Considerations for using MetroCluster-compliant switches

ONTAP MetroCluster

NetApp
July 26, 2024

Table of Contents

Considerations for using MetroCluster-compliant switches

Requirements and limitations when using MetroCluster-compliant switches

Beginning with ONTAP 9.7, MetroCluster IP configurations can use MetroCluster-compliant switches. These are switches that are not NetApp-validated but are compliant with NetApp specifications. However, NetApp does not provide troubleshooting or configuration support services for any non-validated switch. You should be aware of the general requirements and limitations when using MetroCluster-compliant switches.

General requirements for MetroCluster-compliant switches

The switch connecting the MetroCluster IP interfaces must meet the following general requirements:

- The switches must support quality of service (QoS) and traffic classification.
- The switches must support explicit congestion notification (ECN).
- The switches must support a load-balancing policy to preserve order along the path.
- The switches must support L2 Flow Control (L2FC).
- The switch port must provide a dedicated rate and must not be overallocated.
- The cables and transceivers connecting the nodes to the switches must be provided by NetApp. These cables must be supported by the switch vendor. If you are using optical cabling, the transceiver in the switch might not be provided by NetApp. You must verify that it is compatible with the transceiver in the controller.
- The switches connecting the MetroCluster nodes can carry non-MetroCluster traffic.
- Only platforms that provide dedicated ports for switchless cluster interconnects can be used with a MetroCluster-compliant switch. Platforms such as the FAS2750 and AFF A220 cannot be used because MetroCluster traffic and MetroCluster interconnect traffic share the same network ports.
- The MetroCluster-compliant switch must not be used for local cluster connections.
- The MetroCluster IP interface can be connected to any switch port that can be configured to meet the requirements.
- Four IP switches are required, two for each switch fabric. If you use directors, then you can use a single director at each side, but the MetroCluster IP interfaces must connect to two different blades in two different failure domains on that director.
- The MetroCluster interfaces from one node must connect to two network switches or blades. The MetroCluster interfaces from one node cannot be connected to the same network or switch or blade.
- The network must meet the requirements outlined in the following sections:
  - Considerations for ISLs
  - Considerations when deploying MetroCluster in shared layer 2 or layer 3 networks
- The maximum transmission unit (MTU) of 9216 must be configured on all switches that carry MetroCluster IP traffic.
- Reverting to ONTAP 9.6 or earlier is not supported.

Any intermediate switches that you use between the switches connecting the MetroCluster IP interfaces at both sites must meet the requirements and must be configured as outlined in Considerations when deploying MetroCluster in shared layer 2 or layer 3 networks.

Limitations when using MetroCluster-compliant switches

You cannot use any configuration or feature that requires that local cluster connections are connected to a switch. For example, you cannot use the following configurations and procedures with a MetroCluster-compliant switch:

- Eight-node MetroCluster configurations
- Transitioning from MetroCluster FC to MetroCluster IP configurations
- Refreshing a four-node MetroCluster IP configuration
- Platforms sharing a physical interface for local cluster and MetroCluster traffic. Refer to Platform-specific network speeds and switch port modes for MetroCluster-compliant switches for supported speeds.

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

If you are using MetroCluster compliant switches, you should be aware of the platform-specific network speeds and switch port mode requirements.

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

Missing values indicate that the platform cannot be used with a MetroCluster-compliant switch.

Platform	Network Speed (Gbps)	Switch port mode
FAS9500 AFF A900 ASA A900	100Gbps 40Gbps when upgrade PCM from FAS9000 / AFF A700	trunk mode
AFF C800 ASA C800 AFF A800 ASA A800	40Gbps or 100Gbps	access mode
FAS9000 AFF A700	40Gbps	access mode
FAS8300 AFF C400 ASA C400 AFF A400 ASA A400	40Gbps or 100Gbps	trunk mode
AFF A320	40Gbps or 100Gbps	access mode
FAS8200 AFF A300	25Gbps	access mode
FAS500f AFF C250 ASA C250 AFF A250 ASA A250	-	-
FAS2750 AFF A220	-	-
AFF A150 ASA A150	-	-
AFF A70	100Gbps	trunk mode
AFF A90	100Gbps	trunk mode
AFF A1K	100Gbps	trunk mode

Switch port configuration examples

Learn about the various switch port configurations.

The following examples use decimal values and follow the table that applies to Cisco switches. Depending on the switch vendor, you might require different values for DSCP. Refer to the corresponding table for your switch vendor to confirm the correct value.

DSCP value	Decimal	Hex	Meaning
101 000	16	0x10	CS2

011 000	24	0x18	CS3
100 000	32	0x20	CS4
101 000	40	0x28	CS5

Switch port connecting a MetroCluster interface

- Classification for remote direct memory access (RDMA) traffic:
  - Match : TCP port 10006, source, destination, or both
  - Optional match: COS 5
  - Optional match: DSCP 40
  - Set DSCP 40
  - Set COS 5
  - Optional : rate shaping to 20Gbps
- Classification for iSCSI traffic:
  - Match : TCP port 62500, source, destination, or both
  - Optional match: COS 4
  - Optional match: DSCP 32
  - Set DSCP 32
  - Set COS 4
- L2FlowControl (pause), RX and TX

ISL ports

- Classification:
  - Match COS 5 or DSCP 40
    - Set DSCP 40
    - Set COS 5
  - Match COS 4 or DSCP 32
    - Set DSCP 32
    - Set COS 4
- Egress queuing
  - COS group 4 has a minimum configuration threshold of 2000 and a maximum threshold of 3000
  - COS group 5 has a minimum configuration threshold of 3500 and a maximum threshold of 6500.

    Configuration thresholds can vary depending on the environment. You must evaluate the configuration thresholds based on your individual environment.

  - ECN enabled for Q4 and Q5
  - RED enabled for Q4 and Q5

Bandwidth allocation (switch ports connecting MetroCluster interfaces and ISL ports)

- RDMA, COS 5 / DSCP 40: 60%
- iSCSI, COS 4 / DSCP 32: 40%
- Minimum capacity requirement per MetroCluster configuration and network: 10Gbps

If you use rate limits, the traffic should be **shaped** without introducing loss.

Examples for configuring switch ports connecting the MetroCluster controller

The example commands provided are valid for Cisco NX3232 or Cisco NX9336 switches. Commands vary according to the switch type.

If a feature or its equivalent shown in the examples is not available on the switch, the switch does not meet the minimum requirements and cannot be used to deploy a MetroCluster configuration. This is true for any switch connecting to a MetroCluster configuration and for all intermediate switches.

The following examples might only show the configuration for one network.

Basic configuration

A virtual LAN (VLAN) in each network must be configured. The following example shows how to configure a VLAN in network 10.

Example:

```
# vlan 10
The load balancing policy should be set so that order is preserved.
```

Example:

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

Examples for configuring classification

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

In the following example, all TCP traffic to and from the port 65200 is mapped to the storage (iSCSI) class. All TCP traffic to and from the port 10006 is mapped to the RDMA class. These policy-maps are used on switch ports connecting the MetroCluster interfaces.

Example:

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

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

You must configure an ingress policy. An ingress policy maps the traffic as classified to different COS groups. In this example, the RDMA traffic is mapped to COS group 5 and iSCSI traffic is mapped to COS group 4. The ingress policy is used on switch ports connecting the MetroCluster interfaces and on the ISL ports carrying MetroCluster traffic.

Example:

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

NetApp recommends that you shape traffic on switch ports connecting a MetroCluster interface, as shown in the following example:

Example:

```
policy-map type queuing MetroClusterIP_Node_Egress
class type queuing c-out-8q-q7
  priority level 1
class type queuing c-out-8q-q6
  priority level 2
class type queuing c-out-8q-q5
  priority level 3
  shape min 0 gbps max 20 gbps
class type queuing c-out-8q-q4
  priority level 4
class type queuing c-out-8q-q3
  priority level 5
class type queuing c-out-8q-q2
  priority level 6
class type queuing c-out-8q-q1
  priority level 7
class type queuing c-out-8q-q-default
  bandwidth remaining percent 100
  random-detect threshold burst-optimized ecn
```

Examples for configuring the node ports

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

Example:

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

You might need to configure the MetroCluster interface port speed. The following example shows how to configure the speed to **auto** or into 40Gbps mode:

Example:

```
    speed auto

    speed 40000
```

The following example shows a switch port configured to connect a MetroCluster interface. It is an access mode port in VLAN 10, with an MTU of 9216 and is operating in native speed. It has symmetric (send and receive) flow control (pause) enabled and the MetroCluster ingress and egress policies assigned.

Example:

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

On 25Gbps ports, you might need to set the Forward Error Correction (FEC) setting to "off", as shown in the following example.

Example:

```
fec off
```

Examples of configuration of ISL ports throughout the network

A MetroCluster-compliant switch is regarded as an intermediate switch, even it directly connects the MetroCluster interfaces. The ISL ports carrying MetroCluster traffic on the MetroCluster-compliant switch must be configured the same way as the ISL ports on an intermediate switch. Refer to Required settings on intermediate switches for guidance and examples.

Some policy maps are the same for switch ports connecting MetroCluster interfaces and ISLs carrying MetroCluster traffic. You can use the same policy map for both of these port usages.

Copyright information

Trademark information