Technical Report

MetroCluster IP

Solution architecture and design

Stephen B. Galla, NetApp
May 2023 | TR-4689

Abstract

NetApp® MetroCluster is a continuously available storage solution for NetApp ONTAP® running on FAS and AFF systems. MetroCluster IP is the latest evolution that uses an Ethernet-based back-end storage fabric. MetroCluster IP provides a highly redundant configuration to meet the needs of the most critical business applications. Because MetroCluster IP is included in ONTAP, it does not require a separate license, and it provides NAS and SAN connectivity for clients and servers that use ONTAP storage.

TABLE OF CONTENTS

MetroCluster overview **4**

Continuous availability solution overview 4

MetroCluster IP compared to MetroCluster FC 5

MetroCluster IP architecture **5**

Disaster recovery group 7

Replication in MetroCluster IP 7

Network 8

Storage 10

Solution design **11**

Confirming support 11

Hardware components 11

Sizing a solution 15

Platform limits 17

Network configuration 17

Cluster fabric Inter-Switch Link (ISL) 19

MetroCluster IP Inter-Switch Link design 19

Operation and administration **24**

High Availability (HA) and Disaster Recovery (DR) 24

Quorum witness 25

Interoperability **28**

SnapMirror Asynchronous 28

NetApp ONTAP FlexGroup volumes 28

NetApp FlexCache 28

NetApp FabricPool 29

SVM Disaster Recovery (DR) 29

Where to find additional information **30**

Version history **31**

LIST OF TABLES

Table 1) MetroCluster IP controller models. 12

Table 2) NetApp disk shelves feature comparison. 12

Table 3) Netapp AFF controller and disk shelves compatibility. 12

Table 4) Netapp FAS controller and disk shelves compatibility. 13
Table 5) MetroCluster IP switch models. 14
Table 6) NetApp AFF controller and switch compatibility. 14
Table 7) NetApp FAS controller and switch compatibility. 15
Table 8) 40Gb 3m to 5m distance between switches (approximate cable length). 20
Table 9) 100Gb 3m to 5m distance between switches (approximate cable length). 20
Table 10) Short-range optical module for 40GbE switch. 20
Table 11) Short-range optical module for 100GbE switch. 20
Table 12) 40Gb and 100Gb optical cables. 20

LIST OF FIGURES

Figure 1) MetroCluster IP and VMware vSphere Metro Storage Cluster. 5
Figure 2) MetroCluster IP architecture. 6
Figure 3) Storage and server. 6
Figure 4) MetroCluster HA and disaster recovery. 7
Figure 5) MetroCluster IP combined fabric. 9
Figure 6) AFF A700 one site network example. 13
Figure 7) AFF A700 single site without network switch example. 14
Figure 8) Active-passive cluster or site. 16
Figure 9) Active-passive HA. 16
Figure 10) Hardware Universe platform limits. 17
Figure 11) RCF file generator. 18
Figure 12) Site A, a passive DWDM example using 10Gb optical modules and Quad Small Form-factor Pluggable Adapter (QSA). 22
Figure 13) ISL with 10Gb port adapter. 23
Figure 14) MetroCluster Mediator site. 26
Figure 15) MetroCluster Tiebreaker site. 27
Figure 16) Tiebreaker site link failure. 27
Figure 17) Tiebreaker site failure. 28
Figure 18) SVM disaster recovery. 29

MetroCluster overview

NetApp MetroCluster configurations are used by thousands of enterprises worldwide for high availability (HA), zero data loss, and nondisruptive operations both within and beyond the data center. MetroCluster is a free feature of ONTAP software that synchronously mirrors data and configuration between two ONTAP clusters in separate locations or failure domains.

MetroCluster provides continuously available storage for applications by automatically managing two objectives:

- Zero recovery point objective (RPO) by synchronously mirroring data written to the cluster.
- Near zero recovery time objective (RTO) by mirroring configuration and automating access to data at the second site.

MetroCluster provides simplicity with automatic mirroring of data and configuration between the two independent clusters located in the two sites. As storage is provisioned within one cluster, it is automatically mirrored to the second cluster at the second site. NetApp SyncMirror® provides a complete copy of all data with a zero RPO. This means that workloads from one site could switch over at any time to the opposite site and continue serving data without data loss.

MetroCluster manages the switchover process of providing access to NAS and SAN-provisioned data at the second site. The design of MetroCluster as a validated solution contains sizing and configuration that enables a switchover to be performed within the protocol timeout periods or sooner (typically less than 120 seconds). This results in a near zero RPO with the recovery for storage occurring within the storage protocol timeout periods. Applications can continue accessing data without incurring failures.

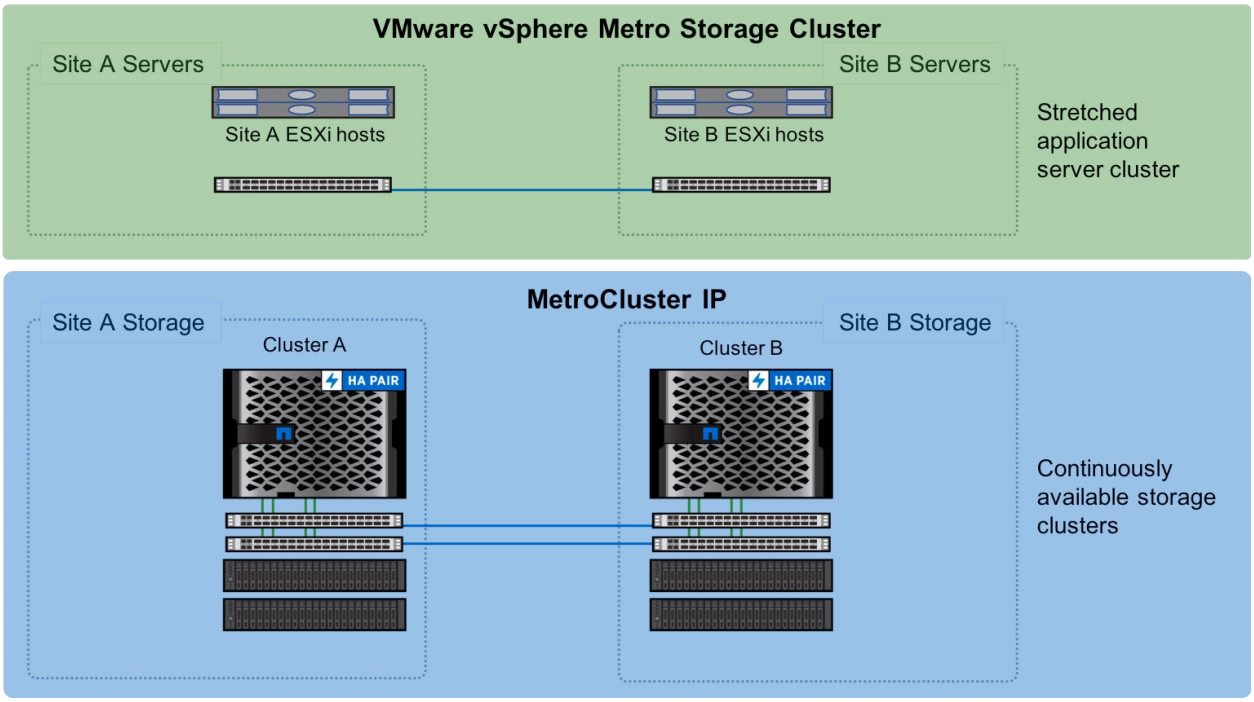
MetroCluster is available in several variations defined by the back-end storage fabric. There are two main types of storage fabric for MetroCluster: FC and Ethernet. The Ethernet storage fabric is referred to as MetroCluster IP.

Continuous availability solution overview

MetroCluster fulfills the need for continuously available storage. When combined with similar application availability products, the complete solution provides a highly resilient architecture that can continue operating even in the event of a site-wide disaster.

One example is using MetroCluster IP with VMware vSphere Metro Storage Cluster (vMSC). Combining the two products creates a highly resilient virtualized infrastructure that addresses the needs of business-critical applications. MetroCluster IP provides storage availability and vMSC provides a cross-site compute cluster that is available to operate even in the event of a complete site outage.

Figure 1) MetroCluster IP and VMware vSphere Metro Storage Cluster.



Similar multisite application solutions are available for databases and other applications that work well with MetroCluster.

MetroCluster IP compared to MetroCluster FC

The following features outline the differences between MetroCluster IP and FC:

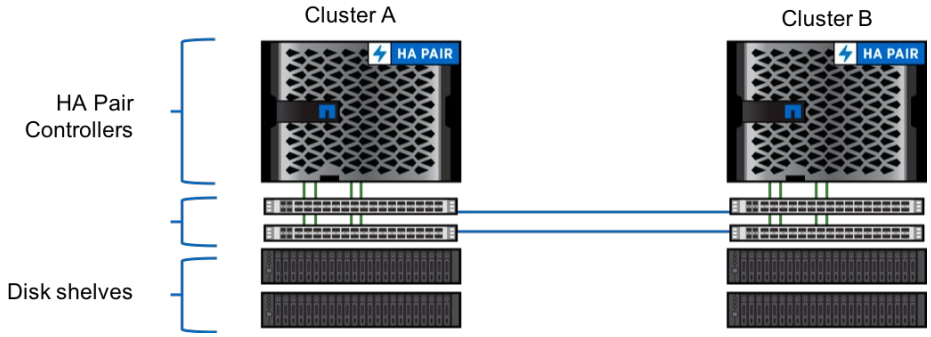
- MetroCluster IP uses an Ethernet back-end storage fabric rather than an FC back-end storage fabric, eliminating the need for dedicated FC switches.
- MetroCluster IP collapses the intercluster switches for both local and remote replication, eliminating the need for FC switches.
- MetroCluster IP does not require SAS bridges.
- MetroCluster IP replicates NVRAM with iWARP, a remote direct access memory protocol.
- MetroCluster IP accesses remote disks using iSCSI protocol with the remote disaster recovery node acting as the iSCSI target, supporting flash systems with integrated storage.

MetroCluster FC is also available in a smaller configuration called Stretched MetroCluster. For more information about MetroCluster FC with ONTAP, see TR-4375: NetApp MetroCluster FC for ONTAP.

MetroCluster IP architecture

MetroCluster IP uses an Ethernet storage fabric. The MetroCluster storage fabric, also referred to as the back-end storage fabric, is used solely by ONTAP. It is a separate dedicated network for ONTAP cluster interconnect, MetroCluster SyncMirror, and MetroCluster NVRAM mirror communications.

Figure 2) MetroCluster IP architecture.



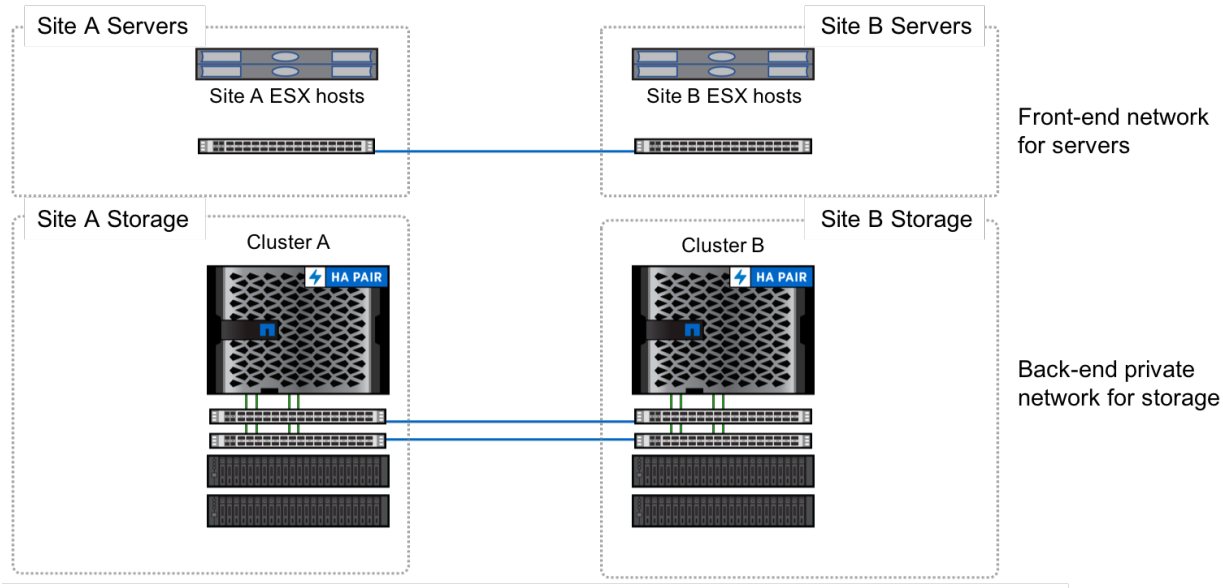
MetroCluster IP hardware summary:

- One HA pair controller per site
- Two high-speed Ethernet switches per site:
  - Collapsed intracluster and intercluster switches for local and remote replication.
- Disk shelves or internal storage

MetroCluster extends the availability of ONTAP by mirroring data between two independent ONTAP clusters. Each cluster is in a site or failure domain and leverages the standard HA features on the FAS or AFF systems. MetroCluster provides the capability to mirror both data and configuration between the two ONTAP clusters. MetroCluster includes validated system parameters and limits designed to provide failover from one site to the other within standard timeout periods for storage protocols.

MetroCluster features and hardware are a certified subset of the typical ONTAP FAS and AFF systems.

Figure 3) Storage and server.



MetroCluster has an architecture that can be broken down logically into several functional areas or components. Understanding how these components, such as replication, operate is important to build a well-architected solution and to administer the solution.

The following are the main MetroCluster components:

- Disaster recovery group
- Replication
- Network
- Storage

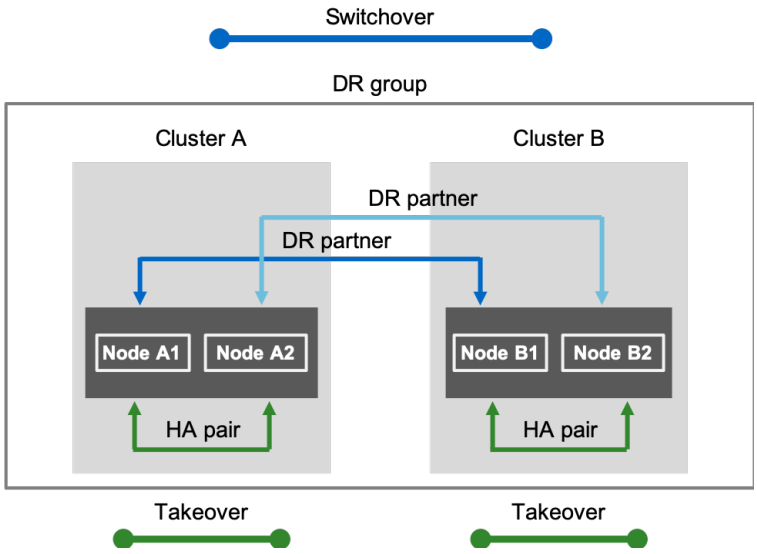
Disaster recovery group

MetroCluster IP uses the concept of disaster recovery for groups and partners to determine the relationship for failover and switchover. The two clusters (site A and site B) are configured together as a disaster recovery group. Within the group, the nodes are associated as disaster recovery partners.

The HA relationship is the same as in a standard cluster. HA protects against single controller faults and performs failover locally. HA is also leveraged for nondisruptive ONTAP updates. For site-wide faults, the disaster recovery relationship is used to switch from site A to site B, which is referred to as switchover.

The disaster recovery partner relationship is configured in the initial MetroCluster setup and does not change. There is one command that assigns one node from cluster A and one node from cluster B as partners. The remaining nodes are automatically assigned to complete the disaster recovery group configuration.

Figure 4) MetroCluster HA and disaster recovery.



Replication in MetroCluster IP

MetroCluster IP utilizes direct-attached storage, which eliminates the need for external SAS bridges to connect disks to the storage fabric. In this configuration, each node within the disaster recovery group functions as a storage proxy or iSCSI target, exporting its disks to the other nodes in the group. The storage transport protocol for the IP fabric is iSCSI (SCSI over TCP/IP), which enables communication between the iSCSI initiator and targets over a TCP/IP fabric.

To access remote storage, each node in the disaster recovery group leverages an iSCSI initiator to establish an iSCSI session with a partner node's iSCSI target. By utilizing iSCSI and direct-attached storage, systems with internal disks can also be used. This configuration allows each node to provide disaster recovery partner nodes with access to both internal storage and storage devices located in external disk shelves.

MetroCluster has three planes of replication:

- Configuration replication - MetroCluster (MC) is a configuration consisting of two ONTAP clusters, each with its own replicated database (RDB) containing its own metadata. The transfer mechanism of metadata objects between the clusters during switchover includes three components: cluster peering, configuration replication service (CRS), and metadata volumes (MDVs). MC enables the replication of configuration objects between the clusters using the peering network, with changes propagated near synchronously to the other cluster's RDB over the configuration replication network. MDVs are used as a fallback mechanism in case the cluster peering network is temporarily unavailable.
- NVRAM replication - NVRAM replication involves copying the local node's NVRAM to the NVRAM of the remote disaster recovery node to protect against data loss in the event of a failover or switchover. NVRAM is mirrored both locally to a local high-availability partner and remotely to a disaster recovery partner, and the nonvolatile cache is split into four partitions for the local, HA partner, DR partner, and DR auxiliary partner. Each node's NVRAM is mirrored twice in a four-node setup, which includes eight-node setups which have two four-node DR groups. The mirroring is done to both the HA partner and DR partner, and updates are transmitted using the iWARP protocol over the ISL (Inter-Switch Link) for MetroCluster IP.
- Storage replication - MetroCluster IP uses RAID SyncMirror (RSM) to mirror the local and remote back-end disks, with each node in a disaster recovery group serving as a remote iSCSI target to present the back-end storage as logically shared. The node accesses its remote back-end disks by going through its remote disaster recovery partner node to access the remote disks that are served through an iSCSI target. Blocks are written to the paired nodes at each site with both NVRAM and SyncMirror, and SyncMirror writes occur in the RAID layer, allowing storage efficiencies such as deduplication and compression to reduce the data written by SyncMirror operations.

For more detailed information about replication within MetroCluster, refer to the following documentation.

- TR-4705: NetApp MetroCluster Solution Architecture and Design

Encryption for replication

MetroCluster does not provide a mechanism to encrypt data being sent between the sites. There are options for encrypting the data at the network and/or storage layers.

- Network layer encryption can be achieved with Wavelength Division Multiplexing (WDM) devices and switch-based encryption. The use of external encryption devices is supported if the round-trip latency remains within requirements.
- Alternatively, you can use host-side encryption of the storage layer. The disadvantage is that this negates any storage efficiencies that ONTAP normally provides.

Note: Please take note that while you can use NetApp Volume Encryption (NVE) to encrypt data written to a volume, any writes are still replicated with NVRAM, including unencrypted block data that is written by the host.

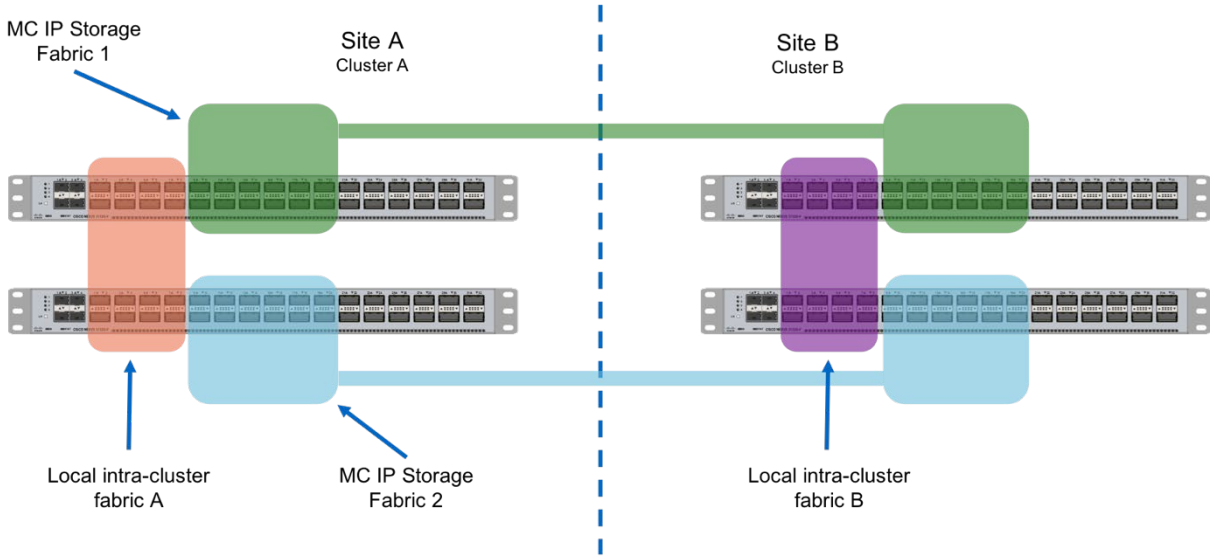
Network

There are two independent storage fabrics for MetroCluster:

- Cluster interconnect
- MetroCluster IP network

Each network is dedicated to certain functions. There are specific virtual LANs (VLAN) that map to each of the networks to create separate data link layers, or layer 2 in the OSI standard.

Figure 5) MetroCluster IP combined fabric.



NetApp provides a standard switch configuration known as a reference configuration file (RCF). Use the RCF generator tool to create the RCF file depending on your switch vendor and the platform models. The NetApp Support Site hosts the download for the RCF File Generator. The RCF files must be used to make any configuration changes to the switches. The RCF file is a bundle of four individual configuration files – one per switch.

The RCF designates the VLAN and the channel group identifiers (IDs). These are used only within the back-end storage switches. There are specific requirements for IP addresses. See the MetroCluster IP Installation and Configuration guide for a worksheet and description of the requirements.

Cluster interconnect

The ONTAP cluster interconnect is a high-speed, local-only network vital for node communication in a NetApp storage system. It does not connect between sites, as all cluster traffic is local to the site where the nodes are located. It ensures efficient, reliable communication and supports scalability.

In a MetroCluster IP setup, dedicated switch ports handle interconnect traffic. Two ISL ports connect site switches, forming a redundant local VLAN. Breakout cables are employed for lower-than-native port speeds.

NetApp interconnects form a dedicated, private network in a storage cluster, separate from data and management networks. They offer redundancy, high-speed connectivity, and scalability by using switches, cables, and NICs or HCAs. These components ensure high performance and reliability in clustered storage environments.

Note: The AFF A150, AFF A250, AFF C250, FAS500f, and FAS2750 share cluster ports with MetroCluster IP networking, using VLANs for separation.

Note: Starting with ONTAP 9.7, the option for switchless interconnects by cross-connecting ports between storage controllers is available. This is used in configurations with NetApp compliant switches.

MetroCluster IP network

In a MetroCluster IP configuration, each site consists of two independent storage fabrics. These local fabrics are connected to corresponding remote fabrics but are not connected to each other, which differentiates them from the cluster interconnect.

Each MetroCluster IP switch has several ports designated for node connectivity. In a standard four-node MetroCluster IP configuration, which includes two nodes per site, only two of these ports are utilized. Each of the two nodes at a site connects to both switches using separate Ethernet interfaces. Typically, standard cluster interconnect ports, either onboard ports or interfaces used for standard ONTAP HA configurations, are employed for these connections.

The MetroCluster IP network employs a specialized Ethernet adapter optimized for the internet-wide area remote direct memory access (RDMA) protocol (iWARP). This adapter features both a TCP offload engine (TOE) and an iSCSI offload capability, facilitating RDMA over high-speed Ethernet. Each node is equipped with two iWARP/iSCSI adapter ports, with one port connecting to each switch. The switches form separate fabrics that remain unconnected locally, as illustrated in Figure 5.

For platforms such as NetApp AFF A150, AFF A250, AFF C250, FAS500f and FAS2750, software iWARP is utilized. These platforms have a fixed number of network ports, and to optimize network port usage for front-end, host-side data access, the two onboard ports (e0a and e0b) typically reserved for cluster interfaces, are combined. This configuration allows cluster traffic and MetroCluster IP traffic to share the same ports, leaving the remaining four network ports available for host-side data access.

Note: iWARP is a standards-based protocol described in IETF RFC 5040 – A Remote Direct Memory Access Protocol Specification.

Storage

Storage for MetroCluster IP is not directly shared across the two sites. While it is not required to configure unique shelf IDs across multiple sites, it is considered a best practice to assign a unique ID to each shelf. This makes it easier to identify and manage shelves within the storage system. The storage at each site is only directly accessible by the local HA pair. The remote storage is made available by the local nodes using iSCSI as described in the Storage Replication section.

SyncMirror

SyncMirror, or RAID SyncMirror (RSM) is the technology used in MetroCluster to mirror the aggregates between the sites. It enables two plexes to be configured in each aggregate, referred to as pool0 and pool1. Pool0 contains the local storage for a node and pool1 contains the remote mirror copy.

ADP

As of ONTAP 9.4, MetroCluster IP supports Advanced Disk Partitioning (ADPv2) on AFF systems using Root-Data-Data (RD2) partitioning. Advanced Drive Partitioning (ADP) is a feature that enhances storage efficiency, allowing for capacity sharing of physical drives between aggregates and controllers within an HA pair. ADP increases usable and effective capacity and improves storage efficiency by up to 10-40% compared to whole disk partitioning. RD2 partitioning divides a drive into one root partition and two data partitions, enabling the capacity and IOPS of a single drive to be used by both controllers in an AFF system. ADP is applied by default at the time of MC initialization.

For more detailed information about ADP, refer to the following documentation.

- TR-4705: NetApp MetroCluster Solution Architecture and Design

Solution design

Proper design of a solution is the key to addressing performance, capacity, and resiliency requirements. The overall steps for designing a solution include checking for supported hosts and platform configurations as well as sizing to meet capacity and performance needs. The following issues must be considered:

- Ensuring support for hosts and protocols
- Sizing of a solution for performance
- Sizing of a solution for capacity: active-active, active-passive configurations for capacity
- Reviewing systems limits
- Sizing ISLs between sites
- Cabling requirements

Confirming support

Review the NetApp Interoperability Matrix Tool (IMT) to verify that the host-side protocol and operating system versions are supported in the same way as any ONTAP design. Check any alerts noted in the results pages to see if they apply to MetroCluster.

The Hardware Universe lists system specifications and supported limits. Starting with ONTAP 9.6, the Hardware Universe also contains interoperability information for ONTAP 9.6 and later.

Hardware components

The following is a broad summary of the primary hardware components included in the MetroCluster IP setup. Please note that specific model and part numbers might vary depending on the exact configuration and requirements of your organization. These are all detailed in the Interoperability Matrix Tool (IMT) and the Hardware Universe. For more information, review the documentation to Install a MetroCluster IP configuration.

Platforms

The following platforms offer the commonly used models in a MetroCluster IP configuration:

- AFF A-Series - NetApp AFF A-Series, designed specifically for flash, deliver industry-leading performance, density, scalability, security, and network connectivity. These systems deliver the industry's lowest latency for an enterprise all-flash array, making them a superior choice for running the most demanding workloads and AI/DL applications.
- AFF C-Series - NetApp AFF C-Series systems help you move more of your data to flash with the latest QLC flash technology. These systems are suited for large-capacity deployment as an affordable way to modernize your data center to all flash and connect to the cloud.
- AFF ASA - NetApp ASA systems are built on NetApp AFF systems, which deliver industry-leading performance and reliability. AFF systems provide an enterprise-class SAN solution for customers who want to consolidate and to share storage resources for multiple workloads.
- FAS - NetApp FAS storage arrays are hybrid storage systems that can handle a mix of flash and hard disk drives. FAS systems provide the optimal balance of capacity and performance for easy deployment and operations while also having the flexibility to handle future growth and cloud integration.

For more information, review the following datasheets.

- AFF A-Series
- AFF C-Series
- FAS Storage Arrays

Controller models

NetApp models available with MetroCluster IP offer varying levels of performance, capacity, and scalability, catering to different business needs and budget requirements. When deploying a MetroCluster IP solution, it is essential to ensure that both the primary and secondary sites have the same type of storage system (either AFF or FAS) and compatible hardware and software components. The following table lists the controller models supported with MetroCluster IP.

Table 1) MetroCluster IP controller models.

Platform	Entry	Mid-Range	High-End
AFF A-Series	A150	A250, A400	A700, A800, A900
AFF C-Series		C250, C400	C800
FAS	FAS27xx	FAS500f, FAS8300, FAS8700	FAS9000, FAS9500

Specific models and features are subject to change, please review the AFF and FAS System Documentation for details.

Disk shelves

Disk shelves serve as storage devices that allow for the expansion of a storage system's capacity. When used in a MetroCluster IP setup, they play a crucial role in ensuring the high availability of data by facilitating data replication between two physically distant locations. The disk shelves are accessible by the nodes (controllers) at both locations, enabling the creation of mirrored volumes between the two sites. Thus, disk shelves are a vital aspect of MetroCluster IP configurations, providing the necessary storage capacity and data replication capabilities to maintain data and application availability across geographically dispersed sites. Disks are available in varying storage capacities and are categorized according to their physical dimensions as either large form factor (LFF) - typically 3.5 inches - or small form factor (SFF) - which is usually 2.5 inches. The following tables compare the available shelves and their compatibility with supported MetroCluster IP controllers.

Table 2) NetApp disk shelves feature comparison.

Feature	NS224	DS224C	DS212C	DS460C
Form factor	2U	2U	2U	4U
Drive per enclosure	24 SFF	24 SFF	12 LFF	60 LFF
Drive types	SSD	SSDs, HDDs	SSDs, HDDs	SSDs, HDDs
I/O module	dual NSM	dual IOM12	dual IOM12	dual IOM12
Connectivity	100Gb/s Eth	12Gb/s SAS	12Gb/s SAS	12Gb/s SAS

Table 3) Netapp AFF controller and disk shelves compatibility.

	AFF A150	AFF A250	AFF A400	AFF A700	AFF A800	AFF A900	AFF C250	AFF C400	AFF C800
DS212C									
DS224C	X	X	X	X	X	X			
DS460C									
NS224	X	X	X	X	X	X	X	X	X

Table 4) Netapp FAS controller and disk shelves compatibility.

	FAS2750	FAS500f	FAS8300	FAS8700	FAS9000	FAS9500
DS212C	X		X	X	X	X
DS224C	X		X	X	X	X
DS460C	X		X	X	X	X
NS224		X				

See the product documentation for further information about the NS224 shelves. Review the SAS shelves section for more information about the DS212C, DS224C and DS460C shelves.

Switches

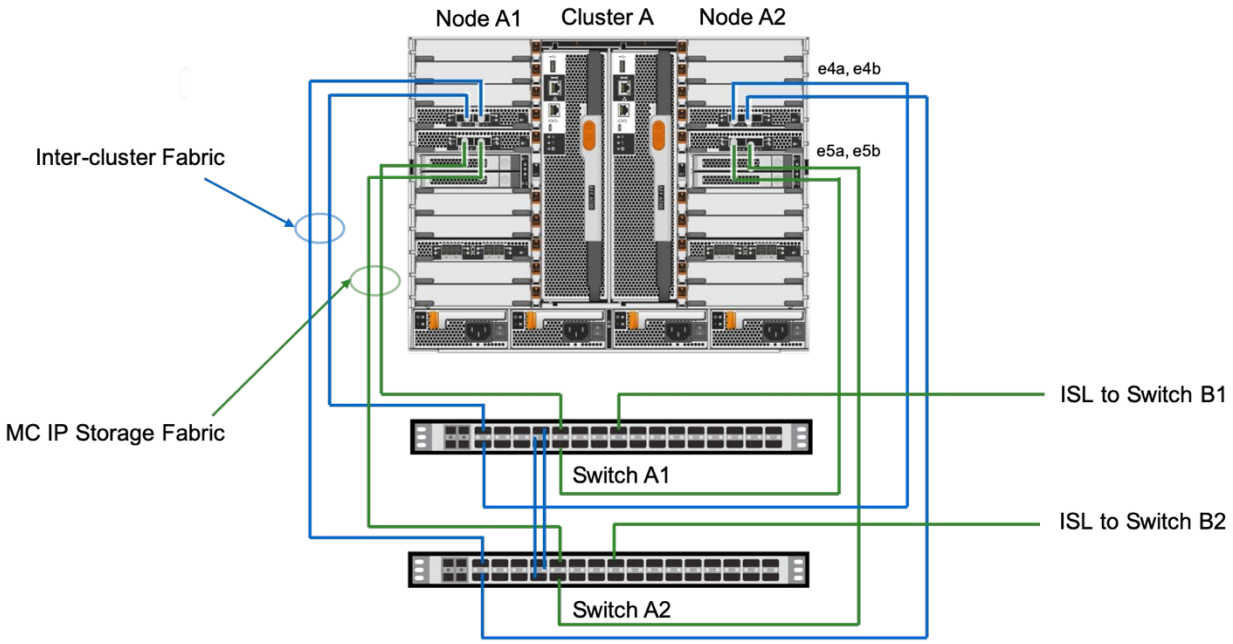
These Layer 3 Ethernet switches are essential for creating robust IP-based interconnects between data centers, ensuring data integrity and uninterrupted operation. They support 10/25/40/50/100-Gigabit Ethernet connectivity for rapid data transfer and minimal latency.

These switches offer features such as jumbo frames, VLAN tagging, and link aggregation, enhancing the MetroCluster environment's performance and reliability. To maintain seamless communication between storage controllers, the switches support routing protocols like OSPF and BGP, ensuring optimal path selection and network load balancing.

Each MetroCluster IP deployment requires four switches—two per site for redundancy—and does not allow mixing switch models. Refer to the Interoperability Matrix Tool on the NetApp Support site and the Hardware Universe for information on supported switch models for specific platforms and ONTAP versions.

Breakout cables are employed for port speeds lower than native port speed, such as connecting optical modules with 10Gb ISL links.

Figure 6) AFF A700 one site network example.

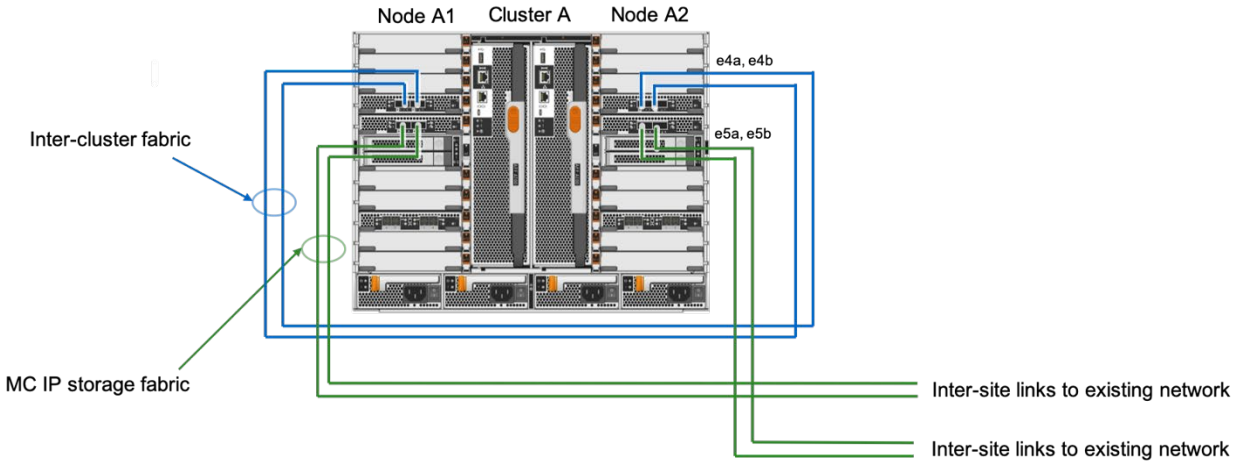


NetApp compliant switches – MetroCluster IP with existing switches

ONTAP 9.7 introduced support for MetroCluster IP on certain platforms without requiring NetApp validated switches. This enables the utilization of existing switches for purposes beyond the MetroCluster IP back-end storage fabric. The solution employs a switchless cluster configuration at each site, consisting of two-node switchless clusters. In this setup, the cluster interconnect interfaces are cross-connected, and the MetroCluster IP interfaces are connected to existing, MetroCluster-compliant switches.

For more information, review the considerations for using MetroCluster-compliant switches.

Figure 7) AFF A700 single site without network switch example.



The following tables compare the available switch models and their compatibility with supported MetroCluster controller models.

Table 5) MetroCluster IP switch models.

Model Description	NetApp PN	Native Port Speed
Broadcom BES-53248	X190005	10/25/100Gb
Cisco Nexus 9336C	X190200-CS-PE	40Gb/100Gb
Cisco Nexus 3232C	X190100	40Gb/100Gb
NVIDIA SN2100	X190006	40Gb/100Gb

Table 6) NetApp AFF controller and switch compatibility.

	AFF A150	AFF A250	AFF A400	AFF A700	AFF A800	AFF A900	AFF C250	AFF C400	AFF C800
Broadcom BES-53248	X	X	X					X	
Cisco Nexus 9336C	X	X	X	X	X	X	X	X	X
Cisco Nexus 3232C	X	X	X	X	X	X	X	X	X
NVIDIA SN2100		X	X	X	X	X	X	X	X
NetApp compliant			X	X	X	X		X	X

Table 7) NetApp FAS controller and switch compatibility.

	FAS2750	FAS500f	FAS8300	FAS8700	FAS9000	FAS9500
Broadcom BES-53248	X	X	X	X		
Cisco Nexus 9336C	X	X	X	X	X	X
Cisco Nexus 3232C	X	X	X	X	X	X
NVIDIA SN2100			X	X	X	X
NetApp compliant			X	X	X	X

Network adapters

MetroCluster IP configurations rely on specialized network adapters to enable efficient data replication and communication over IP networks. These adapters are platform-dependent and provide fast Ethernet with dual ports, allowing for connections to two distinct layer 2 or layer 3 networks. Additionally, the adapters offer iWARP offload capabilities that enable Remote Direct Memory Access (RDMA) over Ethernet networks. This feature enables direct memory-to-memory data transfers between servers and storage systems, minimizing latency and overhead.

These adapters are responsible for the node-to-switch connection used for storage and NVRAM replication. For the cluster interconnect, a separate network adapter or network ports are used.

Different NetApp models implement iWARP and network connections in varying ways, and platform models have different requirements and implementations for network adapters. For instance:

- AFF A150, AFF A250, AFF C250, FAS500f, and FAS2750 utilize onboard ports.
- AFF A400, AFF A700, FAS8300 and FAS9000 utilize a single network adapter.
- AFF A800 utilizes two adapters: one onboard and one adapter.
- AFF A900 and FAS9500 utilize two network adapters.

The AFF A150, AFF A250, AFF C250, FAS500f, and FAS2750 use software iWARP combined on the cluster interfaces, which enables traffic sharing on the onboard e0a and e0b interfaces. This reduces the port count needed for back-end storage and maximizes the number of ports available for host-side data interfaces.

Sizing a solution

A solution can be sized to meet specific storage capacity or performance requirements. MetroCluster sizing is like sizing an HA pair with respect to capacity. With MetroCluster, the storage devices are double the capacity used for an HA pair to provide the mirror copy of the data at the opposite site.

With respect to sizing the performance, the ISLs are a factor that can be accounted for using the ISL sizing spreadsheet.

Active-Passive configurations

Active-passive configurations in NetApp environments utilize two storage nodes or clusters to ensure high availability, minimal downtime, and seamless failover/failback transitions. The active node manages client I/O requests and storage resources, while the passive node monitors the active node's health, ready to take over during failures or maintenance. Key benefits include high availability, non-disruptive failover and failback, and continuous monitoring.

There are two types of active passive configurations:

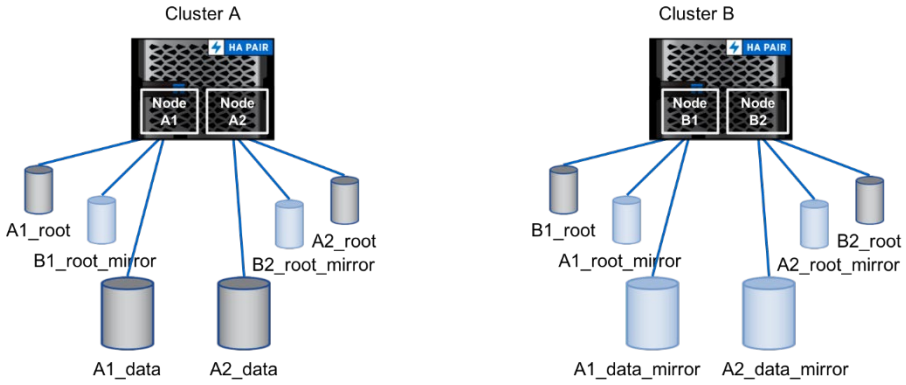
- Active-passive clusters
- Active-passive HA in a cluster

Active-passive cluster

An active-passive cluster or site is when one site is used for production workloads, the active site, and the second site has minimal capacity that is used for failover. This enables smaller storage configurations where the active site contains the active storage and workloads. It is also cost efficient because you do not need disk capacity when the second site is used only for operations in a switchover or site failure situation.

For an active-passive cluster configuration, one cluster has all pool0 disks and the other cluster has all pool1 disks. For an active-passive cluster or site a small data aggregate must be created to host the volume that contains metadata for MetroCluster. Except for root volumes and a small data volume for the volume that contains metadata, the passive site only contains mirror copies of the data.

Figure 8) Active-passive cluster or site.



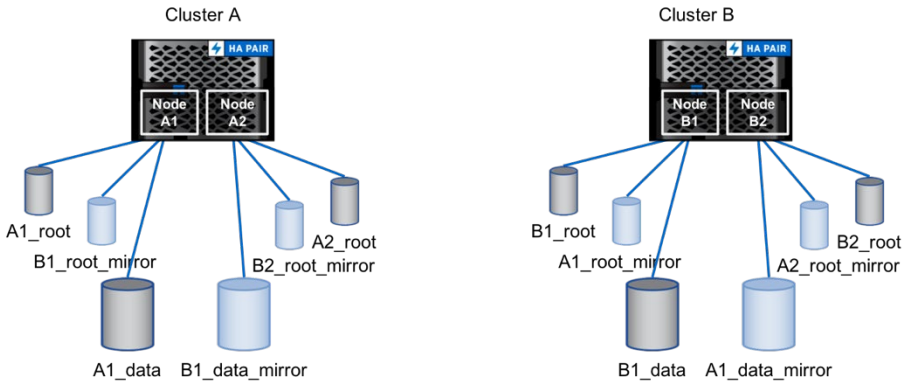
Active-passive HA configuration

An active-passive HA configuration is when the storage is allocated to one of the nodes in an HA pair. This configuration is typically done to maximize capacity in smaller configurations.

Note: For AFF, you must have storage distributed equally between the nodes.

In this example, each node owns disks with root volumes. The local active data volume is hosted on node 1 and the remote mirror copy is hosted on node 2.

Figure 9) Active-passive HA.

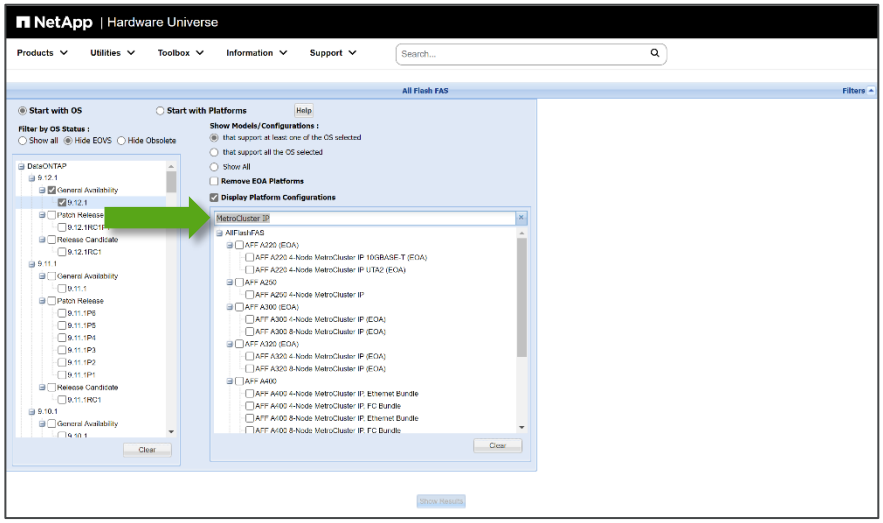


For more information about active-passive HA pairs, see HA pair management overview in the ONTAP 9 Documentation Center.

Platform limits

The Hardware Universe contains ONTAP limits, which can be found under the designated platform and ONTAP version. At the bottom of the results page, you will see the platform restrictions. To view the limits for an HA pair, choose MetroCluster IP from the Platform Configurations menu.

Figure 10) Hardware Universe platform limits.



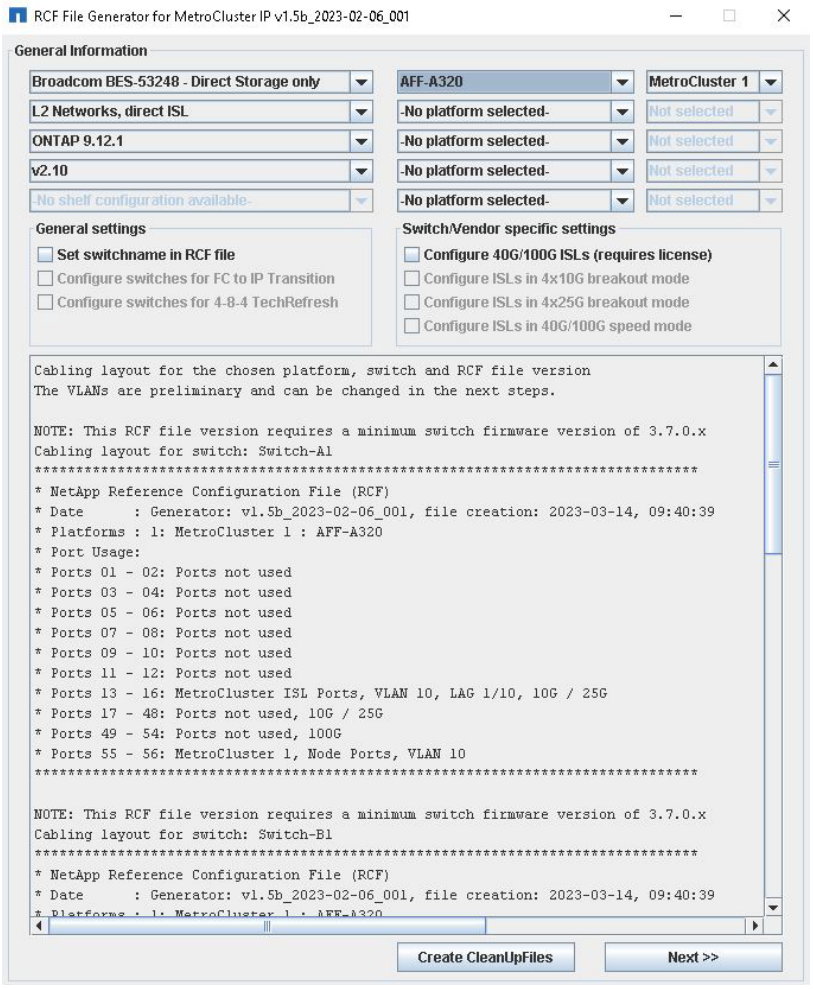
Network configuration

Network configuration requires that VLANs and IP addresses do not overlap with other networks. VLANs are for the private back-end fabric and are assigned automatically in the switch RCF files.

Before ONTAP 9.6, RCF files were created by switch model, platform, and ONTAP release. However, starting with ONTAP 9.6, a new RCF File Generator utility is available that can create RCF files for all MetroCluster IP switch models and supported platforms. The RCF File Generator utility can generate RCF files with customer-provided VLAN IDs in support of a shared layer-2 site-to-site network. In addition, the RCF file generator has support for layer-3 networking configurations. Output from the RCF file generator describes the cabling layout for the chosen platform, switch, and RCF file version

Note: The AFF A220 and FAS 2750 controllers do not enable changing the VLAN IDs in ONTAP 9.7 and earlier. However, in ONTAP 9.8, the VLAN for AFF A220, AFF A250, FAS 2750, and FAS500f can be specified. The defaults are 10 and 20, and the user specified VLAN must be greater than 100 and less than 4096.

Figure 11) RCF file generator.



In n a typical switch configuration, most of the ports are open for future expansion. The switches create two redundant fabrics, and each node has a cluster connection and a MetroCluster IP node connection. There are several choices for cabling depending on the distance of the nodes to the switches. If possible, it is optimal to locate the nodes and switches in the same equipment rack to enable the use of copper Twinaxial cabling instead of requiring optical cabling.

The switches are capable of native port speed, either 10/25Gbps, 40Gbps, or 100Gbps, depending on the model. The native ports are also capable of operating in break-out mode, where the port is divided into four separate lanes that are used as individual interfaces. When operating in break-out mode, a port on the 40-Gb switch can operate as four 10Gbps interfaces that are dependent on the cable or optical module. Not all optical modules support break-out mode.

The 100-Gb switch supports break-out mode as well. When operating at native port speed, a single physical interface operates as four 25Gbp interfaces. Specific cables and optical modules support operating in break-out mode. The RCF files preconfigure the break-out ports for the specific speeds required for each platform.

Note: It is possible to launch the RCF File Generator from the command line. The JAVA_HOME environment variable should be set. To run the RCF File Generator, enter the command as follows:

```
java -jar RcfFileGenerator.jar
```

Cluster fabric Inter-Switch Link (ISL)

The cluster interconnect fabric is a dedicated network used for communication between the nodes of a MetroCluster IP configuration. This network needs specific switch requirements and cable types to ensure reliable and optimal performance.

Switches used for the ISL network must have native ports and should be configured with jumbo frames to support high-speed data transfer. The fabric requires a minimum of two switches.

It is recommended to use either Twinaxial copper cabling or fiber optic cabling. Twinaxial copper cabling is typically used for distances up to 7 meters, while fiber optic cables are used for longer distances. The cables should be high-quality and properly shielded to reduce interference and ensure reliable data transfer.

MetroCluster IP Inter-Switch Link design

To ensure high availability and resilience in a MetroCluster IP configuration, an ISL design is required to connect the two switches with high-speed and low-latency links. The ISL link design involves several critical factors, including sufficient bandwidth to support replication traffic, low latency for quick data transfer, redundancy for high availability, appropriate QoS settings to prioritize replication traffic, and secure authentication to prevent unauthorized access.

The ISL link design process for MetroCluster IP can be complex due to site-specific factors, such as distance and network capabilities between sites. Various elements, such as storage performance requirements, direct fiber availability, multiplexing devices, and existing fiber infrastructure, must be considered to determine the necessary components and supported distances for each link. It is advisable to consult with NetApp to help perform the necessary calculations.

Currently, ISL networking round trip time (RTT) requirements for MetroCluster IP are as follows: latency should be less than or equal to 7ms; jitter should be less than or equal to 3ms and packet loss should be less than 0.01%. These requirements provide a maximum latency of up to 10ms, which is essential for synchronous replication, where there is a strict time limit for data transfer.

When designing an ISL configuration for MetroCluster IP, it is important to select the appropriate optical modules and matching optical cable configurations that provide maximum supported distances. For shorter distances, multimode optics and cables are suitable and cost effective. For longer distances, long-range optics and single-mode fiber are required, and the telecommunications provider should be consulted for maximum distances.

In some cases, standard Ethernet cabling can be used for distances within a single data center separated by racks in separate availability zones, providing the ISLs at the native switch port are at speeds of 25Gb, 40Gb, or 100Gb, depending on the switch and modules.

In summary, careful planning and implementation of the ISL link design for MetroCluster IP, considering site-specific factors, can ensure fast, reliable, and secure replication traffic between the switches.

Rack-to-Rack: short distances

For MetroCluster IP configurations where the separate locations are in proximity (e.g., within a rack or side-by-side racks), it might be possible to use copper cabling for the ISL links. This is often the case with laboratory or test configurations. Table 6 and Table 7 show the part numbers for the switch-to-switch copper cables for both the 40Gb and 100Gb switches.

Table 8) 40Gb 3m to 5m distance between switches (approximate cable length).

Distance	Cable PN	Description
Less than 1m	X66100-3	Cable, copper, QSFP+-QSFP+, 40GbE, 1m
Less than 3m	X66100-3	Cable, copper, QSFP+-QSFP+, 40GbE, 3m
Less than 5m	X66100-5	Cable, copper, QSFP+-QSFP+, 40GbE, 5m

Table 9) 100Gb 3m to 5m distance between switches (approximate cable length).

Distance	Cable PN	Description
Less than 1m	X66211A-1	Cable, copper, QSFP28-QSFP28, 100GbE, 1m
Less than 2m	X66211A-2	Cable, copper, QSFP28-QSFP28, 100GbE, 2m
Less than 5m	X66211A -5	Cable, copper, QSFP28-QSFP28, 100GbE, 5m

Similarly, the use of optical cabling between racks is possible. This enables a simple ISL configuration when the distances are within the specification of the optical modules.

Table 10) Short-range optical module for 40GbE switch.

Distance	Module PN	Description
Up to 400M on OM4	X65401	XCVR, QSFP+, Optical, 40GbE, shortwave

Table 11) Short-range optical module for 100GbE switch.

Distance	Module PN	Description
Up to 100M on OM4	X65405	XCVR, QSFP28, Optical, 100GbE, shortwave

Table 12) 40Gb and 100Gb optical cables.

Length	Module PN	Description
2M	X66200-2	Cable, Optical, OM4, MPO/MPO Type B
5M	X66200-5	Cable, Optical, OM4, MPO/MPO Type B
15M	X66200-15	Cable, Optical, OM4, MPO/MPO Type B
30M	X66200-30	Cable, Optical, OM4, MPO/MPO Type B

Campus links

Campus links that use direct fiber connections between short distances can be similar to using rack-to-rack ISLs. One potential difference is the use of long-range optics and single-mode cabling to achieve longer distances compared to multimode cabling and short-range optics.

Currently, NetApp does not offer long-range optical modules for either the 40GbE or 100GbE switches. For designing links that require long-range optics, see the Cisco support matrix for the specific switch model and the Cisco optical module datasheets to determine distance and connection specifics.

Dedicated fiber links

Dedicated fiber links are more common for campus networks connecting buildings located in proximity. With dedicated fiber links, you might want to multiplex signals from many fiber connections onto fewer fiber links. Doing this can maximize utilization and reduce the required number of fibers between the sites. Multiplexing of optical signals is called wavelength division multiplexing (WDM) and it is available in two types, coarse wavelength division multiplexers (CWDM) and dense wavelength division multiplexers (DWDM).

CWDM can multiplex a smaller number of wavelengths compared to DWDM.

CWDMs are commonly passive devices that optically multiplex and demultiplex the light from the optical modules into a single signal that can be transmitted across a single fiber pair. The optical modules are wavelength specific, sometimes referred to as channel. To multiplex two different fiber signals, each source signal must be generated from an optical module that uses a different wavelength. CWDM optical modules are available from Cisco and support eight different wavelengths. This enables the multiplexing of eight fiber signal links onto a single fiber link. The CWDM multiplexer is passive and only contains optics that multiplex and demultiplex the signals. This typically is a lower cost for multiplexing devices and associated optical modules compared to DWDM.

DWDMs use a similar method to CWDM devices for merging signals. The primary difference is that the optical modules are more precise in the signals they generate, enabling a narrower spectral width for a narrow signal and less spacing between the signals. This enables a higher number of signals to be combined for transmission on a site-to-site fiber link. DWDM devices can be active or passive. Passive devices use the same approach as CWDM where the optical modules transmit a specific wavelength or channel that is merged in the DWDM device to produce a single signal. This signal is transmitted on the longer distance fiber cable between the sites.

DWDM devices are also available as active devices. In this case, the signals between the switch and the DWDM device use standard optics and rely on the DWDM device to produce a signal at the wavelength that can be merged onto the site-to-site fiber link.

For distance, the optical modules provide specifics on the allowable distance and the link characteristics required to meet the specifications. Transmitting over longer distances might require a signal amplifier. There are several types of amplifiers that apply for DWDM such as an optical amplifier. NetApp recommends consulting with a telecommunications specialist to help design the optimal configuration.

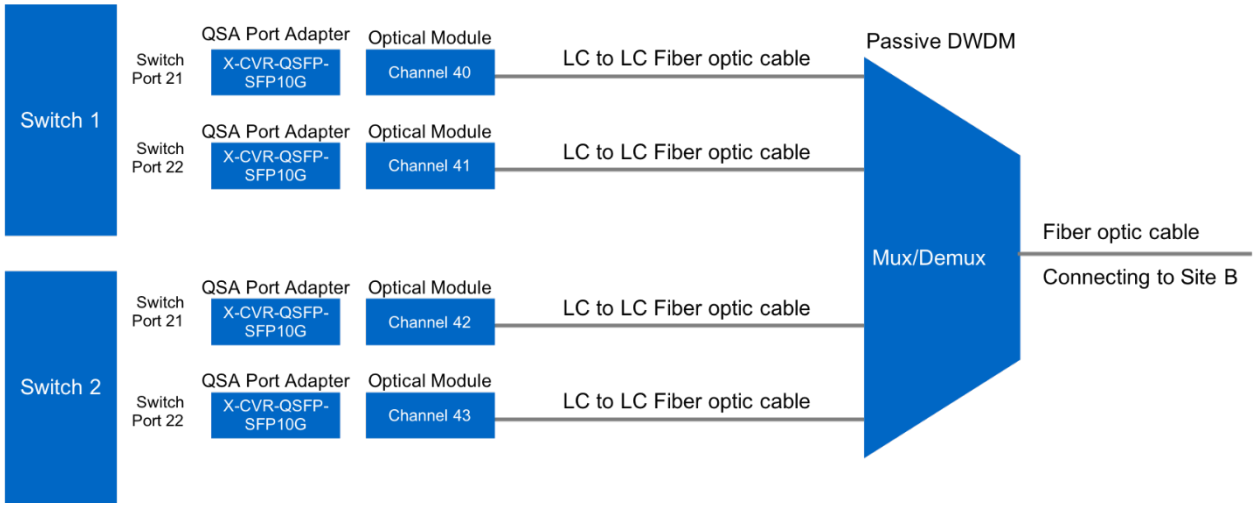
Cisco provides 10Gb SFP+ modules that can be used for coarse or dense wavelength multiplexing. xWDM enables multiple optical signals to be combined or multiplexed on a single fiber pair between sites, then demultiplexed before routing the signal to a switch or device.

Figure 12 is an example of using a passive DWDM and shows a possible mapping for the optical modules on specific channels. All the optical modules for each site use a unique channel. Match the opposite site B optic, same channel.

Example using DWDM modules.

- Site A switch 1 port 21 – Site B switch 1 port 21 using two optical modules on channel 40
- Site A switch 2 port 21 – Site B switch 2 port 21 using two optical modules on channel 41
- Site A switch 1 port 22 – Site B switch 1 port 22 using two optical modules on channel 42
- Site A switch 2 port 22 – Site B switch 2 port 22 using two optical modules on channel 43

Figure 12) Site A, a passive DWDM example using 10Gb optical modules and Quad Small Form-factor Pluggable Adapter (QSA).



The first module in the example is Cisco part number DWDM-SFP10G-45.32, that is a 10GBASE-DWDM SFP+ module operating on the 1545.32-nm wavelength (100-GHz ITU grid) which is ITU channel 40. To complete the configuration in this site, three more modules must be supplied, each corresponding to channels 41, 42 and 43. Site B then contains the exact same configuration of optical modules, port adapters, and passive DWDM.

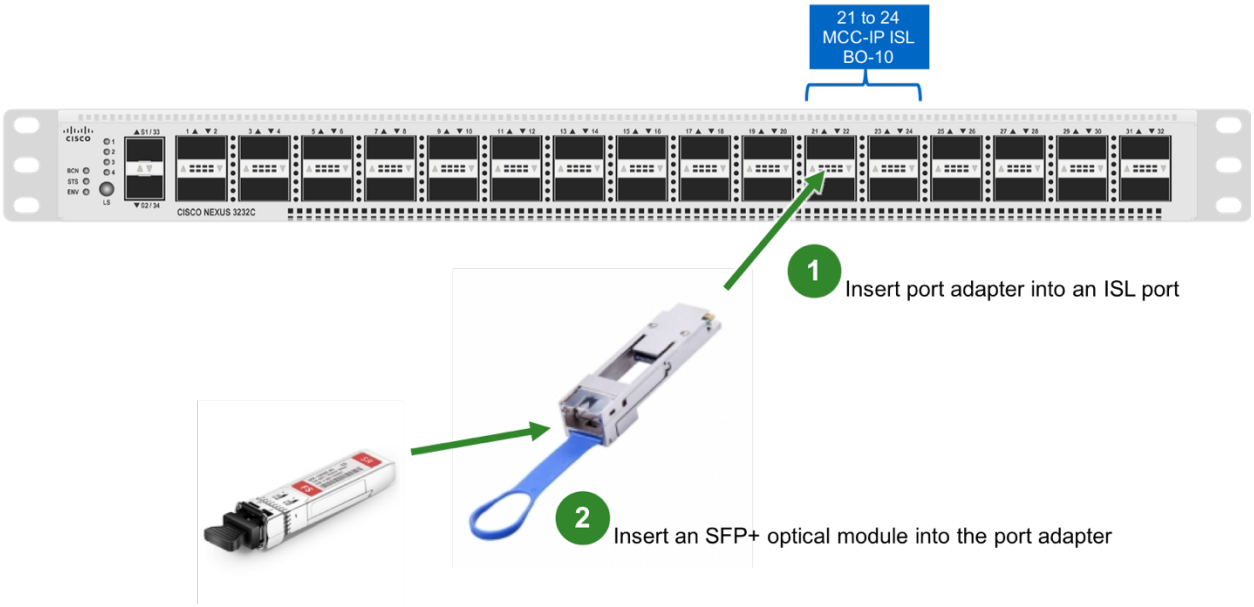
Intercity links

For the longest distance links, active DWDM or Telco circuits are often used. Connection from the switch to most telecommunications or active DWDM devices is done with the same optical modules that would be used in a data center or rack-to-rack configuration. See section “Rack-to-Rack: short distances” for rack-to-rack cabling and modules.

Note: Depending on the equipment, it is also possible that an active DWDM can provide encryption of the ISL traffic.

NetApp recommends that you consult with a telecommunications specialist to help design the optimal configuration for intercity links.

Figure 13) ISL with 10Gb port adapter.



For more information about cabling and optical modules, see the MetroCluster IP switch technical references in Where to Find Additional Information.

Cisco DWDM optical modules are available in three different configurations:

- Cisco DWDM-SFP10G-XX.XX modules
  - DWDM fixed module supports 40 non-tunable ITU 100GHz wavelengths
- Cisco DWDM-SFP10G-C module
  - DWDM tunable module supports 96 tunable ITU 50GHz wavelengths up to 80km
- Cisco DWDM-SFP10G-C-S module
  - Tunable transceiver modules are Ethernet only
  - DWDM tunable module supports 96 tunable ITU 50GHz wavelengths up to 70 km

For more information about supported optical modules and PNs to order from a Cisco partner, see the section "Where to Find Additional Information."

ISL networking: dedicated link, shared layer 2, or layer 3

MetroCluster IP configurations can use dedicated or shared ISLs for replication between sites. In a dedicated ISL configuration, the ISL switches must have native-speed ports that match the speed of the ISL ports on the MetroCluster IP switches. The number and speed of ISLs between the MetroCluster IP switches and the customer network switches do not need to match, but the number and speed of ISLs connecting each MetroCluster switch to the intermediate switch must be the same on both MetroCluster sites.

In a shared layer 2 network configuration, multiple MetroCluster IP configurations can share the same intermediate network for ISLs. However, this requires adequate capacity and appropriate sizing of the ISLs to ensure low latency and high throughput for data replication. Customers must configure class-maps, policy-maps, and QoS access-maps along the path between MetroCluster IP switches to ensure MetroCluster traffic meets required service levels.

Packet loss must be less than or equal to 0.01%, and the supported jitter value is 3ms for a round trip. The maximum theoretical throughput of shared ISLs varies, based on the switch model and port type. The

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

Beginning with ONTAP 9.9.1, MetroCluster IP configurations can also be implemented with IP-routed (layer 3) backend connections. The MetroCluster backend switches are connected to the routed IP network either directly to routers or through other intervening switches. Only four-node MetroCluster configurations are supported, and dynamic routing is not supported for the MetroCluster traffic. Two subnets are required on each MetroCluster site—one in each network.

Overall, customers must carefully review the requirements and recommendations for their specific configuration to ensure proper networking for their MetroCluster IP deployment.

For more information, please read the following documentation:

- Considerations for ISLs
  https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept_considerations_isls.html
- Considerations for sharing private layer 2 networks
  https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept_considerations_layer_2.html
- Considerations for layer 3 wide-area networks
  https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/concept_considerations_layer_3.html

Operation and administration

Operation and administration for MetroCluster includes checking or validating MetroCluster health and monitoring. For most operations, the ONTAP documentation provides the steps to administer storage that includes MetroCluster. For specifics about MetroCluster feature management and operations, see the MetroCluster Management and Disaster Recovery Guide.

High Availability (HA) and Disaster Recovery (DR)

With MetroCluster, there are two primary types of recovery mechanisms to ensure continuous data availability and efficient system functionality during local or cluster (site) level failures. These mechanisms include local level recovery, referred to as takeover and giveback, and cluster-level recovery, known as switchover and switchback.

Local failures and nondisruptive operations, such as ONTAP upgrades, are managed by the HA partner. MetroCluster uses the standard ONTAP terminology for HA operations.

Local level recovery (HA)

Takeover and giveback processes pertain to high-availability (HA) pairs in MetroCluster IP, ensuring fault tolerance and nondisruptive operations.

- Takeover is the process wherein a node (local node) in an HA pair assumes control of its partner's (partner node) storage during an error or halt in data processing. Various situations, such as software or system failures, power loss, heartbeat message interruptions, or manual initiation, can trigger a takeover. The local node detects the partner's failed status and maintains data availability and system functionality by taking over data processing.
- Giveback refers to the process where the partner node resumes control of its storage after recovering from an issue or completing maintenance. The local node returns the storage to the partner node, restoring normal operations. Takeovers can occur automatically in different situations or be manually initiated using the storage failover takeover command.

For more information about takeover and giveback, review the HA pair management overview.

Site level recovery (DR)

Switchover and switchback pertain to disaster recover (DR) processes in MetroCluster IP to enable one cluster site to take over the tasks of another cluster site, facilitating maintenance or recovery from disasters. ONTAP System Manager 9.6 or later supports these operations.

- Switchover is a process that allows a cluster site (Site A) to take over storage control and client access from another cluster site (Site B). This process ensures nondisruptive operations during testing, maintenance, or site failures. Negotiated switchovers (NSO), which are planned, can be used for disaster recovery testing or planned maintenance. Unplanned switchovers (USO) occur in response to a disaster affecting either of the sites. System Manager determines the feasibility of the switchover and aligns the workload accordingly.

After switchover, System Manager completes the healing process for the MetroCluster IP configuration in two phases. The first phase involves resynchronizing mirrored plexes and switching back root aggregates to the disaster site. The second phase prepares the site for the switchback process.

- Switchback is an operation that returns control of storage and client access from Site A to Site B after maintenance and repairs are completed on Site B. For a successful switchback, certain conditions must be met, including powered-on home nodes and storage shelves in Site B, successful completion of the healing phase, mirrored status for all aggregates in Site A, and the completion of all previous configuration changes before performing a switchback operation.

Beginning with ONTAP 9.5, healing is automated during negotiated switchover operations on MetroCluster IP configurations. Beginning with ONTAP 9.6, automated healing after unplanned switchover is supported. This removes the requirement to issue the `metrocluster heal` commands.

For more information about switchover and switchback, review the documentation about MetroCluster switchover and switchback.

Quorum witness

A quorum witness is essential in storage clusters to maintain high availability, prevent data corruption, and facilitate failover and recovery. By acting as a tiebreaker and ensuring a majority of nodes can communicate with each other, the quorum witness helps avoid split-brain scenarios, where cluster nodes operate independently, leading to data inconsistencies. MetroCluster IP supports either MetroCluster Tiebreaker or ONTAP Mediator as its quorum witness.

ONTAP Mediator software

ONTAP 9.7 includes the release of a new ONTAP Mediator software solution for MetroCluster IP. The software resides in a third failure zone and enables MetroCluster IP to perform automated unplanned switchover (AUSO). Additional functionality includes the disabling of AUSO if the two sites have a failure with mirroring data between them. This prevents automatic switchover if the inter-site links are down, enabling an administrator to decide if it is appropriate to switch over manually.

The new ONTAP Mediator service is configured from one of the MetroCluster IP nodes at one site. ONTAP automatically performs the configuration for all the nodes and for the second cluster. The initial release requires MetroCluster IP running ONTAP 9.7 or later and the ONTAP Mediator software, release 1.0 or later.

New commands are provided in ONTAP for the configuration of ONTAP Mediator.

metrocluster configuration-settings mediator add -mediator-address <mediator-ip>

metrocluster configuration-settings mediator remove

When adding a new mediator, the command prompts for valid mediator credentials. The credentials are set during the installation. They can also be changed by logging into the mediator system and using the following commands:

- Change the mediator account name:

/opt/netapp/lib/ontap_mediator/tools/mediator_change_user

- Change the mediator password:

/opt/netapp/lib/ontap_mediator/tools/mediator_change_password

- Check the mediator status:

systemctl status ontap_mediator

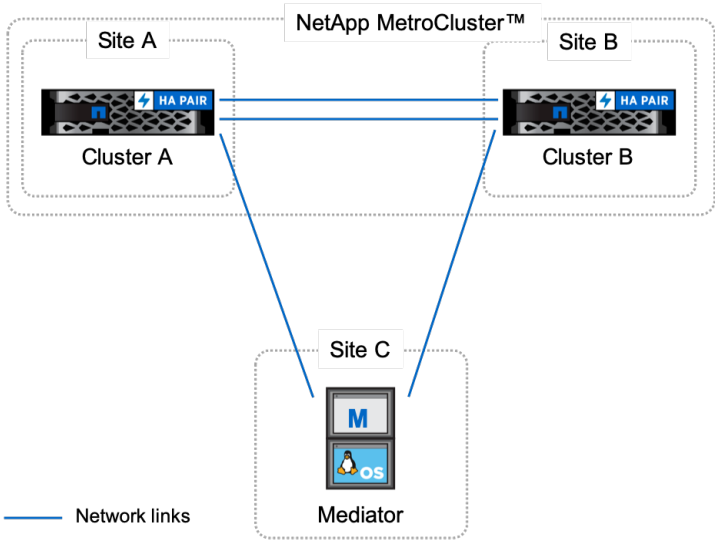
- Run the log collection tool:

/opt/netapp/lib/ontap_mediator/tools/mediator_generate_support_bundle

For more information about the requirements for ONTAP Mediator and details about failures, see the MetroCluster IP Installation and Configuration Guide.

Note: Managing the same MetroCluster configuration with both Tiebreaker and ONTAP Mediator is not supported. Only one of the products can be used to manage a MetroCluster configuration.

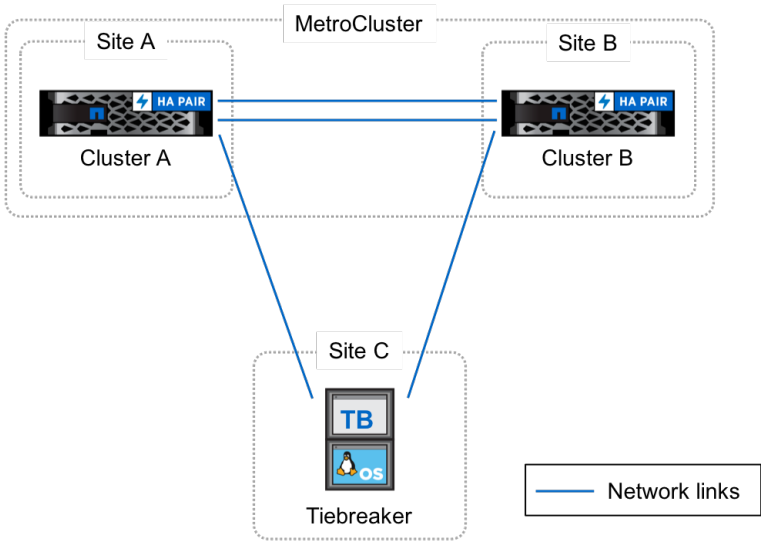
Figure 14) MetroCluster Mediator site.



Tiebreaker software

The MetroCluster Tiebreaker management pack provides monitoring of MetroCluster systems and provides the ability to detect site disasters and ISL failures. Tiebreaker software is installed on a Linux host, typically on a virtual machine, located in a third failure domain separate from the failure domains of either cluster in the MetroCluster solution.

Figure 15) MetroCluster Tiebreaker site.



Tiebreaker software monitors each controller by establishing redundant connections through multiple paths to a node management LIF and to the cluster management LIF.

Note: Managing the same MetroCluster configuration with both Tiebreaker and ONTAP Mediator is not supported. Only one of the products can be used to manage a MetroCluster configuration.

Tiebreaker site failure symptoms

During a site failure, when one cluster is unreachable from the Tiebreaker software and the other cluster is reachable, the cluster that is reachable must also indicate it has lost communication with the partner cluster before Tiebreaker software triggers an alert. If the two clusters can still communicate, Tiebreaker identifies the loss of connectivity in the network between the Tiebreaker software and the cluster that is not reachable.

Figure 16) Tiebreaker site link failure.

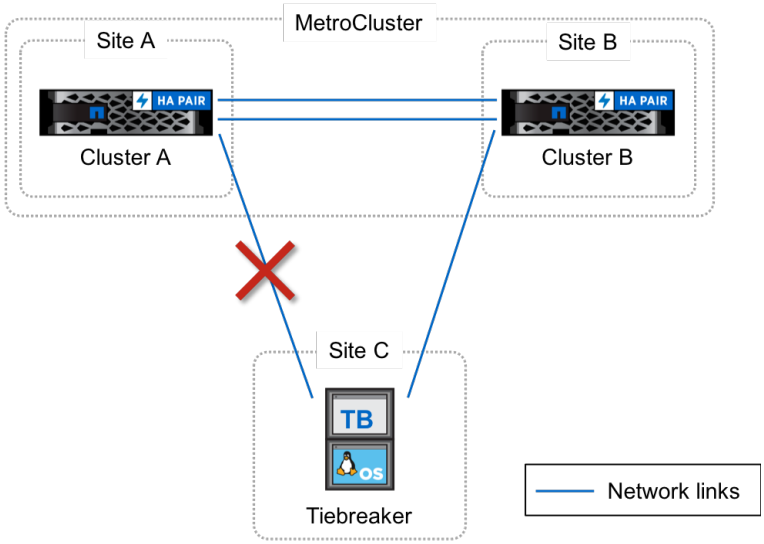
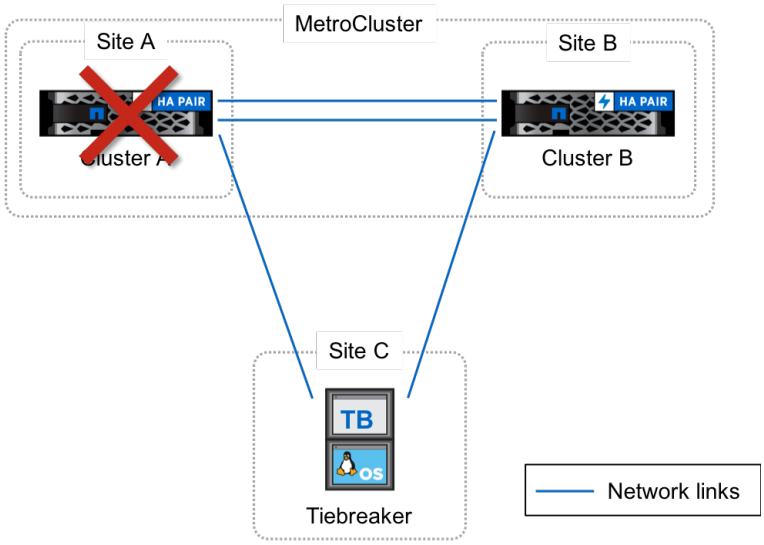


Figure 17) Tiebreaker site failure.



For more information, see the Tiebreaker Software Installation and Configuration Guide.

Interoperability

MetroCluster supports the most common ONTAP features. However, some ONTAP features, such as SnapMirror Synchronous, are not currently supported by MetroCluster. You can see the ONTAP documentation of the feature for guidance on support for MetroCluster.

SnapMirror Asynchronous

SnapMirror Asynchronous is supported with MetroCluster. SnapMirror Asynchronous protection provides scheduled, block replication, mirror protection, between source and destination volumes to a tertiary cluster. MetroCluster systems can be configured as a source and destination for SnapMirror replication relationships. See the following resources for more information.

- Asynchronous SnapMirror disaster recovery basics
- TR-4015: SnapMirror configuration and best practices

NetApp ONTAP FlexGroup volumes

NetApp FlexGroup volumes are supported with MetroCluster starting with ONTAP 9.6. A FlexGroup volume is a scale-out volume that provides high performance along with automatic load distribution and scalability. See the following resources for more information.

- FlexGroup volumes management
- TR-4571: NetApp ONTAP FlexGroup volumes — Best practices and implementation guide

NetApp FlexCache

NetApp FlexCache® technology is supported with MetroCluster IP starting with ONTAP 9.7. FlexCache is a remote caching capability that simplifies file distribution, reduces WAN latency, and lowers WAN bandwidth costs. Cache only the actively-read data, rather than entire files or volumes, either locally within a datacenter or geographically dispersed at remote sites. See the following resources for more information.

- FlexCache volumes management
- TR-4743: FlexCache in NetApp ONTAP

NetApp FabricPool

NetApp FabricPool is supported with MetroCluster IP starting with ONTAP 9.7. FabricPool is a hybrid storage solution in ONTAP that uses an all-flash (SSD) aggregate as a performance tier and an object store in a public cloud service as a cloud tier. See the following resources for more information:

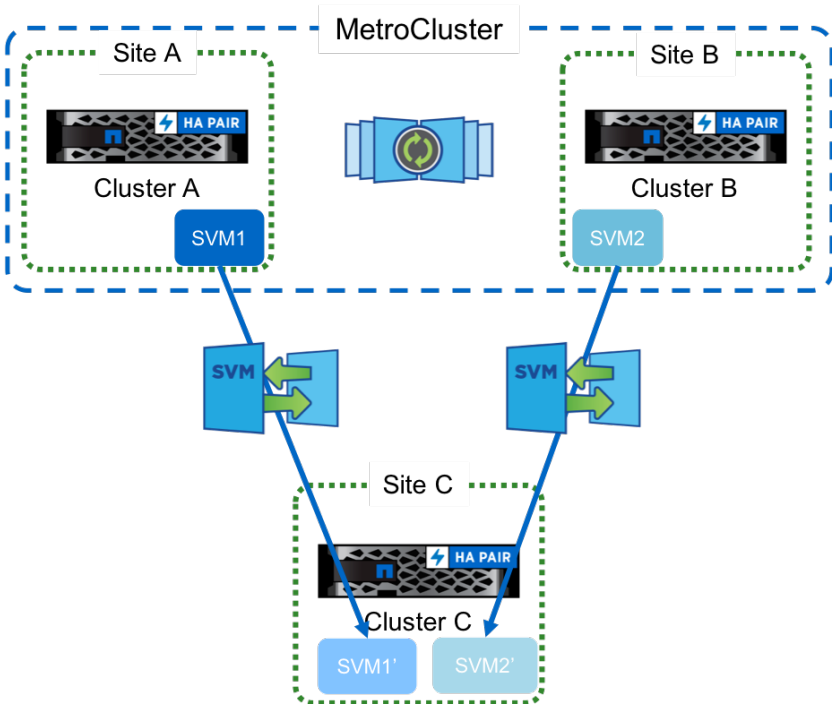
- FabricPool Product Documentation
- TR-4598: FabricPool Best Practices

SVM Disaster Recovery (DR)

Starting from ONTAP 9.5, SVM DR is now compatible with MetroCluster IP. SVM DR is a solution designed to offer data protection and disaster recovery capabilities for SVMs. This feature enables administrators to create a replica of the primary SVM, ensuring an up-to-date copy of the data and configuration is available at a remote site to mitigate against disasters. This is achieved by utilizing SnapMirror asynchronous technology to establish a relationship between a source SVM on a primary site and a destination SVM residing on a DR site. In case of a disaster, the secondary SVM can be activated to provide access to data, thereby minimizing downtime.

Note: With the introduction in ONTAP 9.5, SVMs could only be protected from one site of the MetroCluster configuration. ONTAP 9.11.1 introduced the ability to protect SVMs from both sites of a MetroCluster.

Figure 18) SVM disaster recovery.



Where to find additional information

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

- MetroCluster IP Installation and Configuration Guide
  https://docs.netapp.com/us-en/ontap-metrocluster/install-ip/index.html
- MetroCluster Disaster Recovery Guide
  https://docs.netapp.com/us-en/ontap-metrocluster/disaster-recovery/concept_dr_workflow.html
- ONTAP Product Documentation
  https://docs.netapp.com/us-en/ontap-family/
- TR-4705: NetApp MetroCluster Solution Architecture and Design
  https://www.netapp.com/pdf.html?item=/media/13480-tr4705pdf.pdf
- TR-4375: NetApp MetroCluster FC
  https://www.netapp.com/pdf.html?item=/media/13482-tr4375pdf.pdf
- MetroCluster IP Technical FAQ (NetApp Field Portal; login required)
  https://fieldportal.netapp.com/content/748972
- NetApp Interoperability Matrix Tool
  https://imt.netapp.com/matrix/#welcome
- NetApp MetroCluster Documentation Resources
  https://www.netapp.com/support-and-training/documentation/metrocluster/
- TR-4592: Oracle on MetroCluster
  https://www.netapp.com/pdf.html?item=/media/8583-tr4592pdf.pdf
- VMware vSphere 5.x, 6.x and 7.x support with NetApp MetroCluster (2031038)
  https://kb.vmware.com/s/article/2031038
- TR-4128: vSphere 6 on NetApp MetroCluster 8.3
  https://fieldportal.netapp.com/content/252106 (login required)

For more information about cabling and optical modules, see the following MetroCluster IP switch technical references.

- MetroCluster IP 10/25GbE Broadcom BES-53248 Switch (NetApp Field Portal; login required)
  https://fieldportal.netapp.com/content/886413
- MetroCluster IP 40GbE Cisco N3132Q-V Switch (NetApp Field Portal; login required)
  https://fieldportal.netapp.com/content/729700
- MetroCluster IP 100GbE Cisco N3232C and N9336C Switches (NetApp Field Portal; login required)
  https://fieldportal.netapp.com/content/757495

For more information about supported optical modules and part numbers to order from a Cisco partner, see the Cisco optical module support matrix for the specific model of switch.

- Cisco module and switch support matrix
  https://tmgmatrix.cisco.com/home
- Cisco CWDM SFP 10 Gigabit Ethernet Solution Datasheet
  https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/datasheet-c78-734047.html
- Cisco 10GBASE SFP+ Modules Datasheet
  https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/data_sheet_c78-455693.html
- Cisco 10GBASE Dense Wavelength-Division Multiplexing SFP+ Modules Datasheet
  https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/dwdm-transceiver-modules/data_sheet_c78-711186.html

Version history

Version	Date	Document version history
Version 1.1	February 2019	Includes updates for ONTAP 9.5
Version 1.2	May 2019	Updates for ONTAP 9.6
Version 1.3	November 2019	Updates for ONTAP 9.7
Version 1.4	November 2020	Updates for ONTAP 9.8
Version 1.5	May 2023	Updates for ONTAP 9.12.1

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.

TR-4689-0523

NetApp