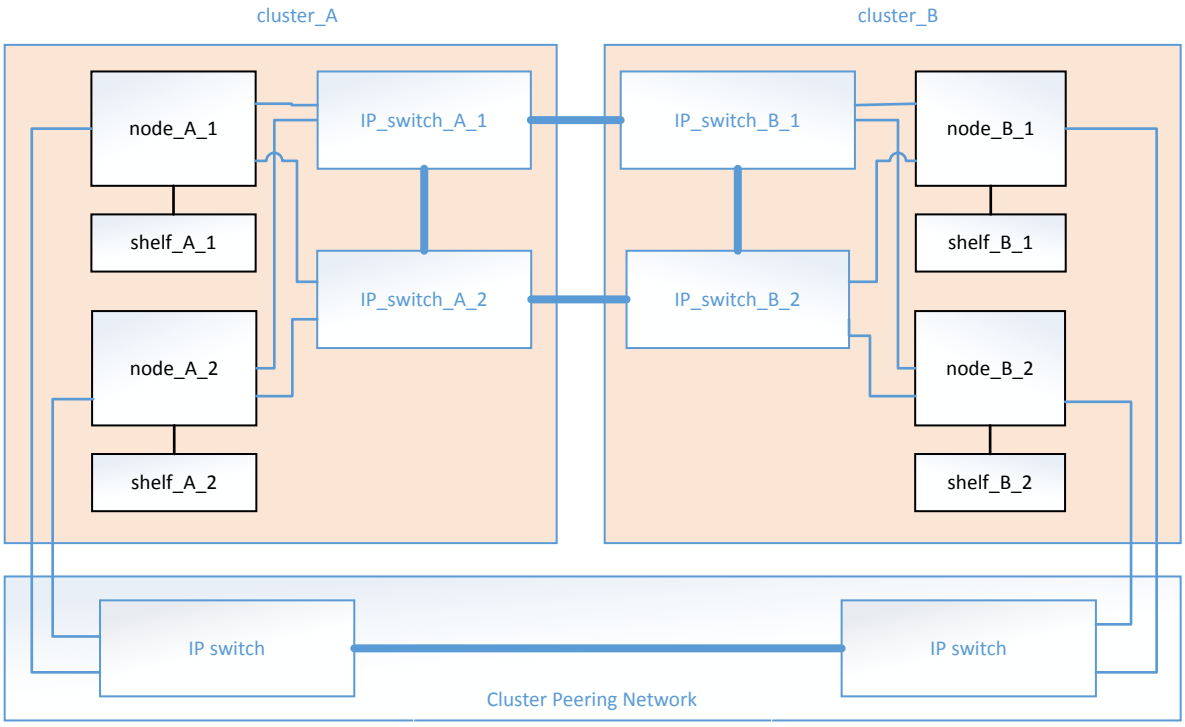


Illustration of the local HA pairs in a MetroCluster configuration

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

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

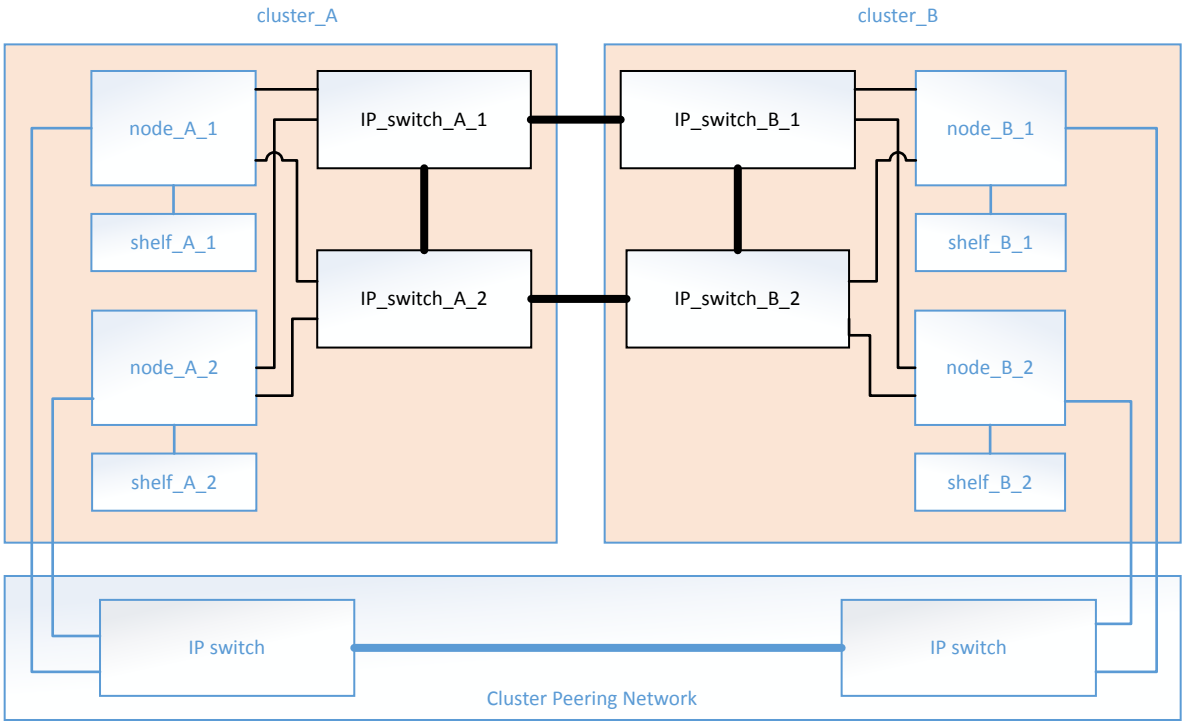


Related information

ONTAP concepts

Illustration of the MetroCluster IP and cluster interconnect network

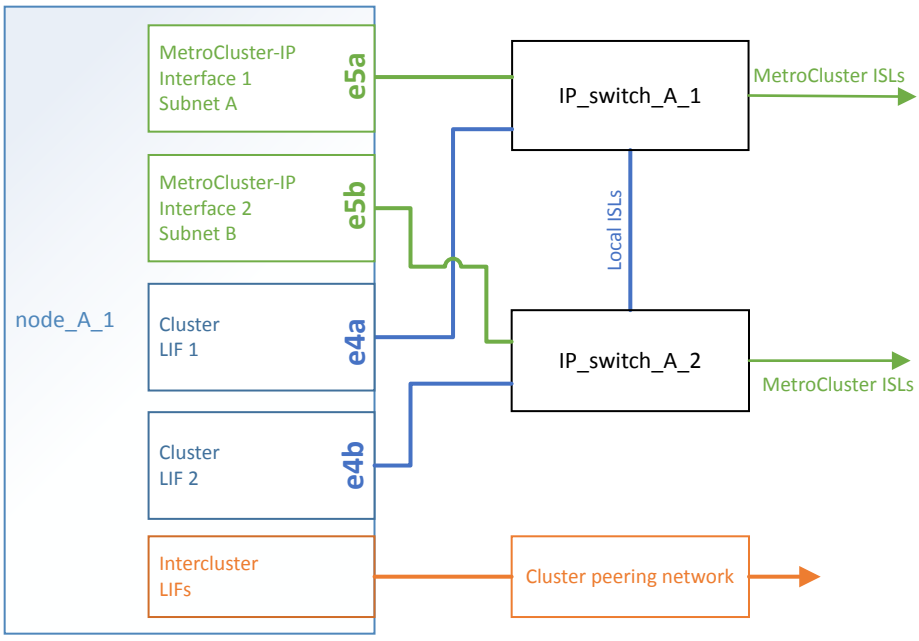
ONTAP clusters typically include a cluster interconnect network for traffic between the nodes in the cluster. In MetroCluster IP configurations, this network is also used for carrying data replication traffic between the MetroCluster sites.



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.



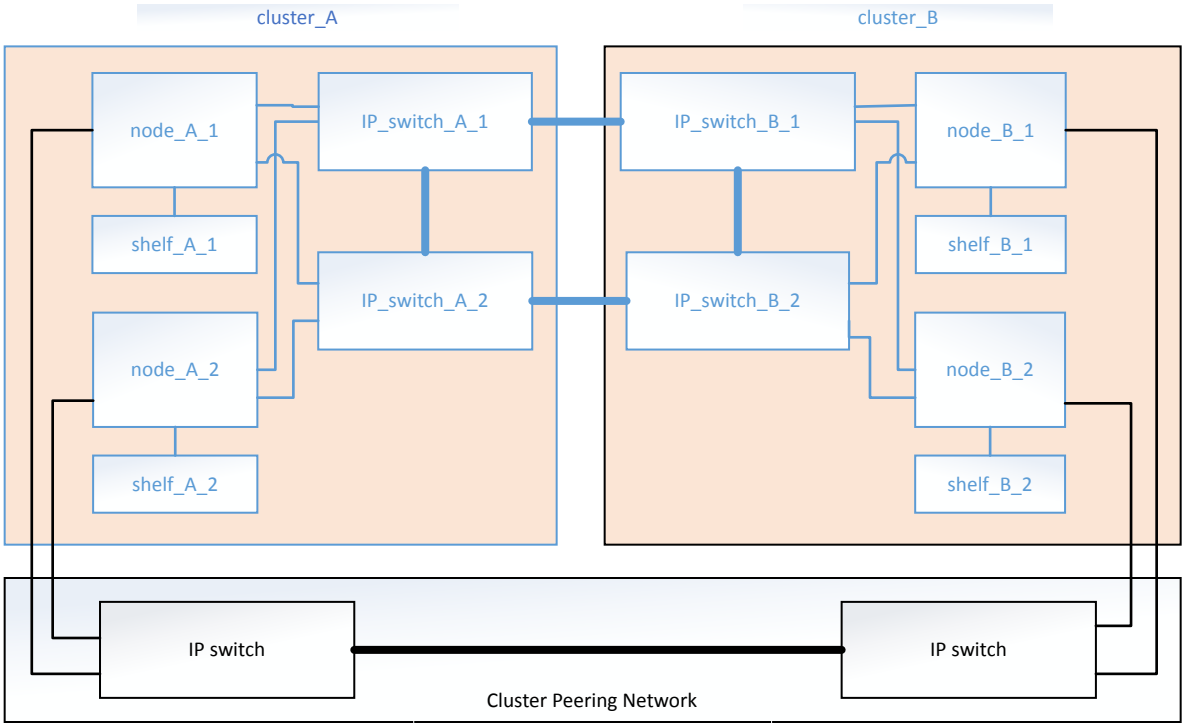
Related concepts

Considerations for MetroCluster IP configuration on page 7

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 concepts

Considerations for configuring cluster peering on page 10

Related tasks

Cabling the cluster peering connections on page 26
Peering the clusters on page 50

Related information

Cluster and SVM peering express configuration

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. For example, one site is referred to as Site A and the other site is referred to as Site B.

Supported software and hardware

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

NetApp Interoperability Matrix Tool

In the IMT, you can use the Storage Solution field to select your MetroCluster solution. You use the **Component Explorer** to select the components and ONTAP version to refine your search. You can click **Show Results** to display the list of supported configurations that match the criteria.

NetApp Hardware Universe

When using All Flash Optimized systems, all controller modules in the MetroCluster configuration must be configured as All Flash Optimized 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 cluster communication among the controller modules in each cluster.

Naming must be unique within the MetroCluster configuration.

Example names:

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

Controller module requirements in a MetroCluster IP configuration

MetroCluster IP configurations require four controller modules.

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

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

- 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.
- Starting with ONTAP 9.4, controller modules configured for ADP are supported.

Example names:

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

40/100-Gbps Ethernet adapter requirements in a MetroCluster IP configuration

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

- In MetroCluster IP configurations on AFF A700 and FAS9000 systems, the X91146A-C 40/100-Gbps Ethernet adapter is required in slot 5 of each controller module.
- In MetroCluster IP configurations on AFF A800 systems, the X1146A 40/100-Gbps Ethernet adapter is required in slot 0 and slot 1.

SAS disk shelf requirements in a MetroCluster IP configuration

Eight SAS disk shelves are recommended (four shelves at each site) to allow disk ownership on a per-shelf basis. A minimum of four disk shelves is required (two shelves at each site).

Shelf IDs must be unique within the MetroCluster IP configuration.

Example names:

- Site A:
  - site_A-shelf_1
  - site_A-shelf_2
  - site_A-shelf_3
  - site_A-shelf_4
- Site B:
  - site_B-shelf_1
  - site_B-shelf_2
  - site_B-shelf_3
  - site_B-shelf_4

Drive location considerations for AFF A800 internal drives

For correct implementation of the ADP feature, theAFF A800 system's 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 quadrants:

- Quadrant one:
  - Bays 0 - 5
  - Bays 24 - 29
- Quadrant two:

- Bays 6 - 11
- Bays 30 - 35

- Quadrant three:
  - Bays 12 - 17
  - Bays 36 - 41
- Quadrant 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: 12, 13, 14, 15
- Four drives in the third quarter: 24, 25, 26, 27
- Four drives in the fourth quarter: 36, 37, 38, 39

Related concepts

Considerations for ADP systems in ONTAP 9.4 on page 8

Installing and cabling MetroCluster components

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

Steps

1. Racking the hardware components on page 23
2. Cabling the IP switches on page 24
3. Cabling the cluster peering connections on page 26
4. Cabling the management and data connections on page 27
5. Configuring the IP switches on page 27

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.

   Each AFF A700 or FAS9000 controller module must have a X91146A-C 40/100-Gbps Ethernet adapter in slot 5.

   Each AFF A800 controller module must have a X1146A 40/100-Gbps Ethernet adapter in slot 0 and slot 1.

   AFF A700 and FAS9000 Installation and Setup Instructions

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

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

   NetApp Documentation: Disk Shelves

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

Cabling the IP switches

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

About this task

- This task must be repeated for each switch in the MetroCluster configuration.
- The port usage shown applies to both Cisco 3232 and Cisco 3132 switches.
- The controller port usage depends on the model of the controller:
  - The AFF A700 and FAS9000 systems use the X91146A-C 40-Gbps Ethernet adapter in slot 5 of each controller.
  - The AFF A800 systems use two X91146A-C 40-Gbps Ethernet adapters in each controller, one in slot 0 and one in slot 1.

Steps

1. Cable the switches to the local nodes.

   The node ports used depend on the platform model.

Site A: IP_switch_A_1 Local interconnect connections				
Switch port	**Node**	**Port**		**Usage**
		AFF A800	**AFF A700 and FAS9000 systems**	
1	node_A_1	e0a	e4a	Local cluster interconnect
2	node_A_2	e0a	e4a	Local cluster interconnect
3	-	-	-	Unused
4	-	-	-	Unused
5	-	-	-	Unused
6	-	-	-	Unused

Site A: IP_switch_A_1 Local interconnect connections				
Switch port	**Node**	**Port**		**Usage**
		AFF A800	**AFF A700 and FAS9000 systems**	
-	-	-	-	
9	node_A_1	e0b	e5a	MetroCluster IP interconnect
10	node_A_2	e0b	e5a	MetroCluster IP interconnect

Site A: IP_switch_A_2 Local interconnect connections				
Switch port	**Node**	**Port**		**Usage**
		AFF A800 systems	**AFF A700 and FAS9000 systems**	
1	node_A_1	e1a	e4b	Local cluster interconnect
2	node_A_2	e1a	e4b	Local cluster interconnect
3	-	-	-	Unused
4	-	-	-	Unused
5	-	-	-	Unused
6	-	-	-	Unused
-	-	-	-	
9	node_A_1	e1b	e5b	MetroCluster IP interconnect
10	node_A_2	e1b	e5b	MetroCluster IP interconnect

2. Cable the switch ISL connections.

   One, two or three 40-Gbps ISLs are supported or up to six 10-Gbps MetroCluster ISLs.

   If using the Cisco 3232C switch in breakout mode, ports 21 - 24 are used as MetroCluster ISLs. In this case these 40-Gbps ports are split into four 10-Gbps ports. You must be using the correct RCF files to support the breakout configuration.

   The switch cannot be configured with both 40-Gbps and 10-Gbps ports.

Site A: IP_switch_A_1 ISL connections		
Switch port	**Switch**	**Usage**
7	IP_switch_A_2	Local cluster ISL
8	IP_switch_A_2	Local cluster ISL
9 - 14	-	-
15	IP_switch_B_1	MetroCluster ISL
16	IP_switch_B_1	MetroCluster ISL

Site A: IP_switch_A_1 ISL connections		
Switch port	**Switch**	**Usage**
17	IP_switch_B_1	MetroCluster ISL
18	IP_switch_B_1	MetroCluster ISL
19	IP_switch_B_1	MetroCluster ISL
20	IP_switch_B_1	MetroCluster ISL
21	IP_switch_B_1	MetroCluster ISL (when using the Cisco 3232C switch in breakout mode).
22	IP_switch_B_1	
23	IP_switch_B_1	
24	IP_switch_B_1	

Site A: IP_switch_A_2 ISL connections		
Switch port	**Switch**	**Usage**
7	IP_switch_A_1	Local cluster ISL
8	IP_switch_A_1	Local cluster ISL
9 - 14	-	-
15	IP_switch_B_2	MetroCluster ISL
16	IP_switch_B_2	MetroCluster ISL
17	IP_switch_B_2	MetroCluster ISL
18	IP_switch_B_2	MetroCluster ISL
19	IP_switch_B_2	MetroCluster ISL
20	IP_switch_B_2	MetroCluster ISL
21		MetroCluster ISL (when using the Cisco 3232C switch in breakout mode).
22		
23		
24		

3. Repeat the previous steps on the partner site, using the same cabling.

Cabling the cluster peering connections

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

About this task

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

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

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

Step

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

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

   Cluster and SVM peering express configuration

Related concepts

Considerations for configuring cluster peering on page 10

Related information

Cluster and SVM peering express configuration

Cabling the management and data connections

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

About this task

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

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

Step

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

   AFF A700 and FAS9000 Installation and Setup Instructions

Configuring the IP switches

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

Steps

1. Copying the switch NX-OS software and RCF files to the MetroCluster IP switches on page 27
2. Installing the IP switch software on page 31

Copying the switch NX-OS software and RCF files to the MetroCluster IP switches

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

Before you begin

You need a transfer protocol, such as FTP, TFTP, SFTP, or SCP, to copy the files to the switches.

About this task

You must use the supported switch software version.

NetApp Interoperability Matrix Tool

Using the Interoperability Matrix Tool to find MetroCluster information on page 14

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	*switch-model*_RCF_v1.2-MetroCluster-IP-switch-A-1.txt
IP_switch_B_1	*switch-model*_RCF_v1.2-MetroCluster-IP-switch-B-1.txt
IP_switch_A_2	*switch-model*_RCF_v1.2-MetroCluster-IP-switch-A-2.txt
IP_switch_B_2	*switch-model*_RCF_v1.2-MetroCluster-IP-switch-B-2.txt

Steps

1. Download the MetroCluster IP RCF files:

   Cisco Cluster and Management Network Switch Reference Configuration File Download for MetroCluster IP

2. 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)

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

Example

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

```
After resetting the switch to factory defaults the configuration wizard should be entered
automatically:
All fields that need to be entered as marked with <value>

         ---- 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.
```

In the next set of prompts, you enter basic information 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 complete 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
  no 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#
```

3. Download the supported NX-OS software file.

NetApp Downloads: Cisco Ethernet Switch

4. Copy the switch software to the switch:

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

   Example

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

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

5. Copy the RCF files to the switches:

   copy sftp://root@*FTP-server-IP-address*/tftpboot/*RCF-filename* bootflash: vrf management

   a. Copy the RCF files to the first switch:

      copy sftp://root@*FTP-server-IP-address*/tftpboot/*RCF-filename* bootflash: vrf management

      Example

      In this example, the `NX3132_RCF_v1.2-MetroCluster-IP-IP_switch_A_1.txt` RCF file is copied from the SFTP server at *`10.10.99.99`* 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://root@10.10.99.99/tftpboot/NX3132_RCF_v1.2-MetroCluster-IP-
      IP_switch_A_1.txt bootflash: vrf management
      root@10.10.99.99's password: sundance
      sftp> progress
      Progress meter enabled
      sftp> get   /tftpboot/NX3132_RCF_v1.2-MetroCluster-IP-switch-A-1.txt  /bootflash/
      NX3132_RCF_v1.2-MetroCluster-IP-switch-A-1.txt
      Fetching /tftpboot/NX3132_RCF_v1.2-MetroCluster-IP-switch-A-1.txt to /bootflash/
      NX3132_RCF_v1.2-MetroCluster-IP-switch-A-1.txt
      /tftpboot/NX3132_RCF_v1.2-MetroCluster-IP-switch-A-1.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 correct RCF file to the corresponding switch.

6. Verify on each switch that the RCF and switch NX-OS files are present in each switch's `bootflash` directory:

   dir bootflash:

   Example

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

```
IP_switch_A_1# dir bootflash:
                   .
                   .
                   .
       5514    Jun 13 22:09:05 2017  NX3132_RCF_v1.2-MetroCluster-IP-
switch-A-1.txt
  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#
```

Installing the IP switch software

You must install the supported version of the switch NX-OS operating system.

About this task

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

Steps

1. Install the switch software:

   `install all nxos bootflash:nxos.`*`version-number`*`.bin`

   The switch will reload (reboot) automatically after the switch software has been installed.

Example

The following example shows the software installation on IP_switch_A_1:

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

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

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

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

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

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

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



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



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


Switch will be reloaded for disruptive upgrade.
```

```
Do you want to continue with the installation (y/n)?  [n] y


Install is in progress, please wait.

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

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

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

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


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

2. 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.
```

3. Verify that the switch software has been installed:

   `show version`

   The following example shows the output:

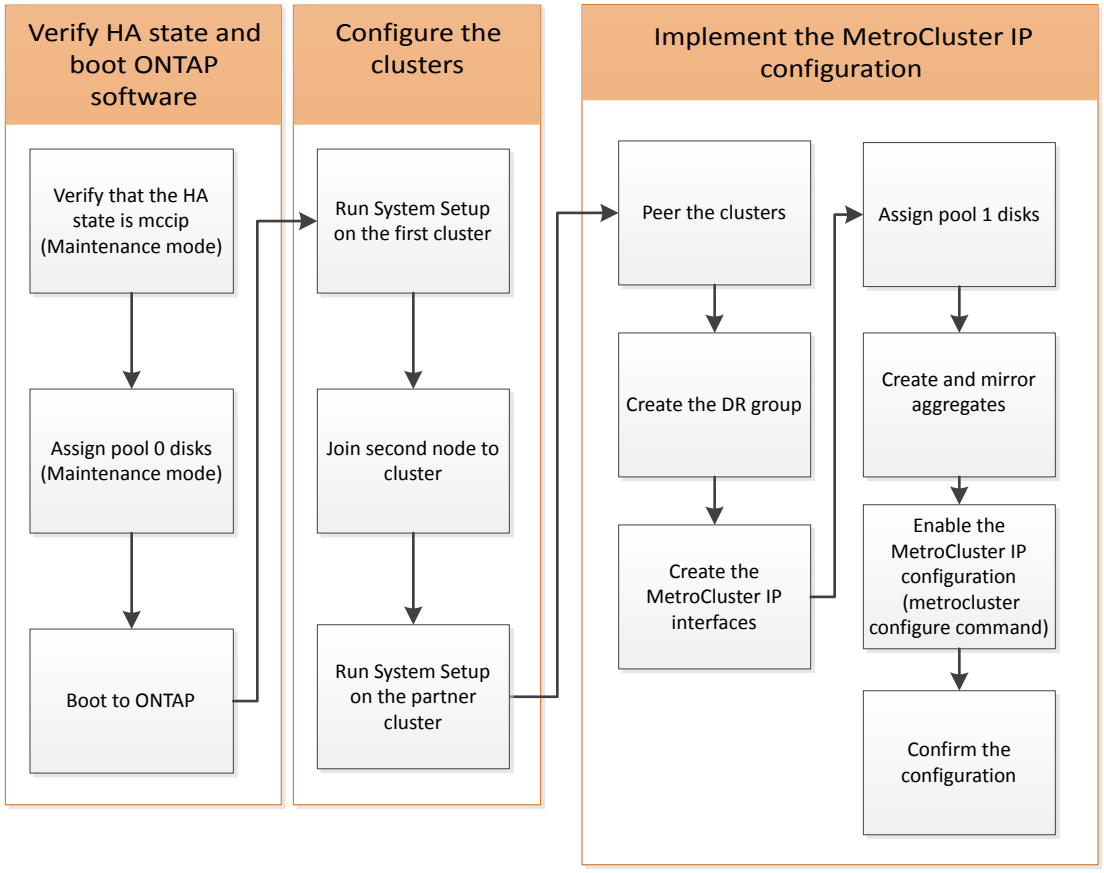
```
IP_switch_A_1# show version
Cisco Nexus Operating System (NX-OS) Software
TAC support: http://www.cisco.com/tac
Copyright (C) 2002-2017, Cisco and/or its affiliates.
All rights reserved.
.
.
.

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#
```

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

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.



Steps

1. Gathering required information on page 35
2. Similarities and differences between standard cluster and MetroCluster configurations on page 39
3. Restoring system defaults on a previously used controller module on page 40
4. Verifying the ha-config state of components on page 41
5. Manually assigning drives to pool 0 on page 41
6. Setting up ONTAP on page 45
7. Configuring the clusters into a MetroCluster configuration on page 49
8. Verifying switchover, healing, and switchback on page 75
9. Installing the MetroCluster Tiebreaker software on page 76
10. Protecting configuration backup files on page 76

Gathering required information

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

IP network information worksheet for site A

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

Site A switch information

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

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

Site A cluster creation information

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

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

Site A node information

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

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

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

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

Considerations for MetroCluster IP configuration on page 7

Node	Port	IP address of MetroCluster IP LIF	Network mask	Default gateway
Node 1 MetroCluster IP LIF 1	e5a			
Node 1 MetroCluster IP LIF 2	e5b			
Node 2 MetroCluster IP LIF 1	e5a			
Node 2 MetroCluster IP LIF 2	e5b			

Site A LIFs and ports for cluster peering

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

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

Site A time server information

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

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

Site A AutoSupport information

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

Type of information		Your values
From email address		
Mail hosts	IP addresses or names	

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

Site A SP information

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

Node	IP address	Network mask	Default gateway
Node 1			

IP network information worksheet for site B

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

Site B switch information

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

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

Site B cluster creation information

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

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

Site B node information

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

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

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

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

Considerations for MetroCluster IP configuration on page 7

Node	Port	IP address of MetroCluster IP LIF	Network mask	Default gateway
Node 1 MetroCluster IP LIF 1	e5a			
Node 1 MetroCluster IP LIF 2	e5b			
Node 2 MetroCluster IP LIF 1	e5a			
Node 2 MetroCluster IP LIF 2	e5b			

Site B LIFs and ports for cluster peering

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

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

Site B time server information

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

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

Site B AutoSupport information

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

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

Site B SP information

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

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

Similarities and differences between standard cluster and MetroCluster configurations

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

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

Configuration step	Standard cluster configuration	MetroCluster configuration
Configure management, cluster, and data LIFs on each node.	Same in both types of clusters	
Configure the root aggregate.	Same in both types of clusters	
Set up the cluster on one node in the cluster.	Same in both types of clusters	
Join the other node to the cluster.	Same in both types of clusters	
Create a mirrored root aggregate.	Optional	Required
Peer the clusters.	Optional	Required

Configuration step	Standard cluster configuration	MetroCluster configuration
Enable the MetroCluster configuration.	Does not apply	Required

Restoring system defaults on a previously used controller module

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

About this task

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

Steps

1. At the LOADER prompt, return the environmental variables to their default setting:

   `set-defaults`

2. Boot the node to the boot menu:

   `boot_ontap menu`

   After you run the 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.

     Note: This process is disruptive.

     The following screen shows the boot menu prompt:

```
Please choose one of the following:

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

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


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

- If your system is not configured for ADP, type `wipeconfig` at the boot menu prompt, and then press Enter.
  The following screen shows the boot menu prompt:

```
Please choose one of the following:

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

Verifying the ha-config state of components

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

Before you begin

The system must be in Maintenance mode.

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.

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.

Choices

- Manually assigning drives for pool 0 (ONTAP 9.4) on page 42
- Manually assigning drives for pool 0 (ONTAP 9.3) on page 43

Manually assigning drives for pool 0 (ONTAP 9.4)

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 disks.

About this task

This procedure applies to configurations running ONTAP 9.4.

Manual assignment is not needed if there are no external shelves and system is only using internal disks. This procedure is required depending on the number of shelves in the configuration:

- Fewer than four external shelves per site.
  The drives must be assigned manually to ensure symmetrical assignment of the drives, with each node having an equal number of drives.
- More than four shelves per site, but the total number of 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.

Considerations for ADP systems in ONTAP 9.4 on page 8

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.

   #### Example

   The following screen shows the boot menu prompt:

   ```
   Please choose one of the following:

       (1) Normal Boot.
       (2) Boot without /etc/rc.
       (3) Change password.
       (4) Clean configuration and initialize all disks.
       (5) Maintenance mode boot.
       (6) Update flash from backup config.
   ```

```
    (7) Install new software first.
    (8) Reboot node.
    (9) Configure Advanced Drive Partitioning.
    Selection (1-9)?  9a
########## WARNING ##########

    This is a disruptive operation and will result in the
    loss of all filesystem data. Before proceeding further,
    make sure that:
    1) This option (9a) has been executed or will be executed
    on the HA partner node (and DR/DR-AUX partner nodes if
    applicable), prior to reinitializing any system in the
    HA-pair (or MCC setup).
    2) The HA partner node (and DR/DR-AUX partner nodes if
    applicable) is currently waiting at the boot menu.

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

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. At the LOADER prompt, assign the root aggregate disks:

   setenv root-configuration "-d 6 -p 4 -s 2"

   This command assigns six data drives, four parity drives, and two spares.
7. Display the boot menu:

   boot_ontap menu
8. Select option 4 from the boot menu and let the system boot.
9. Repeat these steps on the other nodes in the MetroCluster IP configuration.
10. Proceed to *Setting up ONTAP* on page 45.

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. While the node is in Maintenance mode, you must first assign a single disk on the appropriate shelves to pool 0. ONTAP then automatically assign 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.

About this task

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.

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.

      #### Example

      The output should be similar to the following:

      ```
      Welcome to the cluster setup wizard.

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

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

          This system will send event messages and periodic reports to
      ```

```
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.

Example

The prompts are similar to the following:

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

c. Create the cluster by responding to the prompts.

Example

The prompts are similar to the following:

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


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

Existing cluster interface configuration found:

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

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

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

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

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

Retype the password:


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

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

Creating cluster cluster_A

Starting cluster support services ...

Cluster cluster_A has been created.
```

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

Example

The prompts are similar to the following:

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

Enter an additional license key []:


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


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

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

Enter the DNS domain names: 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.

Example

The prompts are similar to the following:

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

```
SFO will be enabled when the partner joins the cluster.


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

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

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

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

Cluster management using System Manager

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

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

5. Confirm that you have four ports configured as cluster interconnects:

network port show

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

Example

The following example shows two cluster ports on each node in cluster_A:

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


Node: cluster_A-01
                                                                        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
2 entries were displayed.
```

6. Repeat these steps on the partner cluster.

After you finish

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

Configuring the clusters into a MetroCluster configuration

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

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

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

About this task

Considerations for ADP systems in ONTAP 9.4 on page 8

Step

1. Disable automatic drive assignment:

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

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

Verifying drive assignment of pool 0 drives

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

About this task

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

Considerations for ADP systems in ONTAP 9.4 on page 8

Step

1. Verify that pool 0 drives are assigned automatically:

   `disk show`

Example

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.

Steps

1. Configuring intercluster LIFs on page 51
2. Creating a cluster peer relationship on page 55

Related concepts

Considerations when using dedicated ports on page 11

Considerations when sharing data ports on page 11

Related information

Cluster and SVM peering express configuration

Configuring intercluster LIFs

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

Choices

- Configuring intercluster LIFs on dedicated ports on page 51
- Configuring intercluster LIFs on shared data ports on page 53

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.

   #### Example

   The following example shows the network ports in **`cluster01`**:

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

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

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

   For complete command syntax, see the man page.

   #### Example

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

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

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

3. Create a failover group for the dedicated ports:

   network interface failover-groups create -vserver *system_SVM* -failover-group *failover_group* -targets *physical_or_logical_ports*

   Example

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

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

4. Verify that the failover group was created:

   network interface failover-groups show

   For complete command syntax, see the man page.

   Example

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

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

   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.

   Example

   The following example creates intercluster LIFs **cluster01_icl01** and **cluster01_icl02** in the failover group **intercluster01**:

   ```
   cluster01::> network interface create -vserver cluster01 -lif cluster01_icl01 -role
   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 -role
   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:

   network interface show –role intercluster

   For complete command syntax, see the man page.

Example

```
cluster01::> network interface show –role 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:

   `network interface show –role intercluster -failover`

   For complete command syntax, see the man page.

Example

The following example shows that the intercluster LIFs **`cluster01_icl01`** and **`cluster01_icl02`** on the SVM **`e0e`** port will fail over to the **`e0f`** port.

```
cluster01::> network interface show -role intercluster –failover
         Logical         Home                  Failover        Failover
Vserver  Interface       Node:Port             Policy          Group
-------- --------------- --------------------- --------------- --------
cluster01-01
         cluster01-01_icl01 cluster01-01:e0e   local-only      intercluster01
                            Failover Targets:  cluster01-01:e0e,
                                               cluster01-01:e0f
         cluster01-01_icl02 cluster01-02:e0e   local-only      intercluster01
                            Failover Targets:  cluster01-02:e0e,
                                               cluster01-02:e0f
```

Related concepts

Considerations when using dedicated ports on page 11

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.

Example

The following example shows the network ports in **`cluster01`**:

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

2. Create intercluster LIFs on the system SVM:

network interface create -vserver *system_SVM* -lif *LIF_name* -role intercluster -home-node *node* -home-port *port* -address *port_IP* -netmask *netmask*

For complete command syntax, see the man page.

Example

The following example creates intercluster LIFs **cluster01_icl01** and **cluster01_icl02**:

```
cluster01::> network interface create -vserver cluster01 -lif cluster01_icl01 -role
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 -role
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:

   network interface show –role intercluster

   For complete command syntax, see the man page.

Example

```
cluster01::> network interface show –role 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:

   network interface show –role intercluster -failover

   For complete command syntax, see the man page.

Example

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

```
cluster01::> network interface show -role 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 concepts

Considerations when sharing data ports on page 11

Creating a cluster peer relationship

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

Before you begin

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

Steps

1. On the data protection destination cluster, create a peer relationship with the data protection source cluster:

   `cluster peer create -generate-passphrase -offer-expiration` *`MM/DD/YYYY HH:MM:SS`*`|1...7days|1...168hours -peer-addrs` *`peer_LIF_IPs`* `-initial-allowed-vserver-peers` *`svm_name`*`,..|* -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.

   Note: If you use a custom IPspace, you cannot later move the relationship into the default IPspace.

   #### Example

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

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

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

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

2. On the data protection source cluster, authenticate the source cluster to the destination cluster:

   `cluster peer create -peer-addrs` *`peer_LIF_IPs`* `-ipspace` *`ipspace`*

   For complete command syntax, see the man page.

   #### Example

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

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

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

   Enter the passphrase:
   ```

```
Confirm the passphrase:

Clusters cluster02 and cluster01 are peered.
```

Enter the passphrase for the peer relationship when prompted.

3. Verify that the cluster peer relationship was created:

   `cluster peer show -instance`

   Example

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

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

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

   `cluster peer health show`

   Example

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

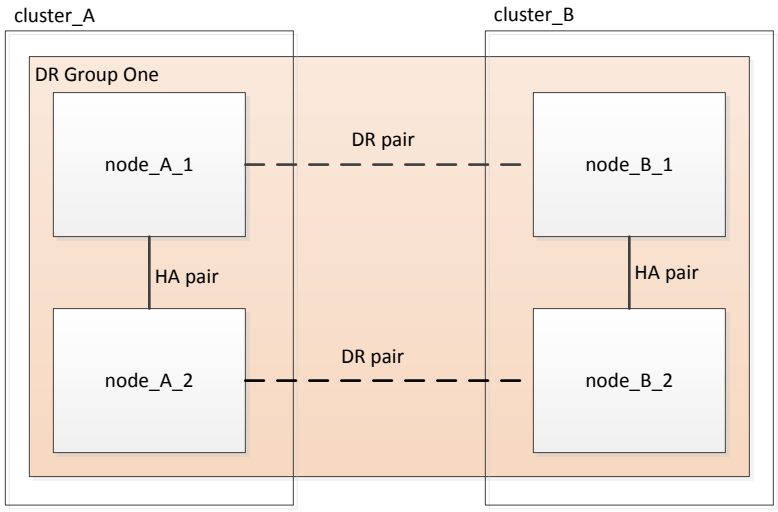
Creating the DR group

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

About this task

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

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



Steps

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

   `metrocluster configuration-settings show-status`

   #### Example

   The command output should show that the nodes are ready:

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

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

2. Create the DR group:

   `metrocluster configuration-settings dr-group create -partner-cluster` ***`partner-cluster-name`*** **`-local-node`** ***`local-node-name`*** **`-remote-node`** ***`remote-node-name`***

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

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

   The following command creates these DR pairs:

- node_A_1 and node_B_1
- node_A_2 and node_B_2

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

Configuring and connecting the MetroCluster IP interfaces

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

About this task

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

Considerations for MetroCluster IP configuration on page 7

You must create two interfaces for each node.

In the examples the following IP addresses and subnets are used:

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 port usage in the following examples used is for an AFF A700 or FAS9000 system. AFF A800 systems use ports e0b and e1b for the MetroCluster IP interfaces.

Steps

1. If you have at least two shelves connected to each node, confirm that each node has disk autoassignment enabled:

   `storage disk option show`

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

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

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

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

metrocluster configuration-settings show-status

Example

All nodes should be ready:

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

3. Create the interfaces on node_A_1.

   a. Configure the interface on port e5a on node_A_1:

   metrocluster configuration-settings interface create -cluster-name *cluster-name* -home-node *node-name* -home-port e5a -address *ip-address* -netmask 255.255.255.0

   Example

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

   ```
   cluster_A::> metrocluster configuration-settings interface create -
   cluster-name cluster_A -home-node node_A_1 -home-port 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 255.255.255.0

   Example

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

   ```
   cluster_A::> metrocluster configuration-settings interface create -
   cluster-name cluster_A -home-node node_A_1 -home-port e5b -address
   10.1.2.1 -netmask 255.255.255.0
   [Job 28] Job succeeded: Interface Create is successful.
   cluster_A::>
   ```

4. Create the interfaces on node_A_2.

   a. Configure the interface on port e5a on node_A_2:

   metrocluster configuration-settings interface create -cluster-name *cluster-name* -home-node *node-name* -home-port e5a -address *ip-address* -netmask 255.255.255.0

Example

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

```
cluster_A::> metrocluster configuration-settings interface create -
cluster-name cluster_A -home-node node_A_2 -home-port e5a -address
10.1.1.2 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

b. Configure the interface on port e5b on node_A_2:

metrocluster configuration-settings interface create -cluster-name *cluster-name* -home-node *node-name* -home-port e5b -address *ip-address* -netmask 255.255.255.0

Example

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

```
cluster_A::> metrocluster configuration-settings interface create -
cluster-name cluster_A -home-node node_A_2 -home-port e5b -address
10.1.2.2 -netmask 255.255.255.0
[Job 28] Job succeeded: Interface Create is successful.
cluster_A::>
```

5. Create the interfaces on node_B_1.

a. Configure the interface on port e5a on node_B_1:

metrocluster configuration-settings interface create -cluster-name *cluster-name* -home-node *node-name* -home-port e5a -address *ip-address* -netmask 255.255.255.0

Example

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

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

b. Configure the interface on port e5b on node_B_1:

metrocluster configuration-settings interface create -cluster-name *cluster-name* -home-node *node-name* -home-port e5b -address *ip-address* -netmask 255.255.255.0

Example

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

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

6. Create the interfaces on node_B_2.

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 255.255.255.0

Example

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

```
cluster_A::> metrocluster configuration-settings interface create -
cluster-name cluster_A -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 255.255.255.0

Example

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

```
cluster_A::> metrocluster configuration-settings interface create -
cluster-name cluster_A -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

Example

```
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

Example

All nodes should be ready:

```
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.

   Example

   The following example shows cluster_A is successfully connected:

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

10. Verify that the connections have been established:

    metrocluster configuration-settings show-status

    Example

    The configuration settings status for all nodes should be completed:

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

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

       Example

       Four MCCIP initiators should be present on each cluster:

```
cluster_A::*> storage iscsi-initiator show

                                                                                Status

Node Type Label    Target Portal      Target Name                       Admin/Op
---- ---- -------- ------------------ --------------------------------- --------

cluster_A-01
     dr_auxiliary
          mccip-aux-a-initiator
                    10.1.1.4:65200     iqn.2016-06.com.netapp:f33b1e3a-23ab-11e8-
b7db-00a098cf668e
                                                                          up/up
          mccip-aux-b-initiator
                    10.1.2.4:65200     iqn.2016-06.com.netapp:f33b1e3a-23ab-11e8-
b7db-00a098cf668e
                                                                          up/up
     dr_partner
          mccip-pri-a-initiator
                    10.1.1.3:65200     iqn.
2016-06.com.netapp:f8c47084-23ab-11e8-8d8b-00a098cf6698
                                                                          up/up
          mccip-pri-b-initiator
                    10.1.2.3:65200     iqn.
2016-06.com.netapp:f8c47084-23ab-11e8-8d8b-00a098cf6698
                                                                          up/up
cluster_A-02
     dr_auxiliary
          mccip-aux-a-initiator
                    10.1.1.3:65200     iqn.
2016-06.com.netapp:f8c47084-23ab-11e8-8d8b-00a098cf6698
                                                                          up/up


                                                                          Status

Node Type Label    Target Portal      Target Name                       Admin/Op
---- ---- -------- ------------------ --------------------------------- --------
cluster_A-02
     dr_auxiliary
          mccip-aux-b-initiator
                    10.1.2.3:65200     iqn.
2016-06.com.netapp:f8c47084-23ab-11e8-8d8b-00a098cf6698
                                                                          up/up
     dr_partner
          mccip-pri-a-initiator
                    10.1.1.4:65200     iqn.2016-06.com.netapp:f33b1e3a-23ab-11e8-
b7db-00a098cf668e
                                                                          up/up
          mccip-pri-b-initiator
                    10.1.2.4:65200     iqn.2016-06.com.netapp:f33b1e3a-23ab-11e8-
b7db-00a098cf668e
                                                                          up/up
8 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`

Example

```
cluster_A::> metrocluster node show
DR                           Configuration  DR
Group Cluster Node           State          Mirroring Mode
----- ------- ------------------ -------------- --------- ----
-     cluster_A
              node_A_1           ready to configure -     -
              node_A_2           ready to configure -     -
2 entries were displayed.
cluster_A::>
```

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

Verifying or manually performing pool 1 drives assignment

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

About this task

Configuration type	Procedure
The 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* on page 64
The configuration does not include four storage shelves per site and is running ONTAP 9.4	*Manually assigning disks for pool 1 (ONTAP 9.4)* on page 66
The systems were not received from the factory and are running ONTAP 9.3 Systems received from the factory are pre-configured with assigned drives.	*Manually assigning disks for pool 1 (ONTAP 9.3)* on page 68

Verifying disk assignment for pool 1 disks

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

Before you begin

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

About this task

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

Considerations for ADP systems in ONTAP 9.4 on page 8

Step

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.

   #### Example

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

```
cluster_B::> disk show -host-adapter 0m -owner node_B_2
                     Usable     Disk              Container   Container
Disk                 Size       Shelf Bay Type    Container   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 disks for pool 1 (ONTAP 9.4)

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 remote pool 1 disks.

About this task

This procedure applies to configurations running ONTAP 9.4.

Manual assignment is not needed if there are no external shelves and the system is only using internal disks. This procedure is required depending on the number of shelves in the configuration:

- Fewer than four external shelves per site.
  The drives must be assigned manually to ensure symmetrical assignment of the drives, with each node having an equal number of drives.
- More than four shelves per site, but the total number of 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 the configuration includes only two external shelves per site, pool 1 disks for each site should be shared from same shelf. For example:

- node_A_1 is assigned disks in bays 0-11 on site_B-shelf_2 (remote)
- node_A_2 is assigned disks in bays 12-23 on site_B-shelf_2 (remote)

Considerations for ADP systems in ONTAP 9.4 on page 8

Step

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

   a. Display the list of unassigned disks:

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

   Example

```
cluster_A::> disk show -host-adapter 0m -container-type unassigned
                     Usable           Disk    Container   Container
Disk                   Size Shelf Bay Type    Type        Name      Owner
---------------- ---------- ----- --- ------- ----------- --------- --------
6.23.0                    -    23   0 SSD     unassigned  -         -
6.23.1                    -    23   1 SSD     unassigned  -         -
.
.
.
node_A_2:0m.i1.2L51       -    21  14 SSD     unassigned  -         -
node_A_2:0m.i1.2L64       -    21  10 SSD     unassigned  -         -
.
.
.
48 entries were displayed.

cluster_A::>
```

   b. Assign ownership of remote disks (0m) to pool1 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`*.

   c. Confirm that the disks have been assigned to pool 1:

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

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

   Example

   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.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.

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. You must first assign a disk on the shelf to pool1. 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 autoassignment of disks.

If you cannot use shelf-level autoassignment, 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 autoassigned to pool1 of node_A_1
- All the disks on site_B-shelf_4 are autoassigned to pool1 of node_A_2
- All the disks on site_A-shelf_2 are autoassigned to pool1 of node_B_1
- All the disks on site_A-shelf_4 are autoassigned to pool1 of node_B_2

You must "seed" the autoassignment by specifying a single disk on each shelf.

Step

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`

   Example

```
cluster_A::> disk show -host-adapter 0m -container-type unassigned
                     Usable           Disk    Container   Container
Disk                   Size Shelf Bay Type    Type
Name      Owner
---------------- ---------- ----- --- ------- -----------
--------- --------
6.23.0                   -    23   0 SSD     unassigned
-         -
6.23.1                   -    23   1 SSD     unassigned
-         -
.
.
.
node_A_2:0m.i1.2L51      -    21  14 SSD     unassigned
-         -
node_A_2:0m.i1.2L64      -    21  10 SSD     unassigned
-         -
.
.
.
48 entries were displayed.

cluster_A::>
```

b. Select a remote disk (0m) and assign ownership of the disk to pool1 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 autoassignment feature assigns all disks on the remote shelf that contains the specified disk.

c. After waiting at least 60 seconds for disk autoassignment 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

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

Example

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

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

About this task

Considerations for ADP systems in ONTAP 9.4 on page 8

Step

1. Enable automatic drive assignment:

storage disk option modify -node *node_name* -autoassign on

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

Mirroring the root aggregates

You must mirror the root aggregates to provide data protection.

About this task

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

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

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

Steps

1. Mirror the root aggregate:

   `storage aggregate mirror aggr_name`

   #### Example

   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.

Before you begin

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

About this task

- Drives are owned by a specific node; when you create an aggregate, all drives in that aggregate must be owned by the same node, which becomes the home node for that aggregate.
  In systems using ADP, aggregates are created using partitions in which each drive is partitioned in to P1, P2 and P3 partitions.
- Aggregate names should conform to the naming scheme you determined when you planned your MetroCluster configuration.
  Disk and aggregate management

Steps

1. Display a list of available spares:

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

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

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

   You can specify the following options:

   - Aggregate's home node (that is, the node that owns the aggregate in normal operation)
   - List of specific drives that are to be added to the aggregate
   - Number of drives to include
   - Checksum style to use for the aggregate
   - Type of drives to use
   - Size of drives to use
   - Drive speed to use
   - RAID type for RAID groups on the aggregate
   - Maximum number of drives that can be included in a RAID group
   - Whether drives with different RPM are allowed

   For more information about these options, see the `storage aggregate create` man page.

   #### Example

   The following command creates a mirrored aggregate with 10 disks:

   ```
   cluster_A::> storage aggregate create aggr1_node_A_1 -diskcount 10 -
   node node_A_1 -mirror true
   [Job 15] Job is queued: Create aggr1_node_A_1.
   [Job 15] The job is starting.
   [Job 15] Job succeeded: DONE
   ```

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

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

Implementing the MetroCluster configuration

You must run the `metrocluster configure` command to start data protection in a MetroCluster configuration.

Before you begin

- There should be at least two non-root mirrored data aggregates on each cluster.
  You can verify this with the `storage aggregate show` command.

  Note: If you want to use a single mirrored data aggregate, then see *step 1* on page 72 for instructions.

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

About this task

You issue the `metrocluster configure` command once, on any of the nodes, to enable the MetroCluster configuration. You do not need to issue the command on each of the sites or nodes, and it does not matter which node or site you choose to issue the command on.

The `metrocluster configure` command automatically pairs the two nodes with the lowest system IDs in each of the two clusters as disaster recovery (DR) partners. In a four-node MetroCluster configuration, there are two DR partner pairs. The second DR pair is created from the two nodes with higher system IDs.

Steps

1. Configure the MetroCluster in the following format:

If your MetroCluster configuration has...	Then do this...
Multiple data aggregates **Note:** The best practice is to have multiple data aggregates.	From any node's prompt, configure MetroCluster: **`metrocluster configure -refresh true`** ***`node-name`***
A single mirrored data aggregate **Note:** If the first DrGroup has only one aggregate and you want to add a DrGroup with one aggregate, you need to move the metadata volume off the single data aggregate. For more information on this procedure, see *Moving a metadata volume in MetroCluster configurations*.	**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`**

Example

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`

Example

The following example shows the network port usage on a four-node MetroCluster configuration:

```
cluster_A::> network port show
                                                        Speed (Mbps)
Node   Port      IPspace   Broadcast Domain Link   MTU    Admin/Oper
------ --------- --------- ---------------- ----- ------- ------------
controller_A_1
       e0a       Cluster   Cluster          up     9000  auto/1000
       e0b       Cluster   Cluster          up     9000  auto/1000
```

```
       e0c       Default   Default          up     1500  auto/1000
       e0d       Default   Default          up     1500  auto/1000
       e0e       Default   Default          up     1500  auto/1000
       e0f       Default   Default          up     1500  auto/1000
       e0g       Default   Default          up     1500  auto/1000
controller_A_2
       e0a       Cluster   Cluster          up     9000  auto/1000
       e0b       Cluster   Cluster          up     9000  auto/1000
       e0c       Default   Default          up     1500  auto/1000
       e0d       Default   Default          up     1500  auto/1000
       e0e       Default   Default          up     1500  auto/1000
       e0f       Default   Default          up     1500  auto/1000
       e0g       Default   Default          up     1500  auto/1000
14 entries were displayed.
```

3. Verify the MetroCluster configuration from both sites in the MetroCluster configuration.

   a. Verify the configuration from site A:

      `metrocluster show`

      Example

      ```
      cluster_A::> metrocluster show

      Configuration: IP fabric

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

   b. Verify the configuration from site B:

      `metrocluster show`

      Example

      ```
      cluster_B::> metrocluster show

      Configuration: IP fabric

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

Checking the MetroCluster configuration

You can check that the components and relationships in the MetroCluster configuration are working correctly. You should do a check after initial configuration and after making any changes to the MetroCluster configuration. You should also do a check before a negotiated (planned) switchover or a switchback operation.

About this task

If the `metrocluster check run` command is issued twice within a short time on either or both clusters, a conflict can occur and the command might not collect all data. Subsequent `metrocluster check show` commands do not show the expected output.

Steps

1. Check the configuration:

   `metrocluster check run`

   #### Example

   The command runs as a background job and might not be completed immediately.

   ```
   cluster_A::> metrocluster check run
   The operation has been started and is running in the background. Wait
   for
   it to complete and run "metrocluster check show" to view the results.
   To
   check the status of the running metrocluster check operation, use the
   command,
   "metrocluster operation history show -job-id 2245"
   ```

   ```
   cluster_A::> metrocluster check show
   Last Checked On: 9/13/2017 20:41:37

   Component           Result
   ------------------- ---------
   nodes               ok
   lifs                ok
   config-replication  ok
   aggregates          ok
   clusters            ok
   5 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.

   #### Example

   The following example shows the `metrocluster check aggregate show` command output for a healthy four-node MetroCluster configuration:

   ```
   cluster_A::> metrocluster check aggregate show

   Last Checked On: 8/5/2014 00:42:58

   Node                Aggregate                Check                    Result
   --------------      -------------------      ---------------------    ---------
   controller_A_1      controller_A_1_aggr0
                                                mirroring-status         ok
                                                disk-pool-allocation     ok
                                                ownership-state          ok
                       controller_A_1_aggr1
                                                mirroring-status         ok
                                                disk-pool-allocation     ok
                                                ownership-state          ok
                       controller_A_1_aggr2
                                                mirroring-status         ok
                                                disk-pool-allocation     ok
                                                ownership-state          ok
   ```

```
controller_A_2       controller_A_2_aggr0
                                         mirroring-status           ok
                                         disk-pool-allocation       ok
                                         ownership-state            ok
                     controller_A_2_aggr1
                                         mirroring-status           ok
                                         disk-pool-allocation       ok
                                         ownership-state            ok
                     controller_A_2_aggr2
                                         mirroring-status           ok
                                         disk-pool-allocation       ok
                                         ownership-state            ok

18 entries were displayed.
```

The following example shows the `metrocluster check cluster show` command output for a healthy four-node MetroCluster configuration. It indicates that the clusters are ready to perform a negotiated switchover if necessary.

```
Last Checked On: 9/13/2017 20:47:04

Cluster               Check                          Result
--------------------- ------------------------------ ---------
mccint-fas9000-0102
                      negotiated-switchover-ready    not-applicable
                      switchback-ready               not-applicable
                      job-schedules                  ok
                      licenses                       ok
                      periodic-check-enabled         ok
mccint-fas9000-0304
                      negotiated-switchover-ready    not-applicable
                      switchback-ready               not-applicable
                      job-schedules                  ok
                      licenses                       ok
                      periodic-check-enabled         ok
10 entries were displayed.
```

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

Installing the MetroCluster Tiebreaker software

You can download and install Tiebreaker software to monitor the two clusters and the connectivity status between them from a third site. Doing so enables each partner in a cluster to distinguish between an ISL failure (when inter-site links are down) and a site failure.

Before you begin

You must have a Linux host available that has network connectivity to both clusters in the MetroCluster configuration.

Steps

1. Go to MetroCluster Tiebreaker Software Download page.

   NetApp Downloads: MetroCluster Tiebreaker for Linux

2. Follow the directions to download the Tiebreaker software and 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 System Administration Guide contains additional information under the section *Managing configuration backups*.

   System administration

Related information

System administration

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 from where the command is issued should switch over to the partner site.
- Nodes at the partner site should shut down gracefully and remain at the LOADER> prompt.

Steps

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

   `metrocluster node show`

   Example

   ```
   cluster_A::>  metrocluster node show

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

2. Begin the switchover operation:

   `metrocluster switchover`

   Example

   ```
   cluster_A::> metrocluster switchover
   Warning: negotiated switchover is about to start. It will stop all
   the data Vservers on cluster "cluster_B" and
   automatically re-start them on cluster "cluster_A". It will finally
   gracefully shutdown cluster "cluster_B".
   ```

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

   `metrocluster node show`

   Example

   ```
   cluster_A::>  metrocluster node show

   Cluster                       Configuration State    Mode
   ----------------------------- ---------------------- ------------------------
   ```

```
Local: cluster_A               configured            switchover
Remote: cluster_B              not-reachable         -
             configured            normal
```

4. Confirm that the switchover operation was successful:

   `metrocluster operation show`

   Example

```
cluster_A::>  metrocluster operation show

cluster_A::> metrocluster operation show
  Operation: switchover
      State: successful
 Start Time: 2/6/2016 13:28:50
   End Time: 2/6/2016 13:29:41
     Errors: -
```

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.

Steps

1. Heal the data aggregate:

   `metrocluster heal aggregates`

   Example

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

2. Heal the root aggregate:

   `metrocluster heal root-aggregates`

   Example

```
cluster_A::> metrocluster heal root-ggregates
[Job 937] Job succeeded: Heal Root Aggregates is successful.
```

3. Verify that healing is completed:

   `metrocluster node show`

Example

```
cluster_A::> metrocluster node show
DR                               Configuration  DR
Group Cluster Node               State          Mirroring Mode
----- ------- ------------------ -------------- ---------
--------------------
1     cluster_A
              node_A_1         configured     enabled   heal roots
completed
      cluster_B
              node_B_2         unreachable    -         switched over
42 entries were displayed.metrocluster operation show
```

4. Verify that all aggregates are mirrored:

   `storage aggregate show`

   Example

```
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. Boot nodes from the disaster site.

6. Check the status of switchback recovery:

   `metrocluster node show`

   Example

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

```
1     cluster_A
              node_A_1            configured     enabled   heal roots
completed
      cluster_B
              node_B_2            configured     enabled   waiting for
switchback
                                                            recovery
2 entries were displayed.
```

7. Perform the switchback:

   metrocluster switchback

   Example

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

8. Confirm status of the nodes:

   metrocluster node show

   Example

```
cluster_A::> metrocluster node show
DR                              Configuration  DR
Group Cluster Node              State          Mirroring Mode
----- ------- ----------------- -------------- ---------
--------------------
1     cluster_A
               node_A_1         configured     enabled   normal
      cluster_B
               node_B_2         configured     enabled   normal

2 entries were displayed.
```

9. Confirm status of the metrocluster operation:

   metrocluster operation show

   Example

   The output should show a successful state.

```
cluster_A::> metrocluster operation show
  Operation: switchback
      State: successful
 Start Time: 2/6/2016 13:54:25
   End Time: 2/6/2016 13:56:15
     Errors: -
```

Verifying operation after power line disruption

You can test the MetroCluster configuration's response to the failure of a PDU.

About this task

The best practice is for each PSU in a component to be connected to separate power supplies. If both PSUs are connected to the same 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 bridge is switched off.
- No failover or loss of service should occur.
- Only one path from the controller module to the drives behind the bridge is available.

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 by using the `system environment sensors show -state fault` and `storage shelf show -errors` commands.

   #### Example

   ```
   cluster_A::> system environment sensors show -state fault

   Node Sensor             State Value/Units Crit-Low Warn-Low Warn-Hi
   Crit-Hi
   ---- -------------------- ------ ----------- -------- --------
   ------- -------
   node_A_1
           PSU1             fault
                               PSU_OFF
           PSU1 Pwr In OK     fault
                               FAULT
   node_A_2
           PSU1             fault
                               PSU_OFF
           PSU1 Pwr In OK     fault
                               FAULT
   4 entries were displayed.

   cluster_A::> storage shelf show -errors
       Shelf Name: 1.1
        Shelf UID: 50:0a:09:80:03:6c:44:d5
    Serial Number: SHFHU1443000059

   Error Type          Description
   ------------------  --------------------------
   Power               Critical condition is detected in storage shelf
   power supply unit "1". The unit might fail.Reconnect PSU1
   ```

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.

Considerations when removing MetroCluster configurations

After removing the MetroCluster configuration, all disk connectivity and interconnects should be adjusted to be in a supported state. If you need to remove the MetroCluster configuration, contact technical support.

Attention: You cannot reverse the MetroCluster unconfiguration. This process should only be done with the assistance of technical support.

Requirements and limitations when using ONTAP in a MetroCluster configuration

When using ONTAP in a MetroCluster configuration, you should be aware of certain requirements and limitations for licensing, peering to clusters outside the MetroCluster configuration, performing volume operations, NVAIL operations, and other ONTAP operations.

- Both sites should be licensed for the same site-licensed features.
- All nodes should be licensed for the same node-locked features.

Job schedules in a MetroCluster configuration

User-created job schedules are automatically replicated between clusters in a MetroCluster configuration. If you create, modify, or delete a job schedule on a cluster, the same schedule is automatically created on the partner cluster, using Configuration Replication Service (CRS).

Note: System-created schedules are not replicated and you must manually perform the same operation on the partner cluster so that job schedules on both clusters are identical.

Cluster peering from the MetroCluster site to a third cluster

Because the peering configuration is not replicated, if you peer one of the clusters in the MetroCluster configuration to a third cluster outside of that configuration, you must also configure the peering on the partner MetroCluster cluster. This is so that peering can be maintained if a switchover occurs.

The non-MetroCluster cluster must be running ONTAP 8.3 or later. If not, peering is lost if a switchover occurs even if the peering has been configured on both MetroCluster partners.

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.

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 concepts

IPspace object replication and subnet configuration requirements on page 84
Requirements for LIF creation in a MetroCluster configuration on page 85
LIF replication and placement requirements and issues on page 85

Related information

Network and LIF management

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 concepts

Requirements for LIF creation in a MetroCluster configuration on page 85
LIF replication and placement requirements and issues on page 85

Requirements for LIF creation in a MetroCluster configuration

You should be aware of the requirements for creating LIFs when configuring your network in a MetroCluster configuration.

You must consider the following guidelines when creating LIFs:

- Fibre Channel: You must use stretched VSAN or stretched fabrics
- IP/iSCSI: You must use layer 2 stretched network
- ARP broadcasts: You must enable ARP broadcasts between the two clusters
- Duplicate LIFs: You must not create multiple LIFs with the same IP address (duplicate LIFs) in an IPspace

Verify LIF creation

You can confirm the successful creation of a LIF in a MetroCluster configuration by running the `metrocluster check lif show` command. If you encounter any issues while creating the LIF, you can use the `metrocluster check lif repair-placement` command to fix the issues.

Related concepts

IPspace object replication and subnet configuration requirements on page 84
LIF replication and placement requirements and issues on page 85

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.

Condition	LIF type: FC	LIF type: IP/iSCSI
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.

Condition	LIF type: FC	LIF type: IP/iSCSI
Port selection	ONTAP categorizes the ports based on attributes such as adapter type and speed, and then selects the ports with matching attributes. If no ports with matching attributes are found, the LIF is placed on a random connected port on the DR partner.	From the ports that are marked as reachable during the reachability check, ONTAP prefers ports that are in the broadcast domain that is associated with the subnet of the LIF. If there are no network ports available on the DR cluster that are in the broadcast domain that is associated with the subnet of the LIF, then ONTAP selects ports that have reachability to the source LIF. If there are no ports with reachability to the source LIF, a port is selected from the broadcast domain that is associated with the subnet of the source LIF, and if no such broadcast domain exists, a random port is selected. ONTAP categorizes the ports based on attributes such as adapter type, interface type, and speed, and then selects the ports with matching attributes.
LIF placement	From the reachable ports, ONTAP selects the least loaded port for placement.	From the selected ports, ONTAP selects the least loaded port for placement.

Placement of replicated LIFs when the DR partner node is down

When an iSCSI or FC LIF is created on a node whose DR partner has been taken over, the replicated LIF is placed on the DR auxiliary partner node. After a subsequent giveback operation, the LIFs are not automatically moved to the DR partner. This can lead to LIFs being concentrated on a single node in the partner cluster. During a MetroCluster switchover operation, subsequent attempts to map LUNs belonging to the storage virtual machine (SVM) fail.

You should run the `metrocluster check lif show` command after a takeover operation or giveback operation to verify that the LIF placement is correct. If errors exist, you can run the `metrocluster check lif repair-placement` command to resolve the issues.

LIF placement errors

LIF placement errors that are displayed by the `metrocluster check lif show` command are retained after a switchover operation. If the `network interface modify`, `network interface rename`, or `network interface delete` command is issued for a LIF with a placement error, the error is removed and does not appear in the output of the `metrocluster check lif show` command.

LIF replication failure

You can also check whether LIF replication was successful by using the `metrocluster check lif show` command. An EMS message is displayed if LIF replication fails.

You can correct a replication failure by running the `metrocluster check lif repair-placement` command for any LIF that fails to find a correct port. You should resolve any LIF replication failures as soon as possible to verify the availability of LIF during a MetroCluster switchover operation.

Note: Even if the source SVM is down, LIF placement might proceed normally if there is a LIF belonging to a different SVM in a port with the same IPspace and network in the destination SVM.

Related concepts

IPspace object replication and subnet configuration requirements on page 84
Requirements for LIF creation in a MetroCluster configuration on page 85

Output for the storage aggregate plex show command is indeterminate after a MetroCluster switchover

When you run the `storage aggregate plex show` command after a MetroCluster switchover, the status of plex0 of the switched over root aggregate is indeterminate and is displayed as `failed`. During this time, the switched over root is not updated. The actual status of this plex can only be determined after the MetroCluster healing phase.

Modifying volumes to set the NVFAIL flag in case of switchover

You can modify a volume so that the NVFAIL flag is set on the volume in the event of a MetroCluster switchover. The NVFAIL flag causes the volume to be fenced off from any modification. This is required for volumes that need to be handled as if committed writes to the volume were lost after the switchover.

About this task

Note: In ONTAP 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`

Monitoring and protecting the file system consistency using NVFAIL

The `-nvfail` parameter of the `volume modify` command enables ONTAP to detect nonvolatile RAM (NVRAM) inconsistencies when the system is booting or after a switchover operation. It also

warns you and protects the system against data access and modification until the volume can be manually recovered.

If ONTAP detects any problems, database or file system instances stop responding or shut down. ONTAP then sends error messages to the console to alert you to check the state of the database or file system. You can enable NVFAIL to warn database administrators of NVRAM inconsistencies among clustered nodes that can compromise database validity.

After the NVRAM data loss during failover or boot recovery, NFS clients cannot access data from any of the nodes until the NVFAIL state is cleared. CIFS clients are unaffected.

How NVFAIL impacts access to NFS volumes or LUNs

The NVFAIL state is set when ONTAP detects NVRAM errors when booting, when a MetroCluster switchover operation occurs, or during an HA takeover operation if the NVFAIL option is set on the volume. If no errors are detected at startup, the file service is started normally. However, if NVRAM errors are detected or NVFAIL processing is enforced on a disaster switchover, ONTAP stops database instances from responding.

When you enable the `NVFAIL` option, one of the processes described in the following table takes place during bootup:

If...	**Then...**
ONTAP detects no NVRAM errors	File service starts normally.
ONTAP detects NVRAM errors	• ONTAP returns a stale file handle (`ESTALE`) error to NFS clients trying to access the database, causing the application to stop responding, crash, or shut down. ONTAP then sends an error message to the system console and log file. • When the application restarts, files are available to CIFS clients even if you have not verified that they are valid. For NFS clients, files remain inaccessible until you reset the `in-nvfailed-state` option on the affected volume.
If one of the following parameters is used: • `dr-force-nvfail` volume option is set • `force-nvfail-all` switchover command option is set.	You can unset the `dr-force-nvfail` option after the switchover, if the administrator is not expecting to force NVFAIL processing for possible future disaster switchover operations. For NFS clients, files remain inaccessible until you reset the `in-nvfailed-state` option on the affected volume. **Note:** Using the `force-nvfail-all` option causes the `dr-force-nvfail` option to be set on all of the DR volumes processed during the disaster switchover.

If...	Then...
ONTAP detects NVRAM errors on a volume that contains LUNs	LUNs in that volume are brought offline. The `in-nvfailed-state` option on the volume must be cleared, and the `NVFAIL` attribute on the LUNs must be cleared by bringing each LUN in the affected volume online. You can perform the steps to check the integrity of the LUNs and recover the LUN from a Snapshot copy or back up as necessary. After all of the LUNs in the volume are recovered, the `in-nvfailed-state` option on the affected volume is cleared.

Commands for monitoring data loss events

If you enable the `NVFAIL` option, you receive notification when a system crash caused by NVRAM inconsistencies or a MetroCluster switchover occurs.

By default, the NVFAIL parameter is not enabled.

If you want to...	Use this command...
Create a new volume with NVFAIL enabled	`volume create -nvfail` **on**
Enable NVFAIL on an existing volume	`volume modify` **Note:** You set the `-nvfail` option to **on** to enable NVFAIL on the created volume.
Display whether NVFAIL is currently enabled for a specified volume	`volume show` **Note:** You set the `-fields` parameter to **`nvfail`** to display the NVFAIL attribute for a specified volume.

See the man page for each command for more information.

Accessing volumes in NVFAIL state after a switchover

After a switchover, you must clear the NVFAIL state by resetting the `-in-nvfailed-state` parameter of the `volume modify` command to remove the restriction of clients to access data.

Before you begin

The database or file system must not be running or trying to access the affected volume.

About this task

Setting `-in-nvfailed-state` parameter requires advanced-level privilege.

Step

1. Recover the volume by using the `volume modify` command with the `-in-nvfailed-state` parameter set to **`false`**.

After you finish

For instructions about examining database file validity, see the documentation for your specific database software.

If your database uses LUNs, review the steps to make the LUNs accessible to the host after an NVRAM failure.

Recovering LUNs in NVFAIL states after switchover

After a switchover, the host no longer has access to data on the LUNs that are in NVFAIL states. You must perform a number of actions before the database has access to the LUNs.

Before you begin

The database must not be running.

Steps

1. Clear the NVFAIL state on the affect volume that hosts the LUNs by resetting the `-in-nvfailed-state` parameter of the `volume modify` command.
2. Bring the affected LUNs online.
3. Examine the LUNs for any data inconsistencies and resolve them.

   This might involve host-based recovery or recovery done on the storage controller using SnapRestore.
4. Bring the database application online after recovering the LUNs.

Where to find additional information

You can learn more about MetroCluster IP configuration and operation from the NetApp documentation library.

MetroCluster and miscellaneous guides

Guide	Content
ONTAP 9 Documentation Center	• All MetroCluster guides
Fabric-attached MetroCluster installation and configuration	• Fabric-attached MetroCluster architecture • Cabling the configuration • Configuring the FC-to-SAS bridges • Configuring the FC switches • Configuring the MetroCluster in ONTAP
Stretch MetroCluster installation and configuration	• Stretch MetroCluster architecture • Cabling the configuration • Configuring the FC-to-SAS bridges • Configuring the MetroCluster in ONTAP
MetroCluster management and disaster recovery	• Understanding the MetroCluster configuration • Switchover, healing and switchback • Disaster recovery
MetroCluster Tiebreaker Software Installation and Configuration Guide	• Monitoring the MetroCluster configuration with the MetroCluster Tiebreaker software
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

Glossary of MetroCluster terms

aggregate

A grouping of physical storage resources (disks or array LUNs) that provides storage to volumes associated with the aggregate. Aggregates provide the ability to control the RAID configuration for all associated volumes.

data SVM

Formerly known as data Vserver. In clustered Data ONTAP, a Storage Virtual Machine (SVM) that facilitates data access from the cluster; the hardware and storage resources of the cluster are dynamically shared by data SVMs within a cluster.

admin SVM

Formerly known as admin Vserver. In clustered Data ONTAP, a Storage Virtual Machine (SVM) that has overall administrative access to all objects in the cluster, including all objects owned by other SVMs, but does not provide data access to clients or hosts.

inter-switch link (ISL)

A connection between two switches using the E-port.

destination

The storage to which source data is backed up, mirrored, or migrated.

disaster recovery (DR) group

The four nodes in a MetroCluster configuration that synchronously replicate each others' configuration and data.

disaster recovery (DR) partner

A node's partner at the remote MetroCluster site. The node mirrors its DR partner's NVRAM or NVMEM partition.

disaster recovery auxiliary (DR auxiliary) partner

The HA partner of a node's DR partner. The DR auxiliary partner mirrors a node's NVRAM or NVMEM partition in the event of an HA takeover after a MetroCluster switchover operation.

HA pair

- In Data ONTAP 8.x, a pair of nodes whose controllers are configured to serve data for each other if one of the two nodes stops functioning.

  Depending on the system model, both controllers can be in a single chassis, or one controller can be in one chassis and the other controller can be in a separate chassis.

- In the Data ONTAP 7.3 and 7.2 release families, this functionality is referred to as an *active/active configuration*.

HA partner

A node's partner within the local HA pair. The node mirrors its HA partner's NVRAM or NVMEM cache.

high availability (HA)

In Data ONTAP 8.x, the recovery capability provided by a pair of nodes (storage systems), called an *HA pair*, that are configured to serve data for each other if one of the two nodes stops functioning. In the Data ONTAP 7.3 and 7.2 release families, this functionality is referred to as an *active/active configuration*.

healing

The two required MetroCluster operations that prepare the storage located at the DR site for switchback. The first heal operation resynchronizes the mirrored plexes. The second heal operation returns ownership of root aggregates to the DR nodes.

LIF (logical interface)

A logical network interface, representing a network access point to a node. LIFs currently correspond to IP addresses, but could be implemented by any interconnect. A LIF is generally bound to a physical network port; that is, an Ethernet port. LIFs can fail over to other physical ports (potentially on other nodes) based on policies interpreted by the LIF manager.

NVRAM

nonvolatile random-access memory.

NVRAM cache

Nonvolatile RAM in a storage system, used for logging incoming write data and NFS requests. Improves system performance and prevents loss of data in case of a storage system or power failure.

NVRAM mirror

A synchronously updated copy of the contents of the storage system NVRAM (nonvolatile random access memory) contents kept on the partner storage system.

node

- In Data ONTAP, one of the systems in a cluster or an HA pair.
  To distinguish between the two nodes in an HA pair, one node is sometimes called the *local node* and the other node is sometimes called the *partner node* or *remote node*.
- In Protection Manager and Provisioning Manager, the set of storage containers (storage systems, aggregates, volumes, or qtrees) that are assigned to a dataset and designated either primary data (primary node), secondary data (secondary node), or tertiary data (tertiary node).
  A *dataset node* refers to any of the nodes configured for a dataset.
  A *backup node* refers to either a secondary or tertiary node that is the destination of a backup or mirror operation.
  A *disaster recovery node* refers to the dataset node that is the destination of a failover operation.

remote storage

The storage that is accessible to the local node, but is at the location of the remote node.

root volume

A special volume on each Data ONTAP system. The root volume contains system files and configuration information, and can also contain data. It is required for the system to be able to boot and to function properly. Core dump files, which are important for troubleshooting, are written to the root volume if there is enough space.

switchback

The MetroCluster operation that restores service back to one of the MetroCluster sites.

switchover

The MetroCluster operation that transfers service from one of the MetroCluster sites.

- A *negotiated* switchover is planned in advance and cleanly shuts down components of the target MetroCluster site.
- A *forced* switchover immediately transfers service; the shut down of the target site might not be clean.

Copyright information

Trademark information

How to send comments about documentation and receive update notifications

You can help us to improve the quality of our documentation by sending us your feedback. You can receive automatic notification when production-level (GA/FCS) documentation is initially released or important changes are made to existing production-level documents.

If you have suggestions for improving this document, send us your comments by email.

doccomments@netapp.com

To help us direct your comments to the correct division, include in the subject line the product name, version, and operating system.

If you want to be notified automatically when production-level documentation is released or important changes are made to existing production-level documents, follow Twitter account @NetAppDoc.

You can also contact us in the following ways:

- NetApp, Inc., 495 East Java Drive, Sunnyvale, CA 94089 U.S.
- Telephone: +1 (408) 822-6000
- Fax: +1 (408) 822-4501
- Support telephone: +1 (888) 463-8277

Index

A

C

D

E

F

G

H

I

L

M

N

O

P

R

S

T

U

V

W