Dell EMC Metro node 7.0.1

Product Guide

7.0.1

July 2021

DELL Technologies

Notes, cautions, and warnings

NOTE: A NOTE indicates important information that helps you make better use of your product.

CAUTION: A CAUTION indicates either potential damage to hardware or loss of data and tells you how to avoid the problem.

WARNING: A WARNING indicates a potential for property damage, personal injury, or death.

Contents

Figures

Tables

1

Preface

As part of an effort to improve its product lines, Dell EMC periodically releases revisions of its software and hardware. Therefore, some functions described in this document might not be supported by all versions of the software or hardware currently in use. The product release notes provide the most up-to-date information on product features.

Contact your Dell EMC technical support professional if a product does not function properly or does not function as described in this document.

NOTE: This document was accurate at publication time. Go to Dell EMC Online Support (https://www.dell.com/support) to ensure that you are using the latest version of this document.

Purpose

This document is part of the VPLEX documentation set, and describes the VPLEX features and use cases, configuration options, VPLEX software and its upgrade, and the hardware overview.

Audience

This guide is intended for use by customers who wish to understand the software and hardware features of VPLEX, the use cases of VPLEX, product offerings, and the configuration options.

Related documents (available on Dell EMC Online Support and SolVe) include:

- *Release Notes for the metro node appliance*
- *Product Guide or the metro node appliance*
- *Metro node Hardware Installation Guide*
- *Configuration and Installation Guide for the metro node appliance*
- *Security Configuration Guide for the metro node appliance*
- *CLI Reference Guide for the metro node appliance*
- *Administration Guide for the metro node appliance*
- Online Help for the metro node appliance
- *REST API v2 for the metro node appliance*
- *Open Source Licenses Guide for the metro node appliance*
- *Hardware Reference Guide for the metro node appliance*
- Procedures provided through the SolVe

Special notice conventions used in this document

Dell EMC uses the following conventions for special notices:

CAUTION: Indicates a hazardous situation which, if not avoided, will result in death or serious injury.

CAUTION: Indicates a hazardous situation which, if not avoided, could result in death or serious injury.

CAUTION: Indicates a hazardous situation which, if not avoided, could result in minor or moderate injury.

NOTE: Addresses practices not related to personal injury.

NOTE: Presents information that is important, but not hazard-related.

Typographical conventions

Dell EMC uses the following type style conventions in this document:

Table 1. Typographical conventions

Bold	Used for names of interface elements, such as names of windows, dialog boxes, buttons, fields, tab names, key names, and menu paths (what the user specifically selects or clicks)
italic	Used for full titles of publications referenced in text
`Monospace`	Used for: • System code • System output, such as an error message or script • Pathnames, filenames, prompts, and syntax • Commands and options
Monospace italic	Used for variables
`Monospace bold`	Used for user input
[ ]	Square brackets enclose optional values
|	Vertical bar indicates alternate selections - the bar means "or"
{ }	Braces enclose content that the user must specify, such as x or y or z
...	Ellipses indicate nonessential information omitted from the example

Where to get help

Dell EMC support, product, and licensing information can be obtained as follows:

Product information

For documentation, release notes, software updates, or information about Dell EMC products, go to Dell EMC Online Support at https://www.dell.com/support.

Technical support

Go to Dell EMC Online Support and click Support. You will see several options for contacting Dell EMC Technical Support. Note that to open a service request, you must have a valid support agreement. Contact your Dell EMC sales representative for details about obtaining a valid support agreement or with questions about your account.

Online communities

Visit Dell EMC Community Network (DECN) at https://www.dell.com/community/Dell-Community/ct-p/English for peer contacts, conversations, and content on product support and solutions. Interactively engage online with customers, partners, and certified professionals for all Dell EMC products.

Your comments

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications. Send your opinions of this document to vplex.doc.feedback@dell.com.

2

Introducing metro node

This chapter introduces the metro node feature.

Topics:

Metro node overview

Metro node virtualizes the data that is on storage arrays to create dynamic, distributed, and highly available data centers.

Use metro node to:

- Move data non-disruptively between Dell EMC storage arrays and non-Dell EMC storage arrays without any downtime for the host.

  Metro node moves data transparently, and the virtual volumes retain the same identities and the same access points to the host. There is no need to reconfigure the host.

- Protect data in the event of disasters or failure of components in your data centers.

  With metro node, you can withstand failures of storage arrays, cluster components, an entire site failure, or loss of communication between sites (when two clusters are deployed) and still keep applications and data online and available.

With metro node, you can transform the delivery of IT to a flexible, efficient, reliable, and resilient service.

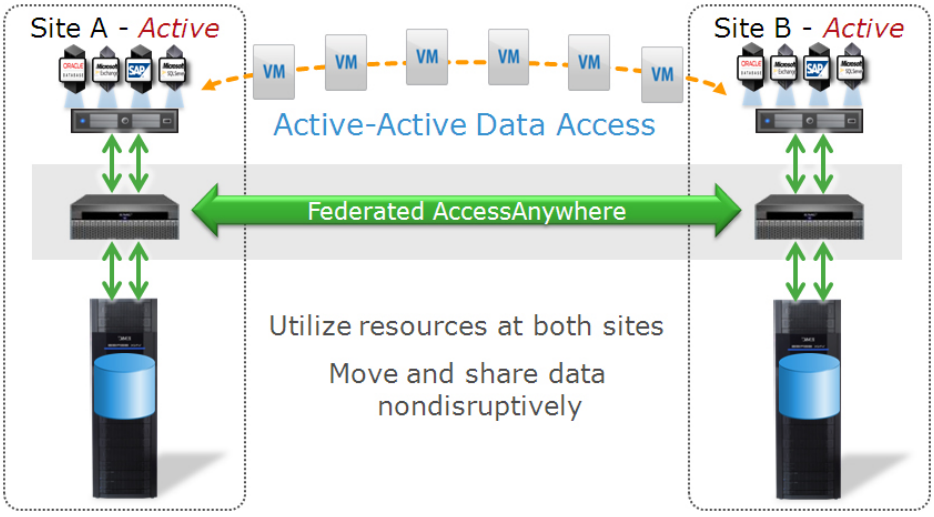

Figure 1. Metro node active-active

Metro node addresses these two primary IT needs:

- **Mobility**: Metro node moves applications and data between different storage installations:
  - Within the same data center or across a campus (metro node Local)
  - Within a geographical region (metro node Metro)
- **Availability:** Metro node creates high-availability storage infrastructure across these same varied geographies with unmatched resiliency.

Metro node offers the following unique innovations and advantages:

- Metro node distributed/federated virtual storage enables new models of application and Data Mobility.

  Metro node is optimized for virtual server platforms (VMware ESX, Hyper-V, Oracle Virtual Machine, AIX VIOS).

  Metro node can streamline or accelerate transparent workload relocation over distances, including moving virtual machines.

- In a Metro configuration, metro node AccessAnywhere provides image consistent active-active access to data across two metro node clusters.

Metro node pools the storage resources in multiple data centers so that the data can be accessed anywhere. With metro node, you can:

- Provide continuous availability and workload mobility.
- Replace your tedious data movement and technology refresh processes with metro node's patented simple, frictionless two-way data exchange between locations.
- Create an active-active configuration for the active use of resources at both sites.
- Provide instant access to data between data centers. metro node allows simple, frictionless two-way data exchange between locations.
- Combine metro node with virtual servers to enable private and hybrid cloud computing.

Metro node product family

The metro node product family includes:

- Metro node Local
- Metro node Metro

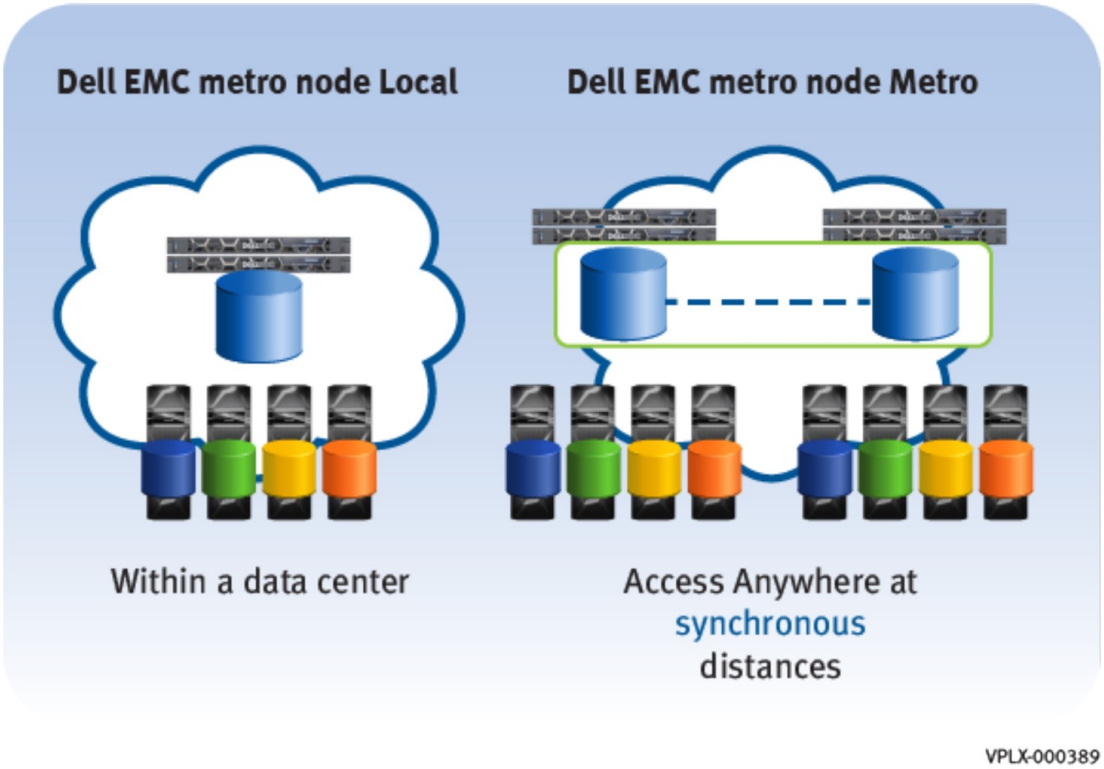

Figure 2. Metro node family: Local and Metro

Metro node Local

Metro node Local consists of a single cluster. Metro node Local:

- Federates Dell EMC storage arrays with non-Dell EMC storage arrays.

Federation allows transparent data mobility between arrays for simple, fast data movement and technology refreshes.

- Standardizes LUN presentation and management using simple tools to provision and allocate virtualized storage devices.
- Improves storage utilization using pooling and capacity aggregation across multiple arrays.
- Increases protection and high availability for critical applications.

  Mirrors storage across mixed platforms without host resources.

  Leverage your existing storage resources to deliver increased protection and availability for critical applications.

Deploy metro node Local within a single data center.

Metro node Metro

Metro node Metro consists of two metro node clusters connected by inter-cluster links with not more than 10ms Round Trip Time (RTT). Metro node Metro:

- Transparently relocates data and applications over distance, protects your data center against disaster.

  Manage all of your storage in both data centers from one management interface.

- Mirrors your data to a second site, with full access at near local speeds.

Deploy metro node Metro within a data center for:

- Additional virtual storage capabilities beyond that of a metro node Local.
- Higher availability.

  Metro clusters can be placed up to 100 km apart, allowing them to be located at opposite ends of an equipment room, on different floors, or in different fire suppression zones; all of which might be the difference between riding through a local fault or fire without an outage.

Deploy metro node Metro between data centers for:

- Mobility: Redistribute application workloads between the two data centers.
- Availability: Applications must keep running in the presence of data center failures.
- Distribution: One data center lacks space, power, or cooling.

Combine metro node Metro virtual storage and virtual servers to:

- Transparently move virtual machines and storage across synchronous distances.
- Improve utilization and availability across heterogeneous arrays and multiple sites.

Distance between clusters is limited by physical distance, by host, and by application requirements. Metro node Metro clusters contain additional I/O modules to enable the inter-cluster WAN communication over IP.

Metro node hardware platforms

Metro node hardware platform is based on the Dell R640 PowerEdge server.

Configuration highlights

A metro node primarily consists of:

- Two hardware nodes
- The two nodes are directly and redundantly connected using two Cat6 shielded cables for management connectivity and two Dell Direct-Attach-Copper (DAC) cables with SFP plugs for local data com connectivity.
- Two 32 Gig, 2 port FC HBA for FE and BE connectivity.
- The metro node management stack runs on each node in the cluster. Each node has a public Ethernet port which provides cluster management services when connected to the network.

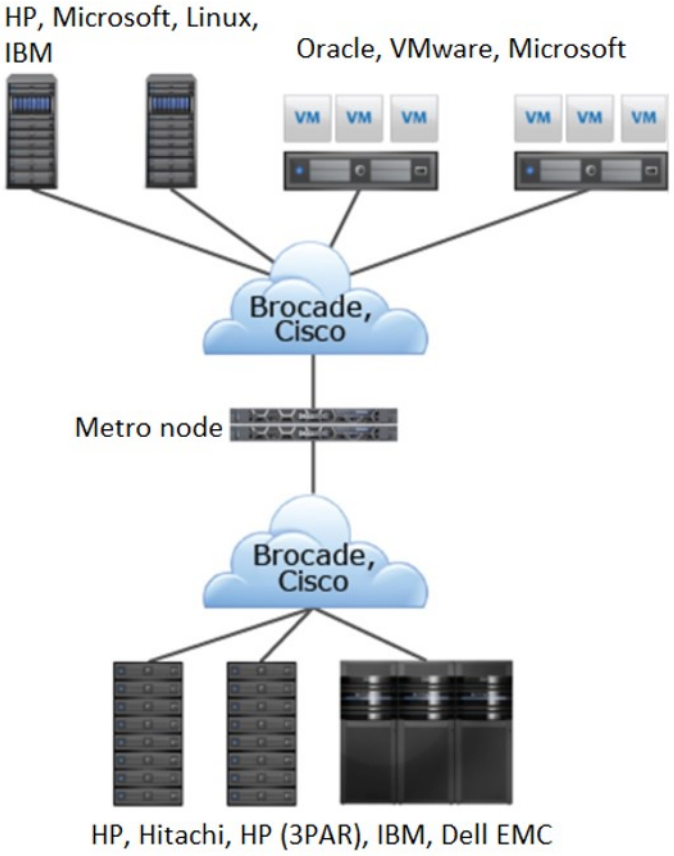

Figure 3. Configuration highlights

Metro node conforms to established world wide naming (WWN) guidelines that can be used for zoning. It also supports Dell EMC storage and arrays from other storage vendors, such as HDS, HP, and IBM. Metro node provides storage federation for operating systems and applications that support clustered file systems, including both physical and virtual server environments with VMware, ESX, and Microsoft Hyper-V. The network fabrics from Brocade and Cisco are supported in metro node.

See the *Dell EMC Simple Support Matrix*, available at http://elabnavigator.EMC.com under the Simple Support Matrix tab.

Management interfaces

In a metro node Metro configuration, both clusters can be managed from any node.

Inside metro node clusters, management traffic traverses a TCP/IP based private management network.

In a metro node Metro configuration, the management traffic between the clusters is secured through https protocol.

Web-based GUI

Metro node web-based user interface (UI) provides an easy-to-use point-and-click management interface.

The following figures show the screen to claim storage:

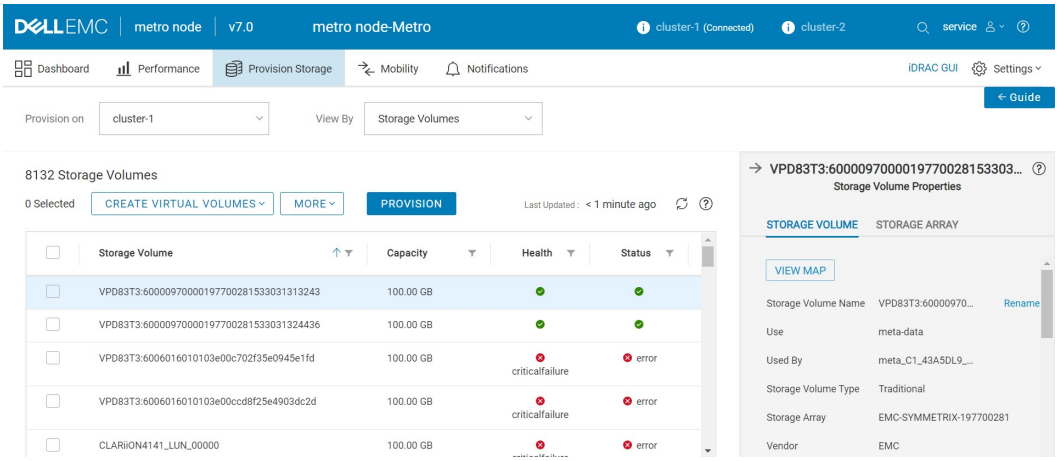

Figure 4. Claim storage using the GUI (for HTML5)

The UI supports most of the metro node operations, and includes Dell EMC Online help for metro node to assist new users in learning the interface.

Metro node operations that are not available in the GUI, are supported by the Command Line Interface (CLI), which supports full functionality.

VPlexcli

The VPlexcli supports all metro node operations.

The CLI is divided into command contexts:

- Global commands are accessible from all contexts.
- Other commands are arranged in a hierarchical context tree, and can be executed only from the appropriate location in the context tree.

The following example shows a CLI session that performs the same tasks as shown in Figure.

Example 1 Claim storage using the CLI:

In the following example, the `claimingwizard` command finds unclaimed storage volumes, claims them as thin storage, and assigns names from a PowerStore hints file:

```
VPlexcli:/clusters/cluster-2/storage-elements/storage-volumes> claimingwizard --file /
home/service/power_store_claim.txt --thin-rebuild
Found unclaimed storage-volume VPD83T3:68ccf0980001c3a3bf416ca07c8c71b5 vendor DellEMC :
claiming and naming power_store_claim_name_2.

Found unclaimed storage-volume VPD83T3:68ccf0980000bb040913bbe50c408407 vendor DellEMC :
claiming and naming power_store_claim_name_1.

Claimed 2 storage-volumes in storage array power_store_claim

Claimed 2 storage-volumes in total.


VPlexcli:/clusters/cluster-2/storage-elements/storage-volumes>
```

The *Dell EMC CLI Guide for the metro node* provides a comprehensive list of metro node commands and detailed instructions on using those commands.

3

Metro node use cases

This chapter describes the general features, benefits, and the important use cases of metro node.

Topics:

General use cases and benefits

The following table summarizes the general metro node use cases and their benefits.

Table 2. General metro node use cases and benefits

General use cases	Benefits
Mobility	• **Migration:** Move data and applications without impact on users. • **Virtual Storage federation:** Achieve transparent mobility and access within a data center and between data centers.
Availability	• **Resiliency:** Mirror across arrays within a single data center or between data centers without host impact. This increases availability for critical applications.

For all metro node deployments, do the following:

- Presents storage volumes from back-end arrays to metro node engines.
- Federates the storage volumes into hierarchies of metro node virtual volumes with user-defined configuration and protection levels.
- Presents virtual volumes to production hosts in the SAN through the metro node front-end.

Mobility

Use metro node to move data between data centers, relocate a data center or consolidate data, without disrupting host application access to the data.

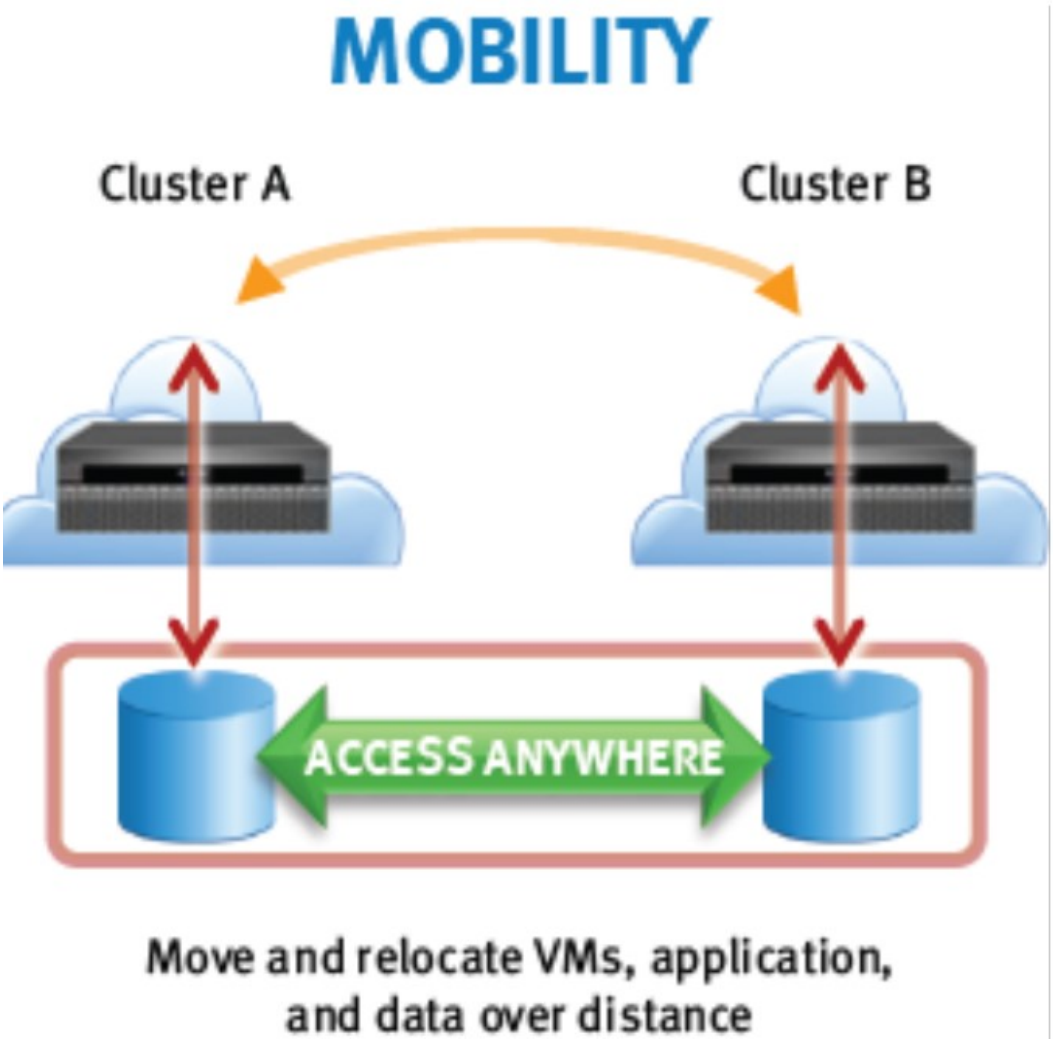

Figure 5. Moving data with metro node

The source and target arrays can be in the same data center (metro node Local) or in different data centers separated by up to 10ms (metro node Metro). The source and target arrays can be heterogeneous.

When you use metro node to move data, the data retains its original metro node volume identifier during and after the mobility operation. No change in volume identifiers eliminates application cut over. The application continues to use the same data, though the data has been moved to a different storage array.

There are many types and reasons to move data:

- Move data from a hot storage device.
- Move the data from one storage device to another without moving the application.
- Move operating system files from one storage device to another.
- Consolidate data or database instances.
- Move database instances.
- Move storage infrastructure from one physical location to another.

With metro node, you no longer need to spend significant time and resources preparing to move data and applications. You do not have to plan for an application downtime or restart the applications as part of the data movement activity. Instead, a move can be made instantly between sites, over distance, and the data remains online and available during the move without any outage or downtime. Considerations before moving the data include the business impact, type of data to be moved, site locations, total amount of data, and schedules.

The data mobility feature of metro node is useful for disaster avoidance, planned upgrade, or physical movement of facilities. The mobility jobs in metro node are as follows:

Table 3. Types of data mobility operations

Device	Moves data from one device to another device (within a cluster and across clusters).
Batch	Moves data using a migration plan file. Create batch migrations to automate routine tasks. ● Use batched device migrations to migrate to dissimilar arrays and to migrate devices within a cluster and between the clusters in a metro node Metro configuration.

Technology refresh

In typical IT environments, migrations to new storage arrays (technology refreshes) require that the data that is being used by hosts be copied to a new volume on the new array. The host must then be reconfigured to access the new storage. This process requires downtime for the host.

Metro node makes it easier to replace heterogeneous storage arrays on the back-end. Migrations between heterogeneous arrays can be complicated and may require additional software or functionality. Integrating heterogeneous arrays in a single environment is difficult and requires a staff with a diverse skill set.

When metro node is inserted between the front-end and back-end redundant fabrics, metro node appears as the target to hosts and as the initiator to storage.

The data resides on virtual volumes in metro node and it can be copied nondisruptively from one array to another without any downtime. There is no need to reconfigure the host; the physical data relocation is performed by metro node transparently and the virtual volumes retain the same identities and the same access points to the host.

In the following figure, the virtual disk is made up of the disks of Array A and Array B. The site administrator has determined that Array A has become obsolete and should be replaced with a new array. Array C is the new storage array. Using Mobility Central, the administrator:

- Adds Array C into the metro node cluster.
- Assigns a target extent from the new array to each extent from the old array.
- Instructs metro node to perform the migration.

Metro node copies data from Array A to Array C while the host continues its access to the virtual volume without disruption.

After the copy of Array A to Array C is complete, Array A can be decommissioned:

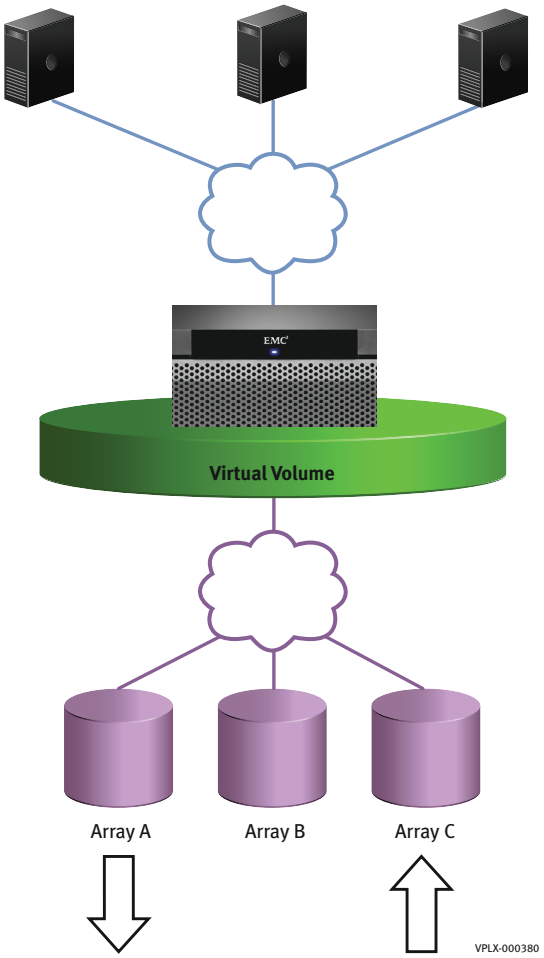

Figure 6. Metro node technology refresh

Because the virtual machine is addressing its data to the abstracted virtual volume, its data continues to flow to the virtual volume without any need to change the address of the data store.

Although this example uses virtual machines, the same is true for traditional hosts. Using metro node, the administrator can move data used by an application to a different storage array without the application or server being aware of the change.

This allows you to change the back-end storage arrays transparently, without interrupting I/O.

Availability

Metro node features allow the highest possible resiliency in the event of an outage. The following figure shows a metro node Metro configuration where storage has become unavailable at one of the cluster sites.

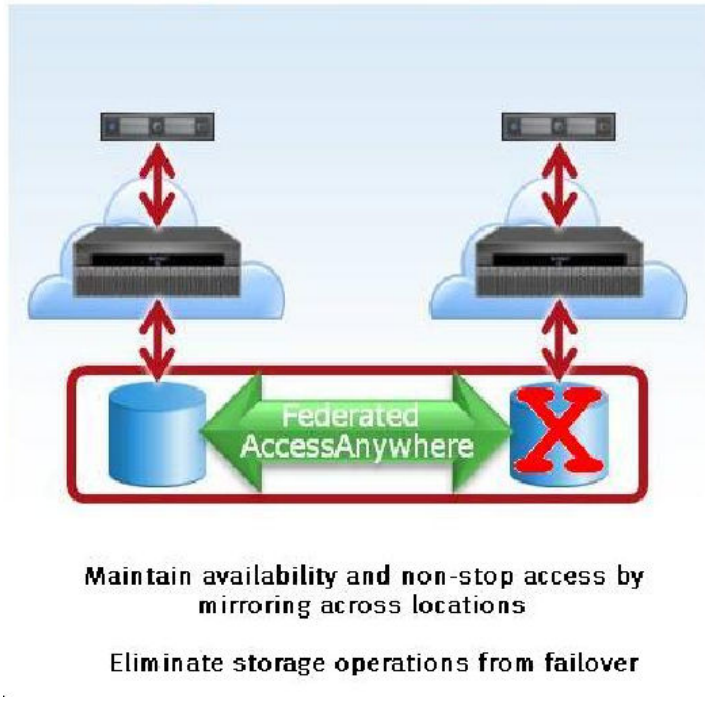

Figure 7. High availability infrastructure example

Metro node redundancy provides zero Recovery Time Objective (RTO) and Recovery Point Objective (RPO) because, metro node AccessAnywhere mirrors all data, applications continue without disruption using the back-end storage at the unaffected site.

With the Federated AccessAnywhere feature of metro node, the data remains consistent, online, and always available. Metro node does not need to ship the entire file back and forth like other solutions. It only sends the changed updates as they are made, greatly reducing bandwidth costs and offering significant savings over other solutions.

To know more about high availability with metro node, see Chapter 4 Integrity and resiliency. .

4

Features in metro node

This chapter describes the specific features of metro node.

Topics:

Metro node security features

The operating systems of the metro node management server and the directors are based on a Novell SUSE Linux Enterprise Server 15 SP1 distribution.

The operating system has been configured to meet Dell EMC security standards by disabling or removing unused services, and protecting access to network services through a firewall.

The metro node security features include:

- LDAP authentication using OS secure SSSD service
- HTTPS to access the metro node UI
- HTTPS inter-cluster link in a metro node Metro configuration
- SCP to copy files
- Support for separate networks for all metro node cluster communication
- Defined user accounts and roles
- Certificate Authority (CA) certificate (default expiration 5 years)
- Two host certificates (default expiration 2 years)
- External directory server support

  CAUTION: The WAN-COM inter-cluster link carries unencrypted user data. To ensure privacy of the data, establish an encrypted VPN tunnel between the two sites.

For more information about security features and configuration see the *Dell EMC Security Configuration Guide for metro node*.

ALUA

Asymmetric Logical Unit Access (ALUA) routes I/O of the LUN directed to non-active/failed storage processor to the active storage processor without changing the ownership of the LUN.

Each LUN has two types of paths:

- **Active/optimized paths** are direct paths to the storage processor that owns the LUN.

  Active/optimized paths are usually the optimal path and provide higher bandwidth than active/non-optimized paths.

- **Active/non-optimized paths** are indirect paths to the storage processor that does not own the LUN through an interconnect bus.

  I/Os that traverse through the active/non-optimized paths must be transferred to the storage processor that owns the LUN. This transfer increases latency and has an impact on the array.

Metro node detects the different path types and performs round robin load balancing across the active/optimized paths.

Metro node supports all three flavors of ALUA on backend arrays:

- **Explicit ALUA** - The storage processor changes the state of paths in response to commands (for example, the Set Target Port Groups command) from the host (the metro node backend).

  The storage processor must be explicitly instructed to change a path's state.

  If the active/optimized path fails, metro node issues the instruction to transition the active/non-optimized path to active/optimized.

  There is no need to failover the LUN.

- **Implicit ALUA** - The storage processor can change the state of a path without any command from the host (the metro node back end).

  If the controller that owns the LUN fails, the array changes the state of the active/non-optimized path to active/optimized and fails over the LUN from the failed controller.

  On the next I/O, after changing the path's state, the storage processor returns a Unit Attention "Asymmetric Access State Changed" to the host (the metro node backend).

  Metro node then re-discovers all the paths to get the updated access states.

- **Implicit/explicit ALUA** - Either the host or the array can initiate the access state change.

Storage processors support implicit only, explicit only, or both.

Provisioning with metro node

Metro node allows easy storage provisioning among heterogeneous storage arrays. Use the web-based UI to simplify everyday provisioning or create complex devices.

There are two ways to provision storage in metro node:

- EZ provisioning
- Advanced provisioning

All provisioning features are available in the Unisphere for metro node UI, CLI, and REST API.

Support for thin volumes and unmapping

Thin provisioning advertises the metro node virtual volumes as thin volumes to the hosts. Thin provisioning dynamically allocates block resources only when they are required. It essentially allows efficient utilization of physical block resources from the storage arrays.

Hosts gather the properties related to the thin provisioning feature of a metro node virtual volume and send SCSI commands to free storage block resources that are not in use. If the blocks of the back end storage volumes are free, the blocks can be mapped to other changed regions. Thin provisioning enables dynamic freeing of storage blocks on storage volumes for which thin provisioning is supported.

NOTE: The *Dell EMC Simplified Support Matrix for metro node* provides more information on the supported storage volumes.

Metro node thin provisioning support includes the following features:

- Discovery of the back-end storage volumes capable for thin provisioning - During the back-end storage volume discovery, metro node gathers all thin provisioning related storage volume properties. Metro node also performs a consistency check on all the properties related to thin-provisioning.
- Reporting thin provisioning enabled metro node virtual volumes to hosts - metro node shares the details of the thin provisioning-enabled virtual volumes with the hosts.
- Reclaiming the unused storage blocks - Through a command, metro node removes the mapping between a deleted virtual machine and its storage volumes and reclaims the storage blocks corresponding to the VMFS blocks used by that virtual machine.
- Handling storage exhaustion - The exhaustion of storage blocks on non-mirrored storage volumes are notified to the host as a space allocation failure. This error notification is posted to the host and the VMware hosts stop the impacted virtual machine.

To prevent potential mapping of all the blocks in the storage volumes that are thin capable, metro node uses thin rebuilds. Thin rebuilds can be configured to be set or unset for any claimed storage volume on which metro node builds virtual volumes. This property controls how metro node does its mirror rebuilding.

The unmap feature reclaims the unused VMFS blocks by removing the mapping between the logical blocks and the physical blocks. This essentially removes the link between a logical block and a physical block that has unknown or unused resources.

Performance monitoring

Metro node performance monitoring provides a customized view into the performance of your system. You decide which aspects of the system's performance to view and compare.

You can view and assess the metro node performance using these methods:

- Storage M&R for VPLEX
- Storage Resource Manager (SRM)
- Unisphere Performance Monitoring Dashboard, which shows real-time performance monitoring data for up to one hour of history.
- Performance statistics collection using the CLI and the API. These methods let you collect and view the statistics, and export them to an external application for analysis.

Unisphere Performance Monitoring Dashboard

The Unisphere Performance Monitoring Dashboard supports these general categories of performance monitoring:

- Current load monitoring that allows administrators to watch CPU load during upgrades, I/O load across the inter-cluster WAN link, and front-end against the back-end load during data mining or back up.
- Long term load monitoring that collects data for capacity planning and load balancing.
- Object-based monitoring that collects data for the virtual volume.

The Unisphere Performance Monitoring Dashboard is a customized view into the performance of the metro node system:

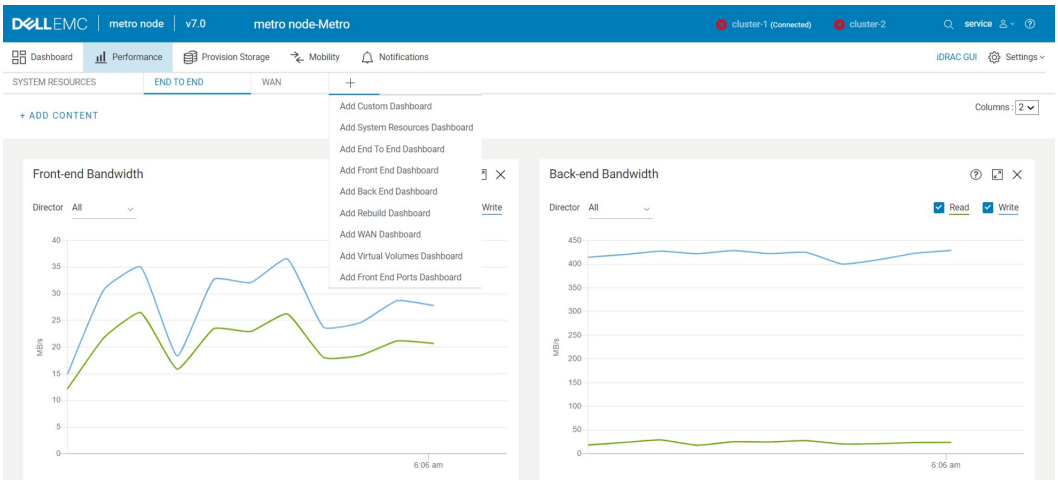

Figure 8. Unisphere Performance Monitoring Dashboard (for HTML5)

You decide which aspects of the system's performance to view and compare:

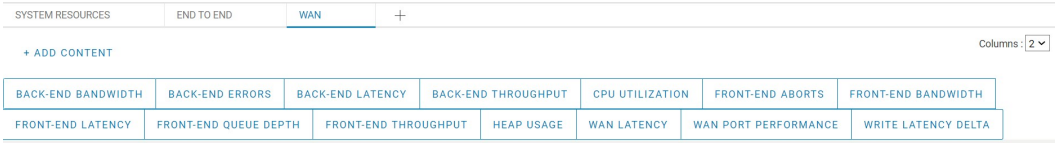

Figure 9. Unisphere Performance Monitoring Dashboard - select information to view (for HTML5)

Performance information is displayed as a set of charts. For example, the following figure shows front-end throughput for all directors (for HTML5):

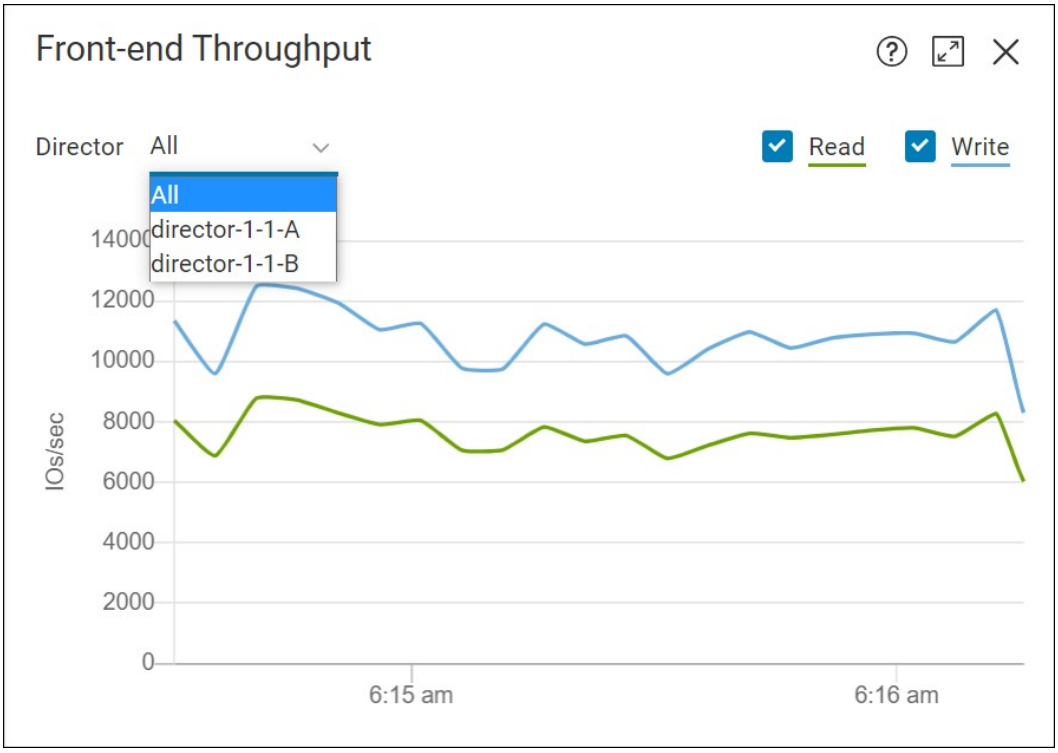

Figure 10. Unisphere Performance Monitoring Dashboard - sample chart (for UI)

For additional information about the statistics available through the Performance Monitoring Dashboard, see the Online Help for the metro node appliance available in the metro node UI.

Performance monitoring using the CLI

The CLI supports current load monitoring, long term load monitoring, object base monitoring, and troubleshooting monitoring. The CLI collects and displays performance statistics using:

monitors - Gather the specified statistic from the specified target at the specified interval.

monitor sinks - Direct the output to the desired destination. Monitor sinks include the console, a file, or a combination of the two.

Use two pre-defined perpetual monitors for each director to collect information to diagnose common problems.

Use the CLI to create a toolbox of custom monitors to operate under varying conditions including debugging, capacity planning, and workload characterization. For example:

The *Dell EMC Administration Guide for metro node* describes the procedure for monitoring metro node performance using the CLI.

Notification

Events provide information about changes happening to the system, which also indicates that there is a problem with the system. Alerts are events that require attention by the system administrator or user. Most alerts indicate that there is a problem with the system that must be rectified to attain the best performance from the system.

The metro node notifications system displays live and historical alerts for platform, hardware (Both iDRAC and metro node Monitor alerts) in the Notification pane that requires user attention and helps to monitor the state of the various components, triage, and troubleshooting issues.

Notification features also allows to send alert notifications to a specified email or SMTP server. To configure SMTP server, see *System Configuration guide*.

5

Integrity and resiliency

This chapter describes how the high availability and the redundancy features of metro node provide robust system integrity and resiliency.

Topics:

About metro node resilience and integrity

With metro node, you get true high availability. Operations continue and data remains online even when a failure occurs. Within synchronous distances (metro node Metro), think of metro node as providing disaster avoidance instead of just disaster recovery.

Metro node Metro provides shared data access between sites. The same data (not a copy), exists at more than one location simultaneously. metro node can withstand a component failure, a site failure, or loss of communication between sites and still keep the application and data online and available. Metro node clusters are capable of surviving any single hardware failure in any subsystem within the overall storage cluster, including host connectivity and memory subsystems. A single failure in any subsystem does not affect the availability or integrity of the data.

Metro node redundancy creates fault tolerance for devices and hardware components that continue operation as long as one device or component survives. This highly available and robust architecture can sustain multiple device and component failures without disrupting service to I/O.

Failures and events that do not disrupt I/O include:

- Unplanned and planned storage outages
- SAN outages
- Metro node component failures
- Metro node cluster failures
- Data center outages

To achieve high availability, you must create redundant host connections and supply hosts with multi path drivers.

NOTE: In the event of a front-end port failure or a director failure, hosts without redundant physical connectivity to a metro node cluster and without multi-pathing software installed could be susceptible to data unavailability.

Site distribution

When two metro node clusters are connected together with metro node Metro, metro node gives you shared data access between sites. Metro node can withstand a component failure, a site failure, or loss of communication between sites and still keep the application and data online and available.

Metro node Metro ensures that if a data center goes down, or even if the link to that data center goes down, the other site can continue processing the host I/O.

In the following figure, despite a site failure at Data Center B, I/O continues without disruption in Data Center A.

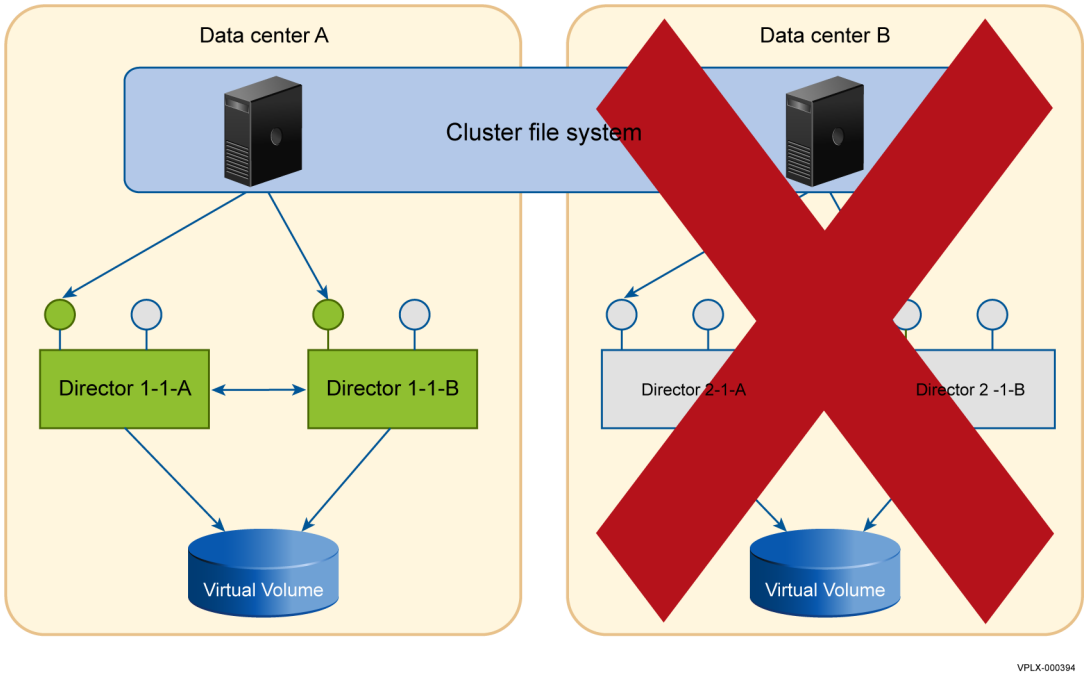

Figure 11. Path redundancy: different sites

Cluster

A metro node is a true cluster architecture. That is, all components are always available and I/O that enters the cluster from anywhere can be serviced by any node within the cluster, while coherency is maintained for all reads and writes.

A metro node cluster provides N–1 fault tolerance, which means that any component failure can be sustained, and the cluster will continue to operate as long as one director survives.

A metro node cluster consists of redundant hardware components.

All hardware resources (CPU cycles, and I/O ports) are pooled.

A two-cluster configuration (Metro) offers true high availability. Operations continue and data remains online even if an entire site fails. It also provides a high availability solution with zero recovery point objective (RPO).

Quorum

Quorum refers to the minimum number of directors required for the cluster to service and maintain operations.

There are different quorum rules for a cluster to become operational and start servicing I/Os when it is booting up, also called “gaining quorum.” Different rules for an operational cluster seeing director failures to either continue servicing operations and I/O after failure handling is called “maintaining quorum.” Stopping servicing operations and I/O is called “losing quorum.” These rules are described below:

- **Gaining quorum** - A non-operational metro node cluster gains quorum and becomes operational when more than half of the configured directors restart and come in contact with each other. In a single engine cluster, it refers to all the directors.
- **Maintaining quorum** - An operational metro node cluster seeing failures will continue operating in the following scenarios:
  - Director failures
    - If less than half of the operational directors with quorum fail.
    - If half of the operational directors with quorum fail, then the remaining directors will check the operational status of the failed directors over the management network and remain alive.

    After recovering from this failure, a cluster can tolerate further similar director failures until only one director is remaining. In a single engine cluster, a maximum of one director failure can be tolerated.
  - Intra-cluster communication failure
    - If there is a split in the middle, that is, half of the operational directors with quorum lose communication with the other half of the directors, and both halves are running, then the directors detect the operational status over the

management network and instruct half with the director with the lowest UUID to keep running and the directors without the lowest UUID to operationally stop.

- **Quorum loss** - An operational metro node cluster seeing failures stops operating in the following scenarios:
  - If more than half of the operational directors with quorum fail at the same time.
  - If half of the operational directors with quorum fail, and the directors are unable to determine the operation status of the other half of the directors (whose membership includes a low UUID).
  - In a dual or quad engine cluster, if all of the directors loose contact with each other.

Metadata volumes

Meta-volumes store metro node metadata, including virtual-to-physical mappings, data about devices, virtual volumes, and system configuration settings.

Metadata is stored in cache and backed up on specially designated external volumes called meta-volumes.

After the meta-volume is configured, updates to the metadata are written to both the cache and the meta-volume when the metro node configuration is modified.

Each metro node cluster maintains its own metadata, including:

- The local configuration for the cluster.
- Distributed configuration information shared between clusters.

At system startup, metro node reads the metadata and loads the configuration information onto each director.

When you make changes to the system configuration, metro node writes these changes to the metadata volume.

If metro node loses access to the metadata volume, the metro node directors continue uninterrupted, using the in-memory copy of the configuration. Metro node blocks changes to the system until access is restored or the automatic backup meta-volume is activated.

Meta-volumes experience high I/O only during system startup and upgrade.

I/O activity during normal operations is minimal.

Backup metadata volumes

Backup metadata volumes are point-in-time snapshots of the current metadata, and provide extra protection before major configuration changes, refreshes, or migrations.

Backup creates a point-in-time copy of the current in-memory metadata without activating it. You must create a backup metadata volume in any of these conditions:

- As part of an overall system health check before a major migration or update.
- If metro node permanently loses access to active meta-volumes.
- After any major migration or update.

Logging volumes

Logging volumes keep track of blocks written:

- During an inter-cluster link outage.
- When one leg of a DR1 becomes unreachable and then recovers.

After the inter-cluster link or leg is restored, the metro node system uses the information in logging volumes to synchronize the mirrors by sending only changed blocks across the link.

Logging volumes also track changes during loss of a volume when that volume is one mirror in a distributed device.

CAUTION: If no logging volume is accessible, then the entire leg is marked as out-of-date. A full re-synchronization is required once the leg is reattached.

The logging volumes on the continuing cluster experience high I/O during:

- Network outages or cluster failures
- Incremental synchronization

When the network or cluster is restored, metro node reads the logging volume to determine what writes to synchronize to the reattached volume.

There is no I/O activity during normal operations.

High availability and metro node hardware

The architectural design of the metro node hardware environment supports high availability.

The metro node hardware is largely designed to withstand technical failures and provide uninterrupted data availability. The critical components in the hardware are redundant to ensure that the failure of a component does not bring the system down.

Directors

A metro node director is the component that process the I/O requests from the hosts in a metro node environment. It interacts with the backend storage arrays for servicing the I/Os.

A director has two I/O modules for servicing I/Os from the arrays; one for the connectivity with the storage arrays on the back end, and another for connecting to the hosts on the front end. The management module in the director is used for management connectivity to the directors and for intra-cluster communication. The local communication module is completely dedicated to intra-cluster communication.

The front-end ports on all directors can provide access to any virtual volume in the cluster. Include multiple front-end ports in each storage view to protect against port failures. When a director port fails, the host multi-pathing software seamlessly fails over to another path through a different port, as shown in the following figure:

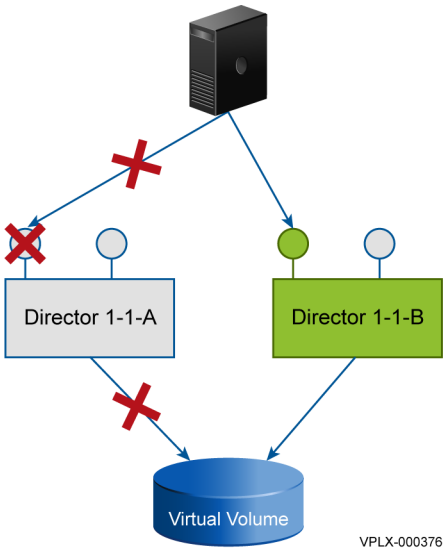

Figure 12. Path redundancy: different ports

Combine multi-pathing software plus redundant volume presentation for continuous data availability in the presence of port failures.

Back-end ports, local COM ports, and WAN COM ports provide similar redundancy for additional resilience.

Each director can service I/O for any other director in the cluster due to the redundant nature of the global directory and cache coherency.

If one director in the engine fails, another director continues to service I/O from the host.

In the following figure, Director 1-1-A has failed, but Director 1-1-B services the host I/O that was previously being serviced by Director 1-1-A.

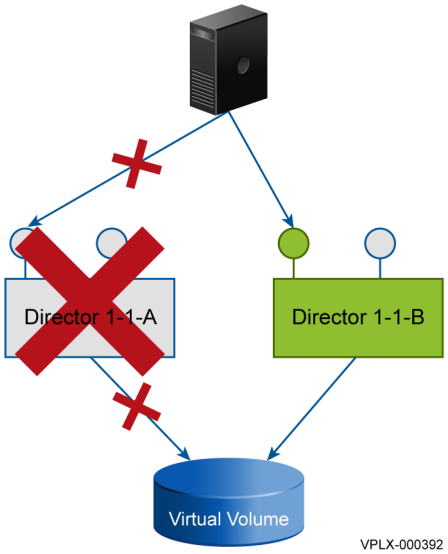

Figure 13. Path redundancy: different directors

Management server

Each metro node server has embedded management server. You can manage both clusters in a metro node Metro configuration from a single management server. The management server acts as a management interfaces to other metro node components in the cluster. Redundant internal network IP interfaces connect the management server to the public network. Internally, the management server is on a dedicated management IP network that provides accessibility to all major components in the cluster.

The larger role of the management server includes:

- Coordinating data collection, metro node software upgrades, configuration interfaces, diagnostics, event notifications, and some director-to-director communication.

High Availability with metro node Witness

Metro node CWS can be deployed on virtual machine. For more details, see the Configuration Guide for Cluster Witness available at https://solveonline.emc.com/solve/home/74

The following figure shows a high level architecture of metro node Witness. The metro node Witness server must reside in a failure domain separate from cluster-1 and cluster-2.

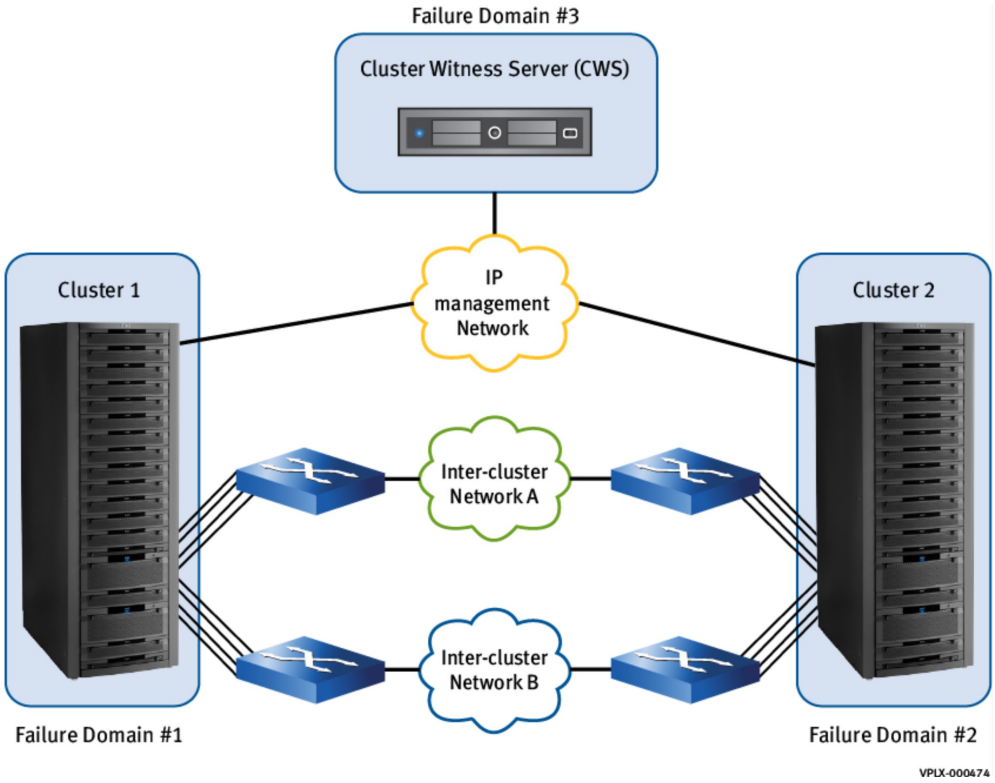

Figure 14. High level metro node Witness architecture

Metro node Metro HA

Metro node Metro High Availability (HA) configurations consist of a metro node Metro system deployed in conjunction with metro node Witness. There are two types of Metro HA configurations:

- Metro node Metro HA can be deployed in places where the clusters are separated by 10 ms latency RTT or less.
- Metro node Metro HA combined with Cross Connect between the metro node clusters and hosts can be deployed where the clusters are separated by 5 ms latency RTT or less (refer to Support Matrix for clusters' limitations).

Metro HA (without cross-connect)

Combine metro node Metro HA with host failover clustering technologies such as VMware HA to create fully automatic application restart for any site-level disaster.

Metro node Metro/VMware HA configurations:

- Significantly reduce the Recovery Time Objective (RTO). In some cases, RTO can be eliminated.
- Ride through any single component failure (including the failure of an entire storage array) without disruption.
- When VMware Distributed Resource Scheduler (DRS) is enabled, distribute workload spikes between data centers, alleviating the need to purchase more storage.
- Eliminate the requirement to stretch the Fiber Channel fabric between sites. You can maintain fabric isolation between the two sites.

In this deployment, virtual machines can write to the same distributed device from either cluster and move between two geographically disparate locations.

If you use VMware Distributed Resource Scheduler (DRS) to automate load distribution on virtual machines across multiple ESX servers, you can move a virtual machine from an ESX server attached to one metro node cluster to an ESX server attached to the second metro node cluster, without losing access to the underlying storage.

Metro HA with cross-connect

Metro node Metro HA with cross-connect (that is, hosts have connectivity to front-end ports of both metro node) can be deployed where the metro node clusters are separated by 5 ms latency RTT or less. Metro node Metro HA combined with cross-connect eliminates RTO for most of the failure scenarios.

Metro HA with cross-connect failure management

This section describes how metro node Metro HA with cross-connect rides through failures of hosts, storage arrays, clusters, metro node Witness, and the inter-cluster link.

Host failure

If hosts at one site fail, then VMware HA restarts the virtual machines on the surviving hosts. Since surviving hosts are connected to the same datastore, VMware can restart the virtual machines on any of the surviving hosts.

Cluster failure

If a metro node cluster fails:

- Metro node Witness guides the surviving cluster to continue.
- VMware re-routes I/O to the surviving cluster.
- No disruption to I/O.

Storage array failure

If one or more storage arrays at one site fail:

- All distributed volumes continue I/O to the surviving leg.
- No disruption to the metro node clusters or the virtual machines.
- I/O is disrupted only to local virtual volumes on the metro node cluster attached to the failed array.

Metro node Witness failure

If metro node Witness fails or becomes unreachable (link outage):

- Both metro node clusters call home to report that metro node Witness is not reachable.
- No disruption to I/O, metro node clusters, or the virtual machines.

Inter-cluster link failure

If the inter-cluster link fails:

- Metro node Witness guides the preferred cluster to continue.
- I/O suspends at the non-preferred cluster.
- VMware re-routes I/O to the continuing cluster.
- No disruption to I/O.

The following table summarizes how metro node HA with cross-connect manages failures.

Table 4. How metro node Metro HA recovers from failure

Failure description	Failure handling
Host failure (Site 1)	VMware HA software automatically restarts the affected applications at Site 2.
Metro node cluster failure (Site 1)	Metro node Witness detects the failure and enables all volumes on the surviving cluster.

Table 4. How metro node Metro HA recovers from failure (continued)

Failure description	Failure handling
Inter-cluster link failure	• If the cross-connects use different physical links from those used to connect the metro node clusters, applications are unaffected. Every volume continues to be available in one data center or the other. • If the cross-connect links use the same physical links as those used to connect the metro node clusters, an application restart is required.
Storage array failure	Applications are unaffected. Metro node dynamically redirects I/O to the mirrored copy on the surviving array. **NOTE:** This example assumes that all distributed volumes are also mirrored on the local cluster. If not, then the application remains available because the data can be fetched or sent from or to the remote cluster. However, each read/write operation now incurs a performance cost.
Failure of metro node Witness	Both clusters call home. As long as both clusters continue to operate and there is no inter-cluster link partition, applications are unaffected. **CAUTION: If either cluster fails or if there is an inter-cluster link partition, the system is at a risk of data unavailability. If the metro node Witness outage is expected to be long, disable the metro node Witness functionality to prevent the possible data unavailability.**

Metro HA without cross-connect failure management

This section describes the failure scenarios for metro node Metro HA without cross-connect.

Metro node cluster failure

In the event of a full metro node cluster outage at one site:

- Metro node Witness guides the surviving cluster to continue.
- VMware at the surviving cluster is unaffected.
- VMware restarts the virtual machines at the site where the outage occurred, redirecting I/O to the surviving cluster.

  VMware can restart because the second metro node cluster has continued I/O without interruption.

Inter-cluster link failure - non-preferred site

If an inter-cluster link outage occurs, the preferred cluster continues, while the non-preferred cluster suspends. If a virtual machine is located at the preferred cluster, there is no interruption of service. If a virtual machine is located at the non-preferred cluster, the storage associated with the virtual machine is suspended. In such a scenario, most guest operating systems will fail. The virtual machine will be restarted at the preferred cluster after a short disruption.

NOTE: The preferred cluster is determined by consistency group detach rules.

If an inter-cluster link outage occurs:

- Metro node Witness guides the preferred cluster to continue.
- VMware at the preferred cluster is unaffected.
- VMware restarts the virtual machines at the non-preferred (suspended) cluster, redirecting I/O to the preferred (uninterrupted) cluster.

VMware can restart because the second metro node cluster has continued I/O without interruption.

Higher availability

Deploy the metro node Witness in a third site failure domain to create even higher availability, as witness software can run on physical host or virtual machine (VMware, Hyper-V, and so on).

Metro node Metro Hardware

To ensure continuous availability across multiple data centers in a metro region, metro node Metro provides an ideal solution with the option of Metro over IP (MetroIP).

Metro node use a metro node Metro with a 10 Gb Ethernet.

6

Software and upgrade

This chapter describes the software that runs on the metro node hardware.

Topics:

Metro node OS

Metro node OS is the operating system that runs on the metro node hardware.

Metro node OS is:

- Designed for highly available, robust operation in geographically distributed environments
- Driven by real-time I/O operations
- Intelligent about locality of access
- Designed to provide the global directory that supports AccessAnywhere

The following table summarizes features provided by metro node OS and AccessAnywhere:

Table 5. Metro node OS AccessAnywhere features

Feature	Description and considerations
Storage volume encapsulation	LUNs on a back-end array can be imported into an instance of metro node and used while keeping their data intact.
	Considerations: The storage volume retains the existing data on the device and leverages the media protection and device characteristics of the back-end LUN.
RAID 1	metro node devices can be mirrored within a site.
	Considerations: Withstands a device failure within the mirrored pair. ● A device rebuild is a simple copy from the remaining device to the newly repaired device. Rebuilds are done in incremental fashion, whenever possible. ● The number of required devices is twice the amount required to store data (actual storage capacity of a mirrored array is 50%). ● The RAID 1 devices can come from different back-end array LUNs providing the ability to tolerate the failure of a back-end array.
Distributed RAID 1	Metro node devices can be mirrored between sites.
	Considerations: Provides protection from site disasters and supports the ability to move data between geographically separate locations.
Migration	Volumes can be migrated non-disruptively to other storage systems.
	Considerations: Use migration for changing the quality of service of a volume or for performing technology refresh operations.

Table 5. Metro node OS AccessAnywhere features (continued)

Feature	Description and considerations
Global Visibility	The presentation of a volume from one metro node cluster where the physical storage for the volume is provided by a remote metro node cluster.
	Considerations: Use Global Visibility for AccessAnywhere collaboration between locations. The cluster without local storage for the volume will use its local cache to service I/O but non-cached operations incur remote latencies to write or read the data.

Non-disruptive upgrade (NDU)

Metro node management server software and metro node OS can be upgraded without disruption.

Metro node hardware can be replaced, the engine count in a cluster increased, and a metro node Local can be expanded to metro node Metro without disruption.

Metro node never has to be completely shut down.

Storage, application, and host upgrades

Metro node enables the easy addition or removal of storage, applications, and hosts.

When metro node encapsulates back-end storage, the block-level nature of the coherent cache allows the upgrade of storage, applications, and hosts.

You can configure metro node so that all devices within metro node have uniform access to all storage blocks.

Software upgrades

Metro node is fully redundant for:

- Ports
- Paths
- Nodes

This redundancy allows GeoSynchrony on metro node Local and Metro to be upgraded without interrupting host access to storage, it does not require service window or application disruption.

NOTE: You must upgrade the metro node management server software before upgrading GeoSynchrony. Management server upgrades are non-disruptive.

Simple support matrix

Dell EMC publishes storage array interoperability information in a Simple Support Matrix available on Dell EMC Elab navigator. This information provides details on the tested, compatible combinations of storage hardware and applications that metro node supports. The Simple Support Matrix can be located at: https://elabnavigator.dell.com/eln/elnhome.