Dell EMC PowerVault ME4 Series and Linux

Abstract

This document provides best practices for Dell EMC™ PowerVault™ ME4 Series storage with Red Hat® Enterprise Linux® (RHEL).

November 2019

Revisions

Date	Description
September 2018	Initial release
November 2019	Updated reference to the support matrix

Acknowledgments

Author: Henry Wong

Table of contents

DELLEMC

Executive summary

Dell EMC™ PowerVault™ ME4 Series entry-level storage systems offer many robust hardware and software features that rival those found in more expensive storage systems. The Linux® operating system is an extremely robust, scalable, and popular enterprise-class operating system (OS) used by many organizations. PowerVault ME4 Series supports various Linux distributions. Refer to the ME4 Series *Storage System Support Matrix* on Dell.com/support for a complete list of supported operating systems.

When configured using the best practices presented in this paper, the Linux operating systems provide an optimized experience with ME4 Series storage. The recommendations in this document include guidelines for volume discovery, multipath configuration, logical volume manager, and file-system and snapshot data protection.

This paper uses Red Hat® Enterprise Linux® (RHEL) to provide administration and configuration examples. These examples are intended to provide a starting point for storage and system administrators.

Audience

This document is intended for ME4 Series administrators, system administrators, and anyone responsible for configuring the storage systems. It is assumed that readers have prior experience with or training in the following:

- SAN storage systems
- Linux operating environment
- Multipath software

We welcome your feedback along with any recommendations for improving this document. Send comments to StorageSolutionsFeedback@dell.com.

DELLEMC

1 ME4 Series overview

The ME4 Series is a next-generation, entry-level storage system that is available in three models: PowerVault ME4012, ME4024, and ME4084. Three disk expansion enclosure models are also available to provide expanded capacity: PowerVault ME412, ME424, and ME484.

1.1 Hardware features

ME4 Series arrays are available with single or dual controllers. A dual-controller configuration offers a higher level of availability, protection, and double the I/O modules for connectivity and performance. In this configuration, if a partner controller fails, the storage system will fail over and run on a single controller until the redundancy is restored. With proper cabling is in place and MPIO configured, the failover should not cause any service interruption to the hosts and applications.

ME4 Series arrays can be configured in all-flash, all-HDD, or hybrid configurations, and are designed to be installed by customers. To find more information about ME4 Series features and hardware specifications, go to Dell.com.

1.2 Software features

ME4 Series storage includes a comprehensive set of software features in an all-inclusive license. The following sections provide a summary of the ME4 Series software features. For more information, refer to ME4 Series *Administrator's Guide* on Dell.com/support.

1.2.1 Virtual and linear storage

ME4 Series arrays support two storage technologies: virtual and linear storage.

Virtual storage is the foundation for data-management features such as thin provisioning, automated tiered storage (ATS), and other virtualization features. Virtual storage virtualizes the physical devices and creates virtual volumes that reside in virtual disk groups. A virtual volume is a set of pages that are mapped to the physical devices. A virtual volume can span across more than 16 physical devices which is highly flexible and efficient. It also offers excellent performance because it scales with the number of physical devices in the pool.

Linear storage does not virtualize the logical data on the physical devices. It stores data in sequential, fully allocated physical blocks and the mapping is static. While it offers better performance for certain sequential workloads, it is inflexible, and many virtualization features such as tiering, snapshots, and replications are not available on linear pools. The most common use case for linear storage is for video archiving.

In most cases, Dell EMC recommends using virtual storage with ME4 Series systems.

1.2.2 RAID data protection levels

ME4 Series arrays support RAID data protection levels NRAID, 0, 1, 10, 3, 5, 50, 6 and ADAPT. ADAPT is a special RAID implementation that offers some unique benefits. It can withstand two drive failures with very fast rebuilds. Spare capacity is distributed across all drives instead of dedicated spare drives. ADAPT disk groups can have up to 128 drives and allow mixing different drive sizes. Data is stored across all disks evenly. The storage system automatically rebalances the data when new drives are added or when the distribution of data has become imbalanced.

It is recommended to choose the right RAID level that best suits the type of workloads in the environment. Review the information in ME4 Series *Administrator's Guide* which details the benefits of each RAID level, the minimum and maximum disks requirements, and the recommendation of RAID levels for popular workloads.

1.2.3 Thin provisioning

Virtual volumes are thin provisioned in the storage system where pages are allocated only when the data is written to the volume. This allows the storage system to appear to have more storage than what is actually allocated. It also allows more efficient usage because the consumption is based on the actual amount of data written and not on the defined volume size.

1.2.4 Automated tiered storage

When the storage system is configured with different types of disks, the automated tiered storage (ATS) feature automatically moves data between different disk classes based on data-access patterns to optimize performance and space consumption. There are three storage tiers: Performance, Standard and Archive. The performance tier uses SSDs, the standard tier uses lower-cost HDD SAS disks, and the archive tier uses lowest-cost, large-capacity NL-SAS disks.

1.2.5 Snapshots

The storage system can create point-in-time copies of virtual volumes. Snapshot creation is fast and efficient because only changed data in the snapshot consume space. See section 3.6 about data protection with snapshots.

1.2.6 Replication

Virtual volumes and snapshots can be replicated asynchronously to a remote ME4 Series system. This feature can be used for disaster recovery, to distribute data, or to archive data to an offsite location. Replication requires dedicated I/O ports and therefore reduces the ports that can be used for host connectivity.

1.2.7 Full-disk encryption

ME4 Storage System supports full-disk encryption (FDE) using self-encrypting drives (SEDs). The data and system are secured on the disks and are accessed using passphrases and lock keys that are only available on the storage system. FDE operates on a per-system basis. To use FDE, all disks in the system must be FDE-capable.

1.2.8 ME Storage Manager

ME Storage Manager is a web-based interface for configuring, monitoring, and managing the storage system. It can be accessed through each controller's embedded web server.

D∈LLEMC

1.2.9   Command-line interface

The storage system provides a full-feature command-line interface (CLI) that enables administrators to manage and configure the storage system. The CLI software is embedded in the controller modules and can be accessed through HTTPS, HTTP, SSH, Telnet over the network, or by directly connecting to the controller's serial CLI port. In additional to the interactive CLI console format, the CLI API format allows external programs to interact with the system through XML and JSON formats. To script CLI commands, the CLI API format is recommended because it is expected to remain consistent from release to release. Refer to the ME4 Series *CLI Guide* on Dell.com/support for a complete description of the CLI commands.

DELLEMC

2 Connectivity considerations

In a dual-controller configuration, each ME4 Series system provides up to eight host ports, four per controller, that support the following host connectivity options:

- All Fibre Channel
- All iSCSI
- All SAS
- FC/iSCSI combination

    - FC: ports A0, A1, B0, and B1
    - iSCSI: ports A2, A3, B2, and B3

Note: To make sure the host HBAs, and FC or network switches meet the support requirements, refer to the ME4 Series *Support Matrix* at Dell.com/support for supported hardware, firmware, and driver versions.

2.1 Direct-attached storage

ME4 Series storage supports connecting hosts directly to one or more IOM ports. The number of hosts supported depends on the number of IOM ports available (eight ports in a dual-controller configuration), and whether the hosts have multiple redundant connections. This configuration is simple and does not require additional SAN switches, but limits the number of attached servers. It is highly recommended to use a dual-controller, redundant-path configuration for any critical or production environments.

ME4 Series storage supports FC arbitrated loop (public or private) or point-to-point topologies. When directly connecting the host and ME4 Series system, use either loop protocol or point-to-point protocol. Point-to-point direct connection is only supported at 16 Gb/s.

2.2 SAN-attached storage

Given the limited host ports available on a ME4 Storage System, attaching the storage system and hosts to a SAN allows more hosts to connect to a single ME4 Series system concurrently. While this can significantly improve the usage efficiency of the storage system, administrators should carefully balance the total workload driven by all connected hosts and the resource availability on the storage system.

2.2.1 Fibre Channel

Fibre Channel zones are used to segment the fabric to restrict access. A zone contains paths between initiators (server HBAs) and targets (storage array front-end ports). Either physical ports (port zoning) on the Fibre Channel switches or the WWNs (name zoning) of the end devices can be used in zoning. It is recommended to use name zoning because it offers better flexibility. With name zoning, server HBAs and storage array ports are not tied to specific physical ports on the switch.

Zoning Fibre Channel switches for Linux hosts is essentially no different than zoning any other hosts to the ME4 Series array.

D&LLEMC

The following list includes key zoning rules and recommendations.
- Use the point-to-point protocol to connect to a fabric switch.
- Both the ME4 Series system and the Linux hosts should be connected to two different Fibre Channel switches (fabrics) for high availability and redundancy.
- Name zoning using WWNs is recommended. See 2.2.1.1 and 2.2.1.2 for instruction on identifying the WWNs on hosts and the ME4 Series array.
- When defining the zones, it is a best practice to use single-initiator (host port), multiple-target (ME4 ports) zones. For example, for each Fibre Channel HBA port on the server, create a server zone that includes the HBA port WWN and all the physical WWNs on the ME4 Series system controllers on the same fabric. See Table 1 for an example.

Table 1   Fibre Channel zoning examples

Fabrics (dual-switch configuration)	FC HBA port (dual-port HBA configuration)	ME4 Series FC ports	
		All-FC port configuration	FC/iSCSI port configuration
Fabric one zone	Port 0	A0, B0, A2, B2	A0, B0*
Fabric two zone	Port 1	A1, B1, A3, B3	A1, B1*

*In an FC/iSCSI configuration, only ports A0, A1, B0, and B1 are available. Ports A2, A3, B2, and B3 are reserved for iSCSI.

Note: Dell EMC recommends using name zoning and creating single-initiator, multiple-target zones.

2.2.1.1   Identify host FC WWNs

When creating a host object on the ME4 Series array, the host FC WWNs are required. Use the following procedure to identify the HBA WWNs on the Linux host. The **fcshow.sh** bash shell script is provided as an example to extract useful information about the host HBA information.

```
# cat fcshow.sh

printf "%-10s %-20s %-10s %-s\n" "Host-Port" "WWN" "State" "Speed"
printf "%50s\n" |tr ' ' -

ls -1d /sys/class/fc_host/host* | while read host
do
   (echo ${host##*/}: ; cat $host/port_name $host/port_state $host/speed)| xargs
-n 5 printf "%-10s %-20s %-10s %s %s\n"
done

# bash fcshow.sh

Host-Port  WWN                 State      Speed
-------------------------------------------------
host14:    0x2001000e1ec2ada0  Online     16 Gbit
host15:    0x2001000e1ec2ada1  Online     16 Gbit
host16:    0x2001000e1ec2d342  Online     16 Gbit
host17:    0x2001000e1ec2d343  Online     16 Gbit
host18:    0x2001000e1e09b7b8  Online     16 Gbit
host19:    0x2001000e1e09b7b9  Online     16 Gbit
```

2.2.1.2 Identify ME4 Series FC WWNs

Locate the WWN information in the ME Storage Manager or by running the interactive CLI command. Refer to the ME4 Series *CLI Guide* for a detailed explanation of all available CLI commands.

Identify WWNs using ME Storage Manager

1. Log in to ME Storage Manager.
2. Navigate to the **Home** topic.
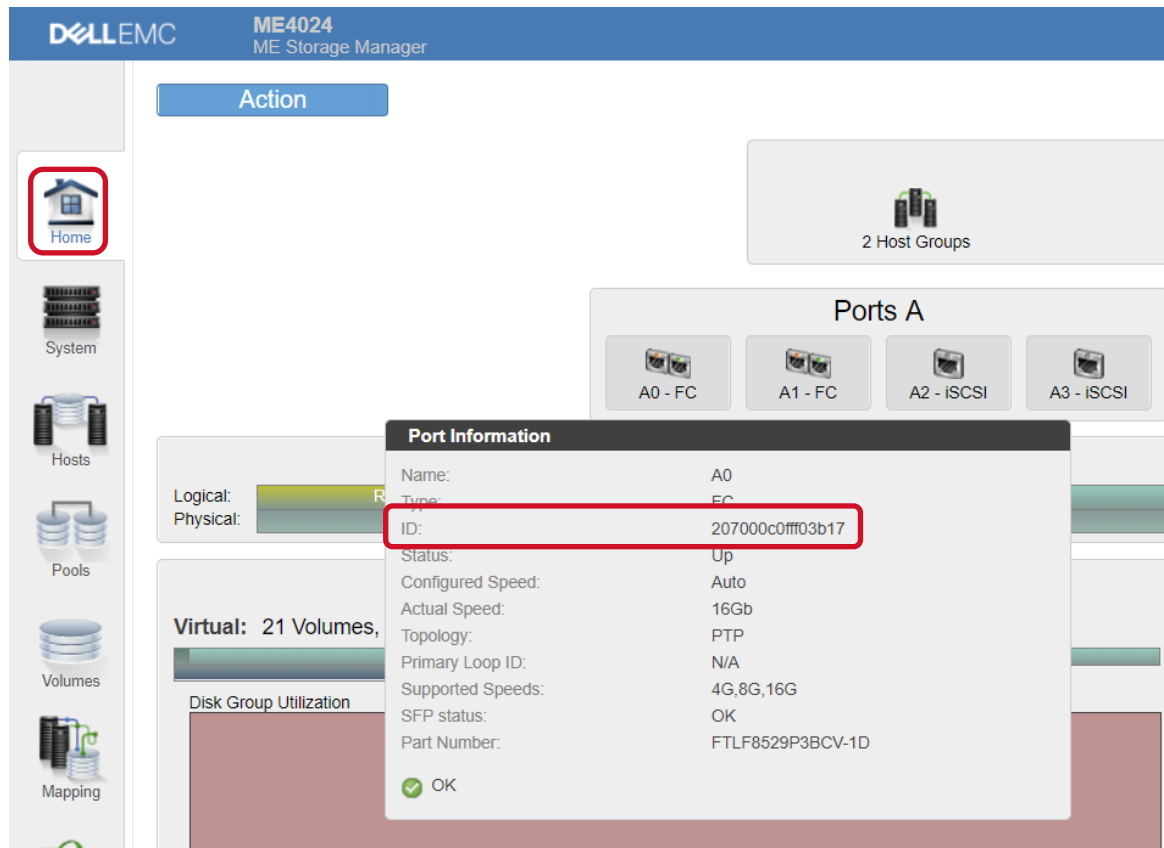3. Hover over the host ports to display the configuration information. See Figure 1.

Figure 1   Display FC port WWN in ME Storage Manager

Identify WWNs using interactive CLI commands

1. Log in to the ME4 Series controller using **ssh**.

```
ssh manage@IP-address
```

2. Execute the following CLI command to show the host-port information. The following shows information for one FC port.

```
# show ports detail
Ports Media    Target ID                                Status
Speed(A) Health    Reason Action
----------------------------------------------------------------------------
---------
A0    FC(P)    207000c0fff03b17                        Up              16Gb
OK

    Topo(C) PID SFP Status          Part Number     Supported Speeds
    ----------------------------------------------------------------
    PTP     N/A OK                  FTLF8529P3BCV-1D 4G,8G,16G
```

2.2.2 iSCSI

The following iSCSI network best practices are recommended:

- Use dual network switches to provide network-level redundancy.
- Use two dedicated iSCSI networks to isolate iSCSI traffic between the hosts and the ME4 Series array from other public or application traffic. See Table 2 for an example.
- Use two dedicated network ports on the host for path redundancy.
- Use 10 Gb network HBAs and switches for better performance.
- Use auto-negotiate for all interfaces that negotiate at full-duplex and at the maximum speed of the connected port.
- Enable flow control on all servers and switch ports that handle iSCSI traffic. A minimum of receive (RX) flow control should be enabled for all switch interfaces used by the servers and storage systems for iSCSI traffic.
- Enable symmetric flow control for all server interfaces used for iSCSI traffic.
- For multipathed iSCSI, use two separate VLANs and separate IP subnets.
- Disable unicast storm control on iSCSI switches.
- Disable multicast or set multicast storm control to enabled at the switch level for all iSCSI VLANs.
- If using Jumbo frames for improved performance, make sure all devices on the transmission path have Jumbo frames enabled and support 9,000 as the maximum transmission unit (MTU).
- If the switches do not support both flow control and Jumbo frames at the same time, Dell EMC recommends choosing flow control over Jumbo frames.
- Refer to the ME4 Series *Deployment Guide* on Dell.com/support for additional iSCSI best-practice information.

Table 2   iSCSI port assignment examples

Network/VLAN	iSCSI HBA port	ME4 Series iSCSI ports	
		All-iSCSI port configuration	FC/iSCSI port configuration
Network 1/VLAN 1	eth0	A0, B0, A2, B2	A2, B2*
Network 2/VLAN2	eth1	A1, B1, A3, B3	A3, B3*

*In an FC/iSCSI configuration, only ports A2, A3, B2, and B3 are available. Ports A0, A1, B0, and B1 are reserved for FC.

DELLEMC

2.2.2.1 Identify ME4 Series iSCSI port network IP addresses

If the iSCSI ports are set up during the initial Easy Start wizard, the iSCSI port network information can be found under the **Home** topic in ME Storage Manager. Hover over the iSCSI port to see the configuration information. See Figure 2.
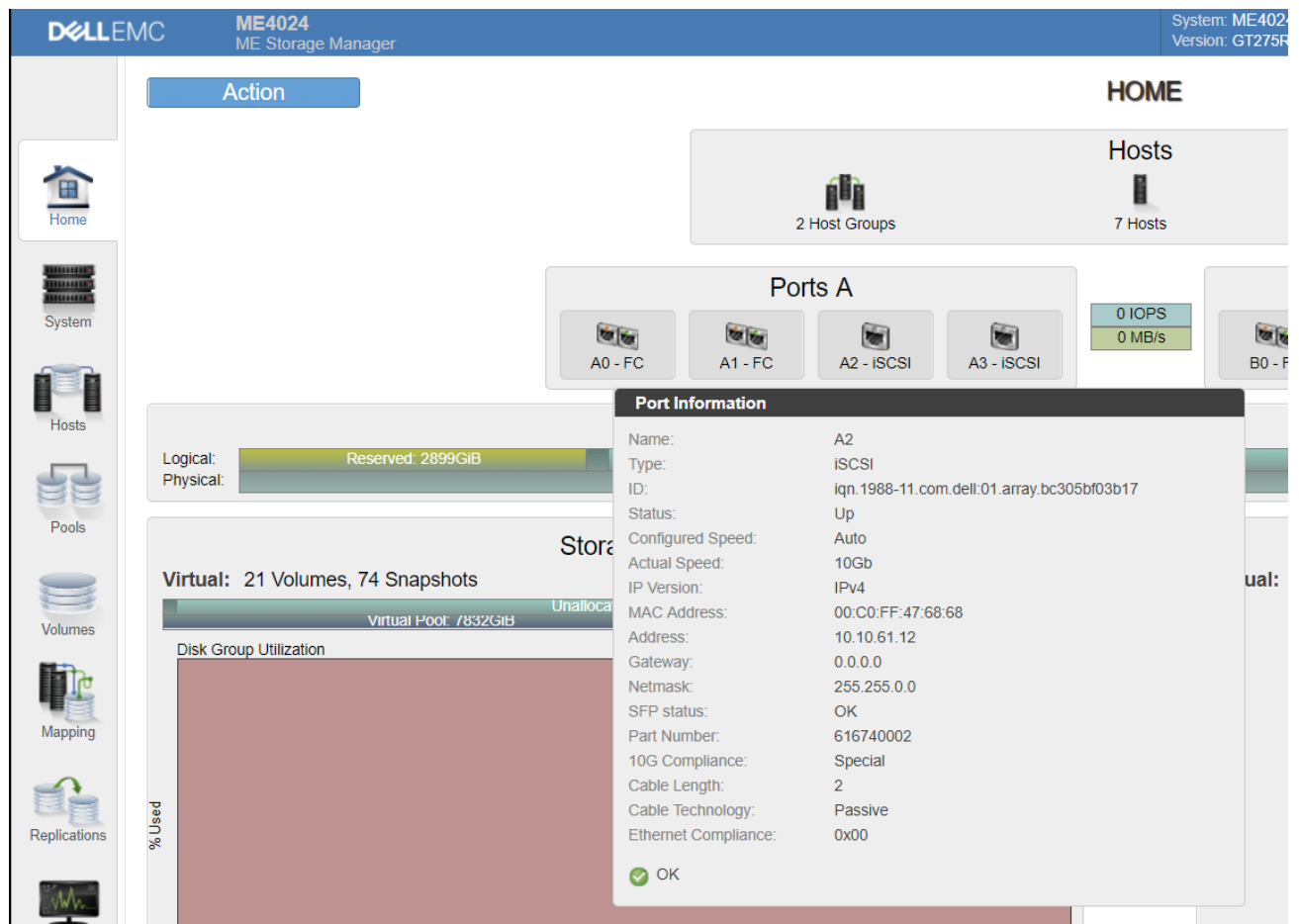

Figure 2   ME Storage Manager iSCSI port information

Alternatively, use the ME4 Series CLI command to display the port information.

1. Log in to the ME4 Series controller using **ssh**.

    ```
    # ssh manage@IP-address
    ```

2. Execute the following CLI command to show all host-port information. The following shows the output for one iSCSI port.

```
# show ports detail
Ports Media    Target ID                                    Status     Speed(A)
Health   Reason Action
----------------------------------------------------------------------------
---------------------
A2    iSCSI    iqn.1988-11.com.dell:01.array.bc305bf03b17  Up         10Gb
OK

    Port Details
    ------------
    IP Version: IPv4
    IP Address: 10.10.61.12
    Gateway: 0.0.0.0
    Netmask: 255.255.0.0
    MAC: 00:C0:FF:47:68:68
    SFP Status: OK
    Part Number: 616740002
    10G Compliance: Special
    Ethernet Compliance: 0x00
    Cable Technology: Passive
    Cable Length: 2
```

2.2.2.2 Configure host network adapter and iSCSI

To configure the host iSCSI initiator, use the following steps to make the recommended settings and activate the iSCSI transport.

1. Edit the **/etc/iscsi/iscsid.conf** file and set the following values. These values dictate the failover timeout and queue depth. The values shown serve as a starting point and might require adjustment depending on the environment.

   ```
   node.session.timeo.replacement_timeout = 5
   node.session.cmds_max = 1024
   node.session.queue_depth = 128
   ```

2. Perform iSCSI discovery against one of the ME4 Series iSCSI port IPs. This sets up the nodedb in the **/var/lib/iscsi** directory. See section 2.2.2.1 to identify the ME4 Series iSCSI port IP addresses.

   ```
   # iscsiadm -m discovery -t st -p $ME4_A2_IP
   ```

3. Log in to the discovered iSCSI qualified name (IQN).

   ```
   # iscsiadm -m node -L all
   ```

4. Verify the host has established the iSCSI sessions to all the ME4 Series iSCSI ports. The following command should display a session for each ME4 Series iSCSI port. In this example, 10.10.61.12 corresponds to A2, 10.20.61.12 corresponds to B2, 10.10.61.13 corresponds to A3, and 10.20.61.13 corresponds to B3.

```
# iscsiadm -m node
10.10.61.12:3260,5 iqn.1988-11.com.dell:01.array.bc305bf03b17
10.20.61.12:3260,6 iqn.1988-11.com.dell:01.array.bc305bf03b17
10.20.61.13:3260,7 iqn.1988-11.com.dell:01.array.bc305bf03b17
10.10.61.13:3260,8 iqn.1988-11.com.dell:01.array.bc305bf03b17
```

5. Identify the host iSCSI initiator IQN. The IQN information is required to create the host object in ME Storage Manager.

```
# cat /etc/iscsi/initiatorname.iscsi
```

2.2.2.3 Host TCP settings and other tuning considerations

Kernel parameters that can be tuned for performance are found in the **/proc/sys/net/core** and **/proc/sys/net/ipv4** kernel parameters. Once optimal values are determined, permanently set these in the **/etc/sysctl.conf** file. Like most other modern operating system platforms, Linux can efficiently auto-tune TCP buffers. However, by default, some settings such as buffer size are conservatively low. Experimenting with the following kernel parameters can lead to improved network performance, and subsequently improve iSCSI performance.

To set these parameters permanently, enter them in the **/etc/sysctl.conf** file and reboot the servers.

Table 3   An example of TCP settings for 10 GB network

Parameter	Value	Description
net.core.rmem_max	134217728	Maximum receive buffer size used by each TCP socket
net.core.wmem_max	134217728	Maximum send buffer size used by each TCP socket
net.ipv4.tcp_rmem	4096 87380 134217728	Auto-tune TCP buffer limits: min, default, and max size of the receive buffer used by each TCP socket
net.ipv4.tcp_wmem	4096 65536 134217728	Auto-tune TCP buffer limits: min, default, and max size of the send buffer used by each TCP socket
net.core.netdev_max_backlog	300000	Maximum number of incoming connections backlog queue
net.ipv4.tcp_moderate_rcvbuf	1	Auto-tune the receiver buffer size
net.bridge.bridge-nf-call-iptables	0	netfilter
net.bridge.bridge-nf-call-arptables	0	netfilter
net.bridge.bridge-nf-call-ip6tables	0	netfilter

DELLEMC

3 Linux setup and configuration

The following subsections describe best practices when working with ME4 Series arrays on Linux operating systems. This paper uses Red Hat Enterprise Linux to illustrate various Linux setup and configuration tasks.

3.1 Discover and identify ME4 Series volumes

After creating and enabling host access of the LUNs in the ME4 Series system, the host operating system needs to scan for these new LUNs before they can be used. On Linux, install the following rpm packages which contain useful utilities to discover and identify LUNs: **sg3_utils**, **lsscsi**, and **blkid**.

3.1.1 Scan and identify new volumes on Linux

New volumes are added to the Linux host by scanning the SCSI buses that have access to the ME4 Series system. The **sg3_utils** package provides the **rescan-scsi-bus.sh** script that helps simplify the process. It is recommended to scan for all targets. The following command works for all LUNs except for LUN 0. See section 3.1.2 for information on scanning LUN 0.

```
# /usr/bin/rescan-scsi-bus.sh -a
```

3.1.2 Scan LUN 0

By default, LUN 0 is assigned to the ME4 Series system. This assignment is completed as soon as the host has visibility to the ME4 Series system (for example, when FC zones are created for the host and ME4 Series system, or when the host iSCSI initiator logs in to ME4 Series iSCSI ports). Therefore, the very first volume mapped to the Linux host starts with LUN ID 1.

```
# lsscsi -is
[0:0:0:0]    cd/dvd QEMU     QEMU DVD-ROM       0.10  /dev/sr0   -       -
[6:0:0:0]    enclosu DellEMC  ME4               G275  -          -       -
[6:0:0:1]    disk    DellEMC  ME4               G275  /dev/sdl
3600c0ff0003c589fe452605b01000000  99.9GB
[7:0:0:0]    enclosu DellEMC  ME4               G275  -          -       -
[7:0:0:1]    disk    DellEMC  ME4               G275  /dev/sdn
3600c0ff0003c589fe452605b01000000  99.9GB
[8:0:0:0]    enclosu DellEMC  ME4               G275  -          -       -
[8:0:0:1]    disk    DellEMC  ME4               G275  /dev/sdj
3600c0ff0003c589fe452605b01000000  99.9GB
[9:0:0:0]    enclosu DellEMC  ME4               G275  -          -       -
[9:0:0:1]    disk    DellEMC  ME4               G275  /dev/sdk
3600c0ff0003c589fe452605b01000000  99.9GB
[10:0:0:0]   enclosu DellEMC  ME4               G275  -          -       -
[10:0:0:1]   disk    DellEMC  ME4               G275  /dev/sdm
3600c0ff0003c589fe452605b01000000  99.9GB
[11:0:0:0]   enclosu DellEMC  ME4               G275  -          -       -
[11:0:0:1]   disk    DellEMC  ME4               G275  /dev/sdi
3600c0ff0003c589fe452605b01000000  99.9GB
[12:0:0:0]   enclosu DellEMC  ME4               G275  -          -       -
[12:0:0:1]   disk    DellEMC  ME4               G275  /dev/sdh
3600c0ff0003c589fe452605b01000000  99.9GB
[13:0:0:0]   enclosu DellEMC  ME4               G275  -          -       -
[13:0:0:1]   disk    DellEMC  ME4               G275  /dev/sdg
3600c0ff0003c589fe452605b01000000  99.9GB
```

To reclaim LUN 0 for a regular volume, manually unmap and remap the volume in ME Storage Manager. In **ME Storage Manager** under the **Volumes topic**, right-click the desired volume and click **Map Volumes**. On the **Map** volume interface, override the LUN ID.
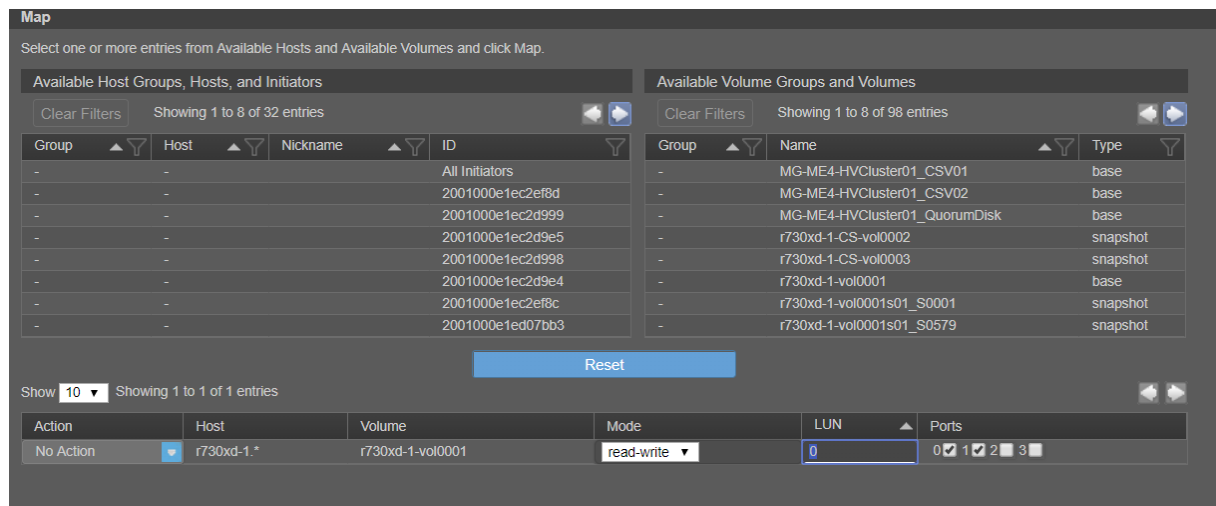

Figure 3 Map wizard

On the Linux host, perform a force rescan on all SCSI buses. This removes the enclosure entries and the volume is assigned to LUN 0.

```
# rescan-scsi-bus.sh --forcerescan

# lsscsi -is
[0:0:0:0]    cd/dvd  QEMU      QEMU DVD-ROM      0.10  /dev/sr0    -        -
[6:0:0:0]    disk    DellEMC   ME4               G275  /dev/sdg
3600c0ff0003c589fe452605b01000000   99.9GB
[7:0:0:0]    disk    DellEMC   ME4               G275  /dev/sdh
3600c0ff0003c589fe452605b01000000   99.9GB
[8:0:0:0]    disk    DellEMC   ME4               G275  /dev/sdi
3600c0ff0003c589fe452605b01000000   99.9GB
[9:0:0:0]    disk    DellEMC   ME4               G275  /dev/sdj
3600c0ff0003c589fe452605b01000000   99.9GB
[10:0:0:0]   disk    DellEMC   ME4               G275  /dev/sdk
3600c0ff0003c589fe452605b01000000   99.9GB
[11:0:0:0]   disk    DellEMC   ME4               G275  /dev/sdl
3600c0ff0003c589fe452605b01000000   99.9GB
[12:0:0:0]   disk    DellEMC   ME4               G275  /dev/sdm
3600c0ff0003c589fe452605b01000000   99.9GB
[13:0:0:0]   disk    DellEMC   ME4               G275  /dev/sdn
3600c0ff0003c589fe452605b01000000   99.9GB
```

Note: Only use the **--forcerescan** option for removing the enclosure entry and reclaiming LUN 0.

3.1.3 Identify ME4 Series volumes on Linux

ME4 Series storage automatically assigns LUN IDs when the LUN is mapped to the Linux host. Assigning specific LUN IDs to volumes is possible but requires un-mapping and re-mapping the volumes manually. In the Map Volume wizard, administrators can explicitly override the LUN IDs.

After scanning and adding the volumes on a Linux host, the volumes can be identified by either their LUN IDs or their SCSI IDs.

3.1.3.1 Identify volumes by LUN ids

The LUN ID information can be found in ME Storage Manager in the **Volumes** topic. Select the desired volume and click the **Maps** tab. The **LUN** column reflects the assigned ID. If the volume belongs to a **Volume Group**, right-click the group and select **View Map Details**. See Figure 4.
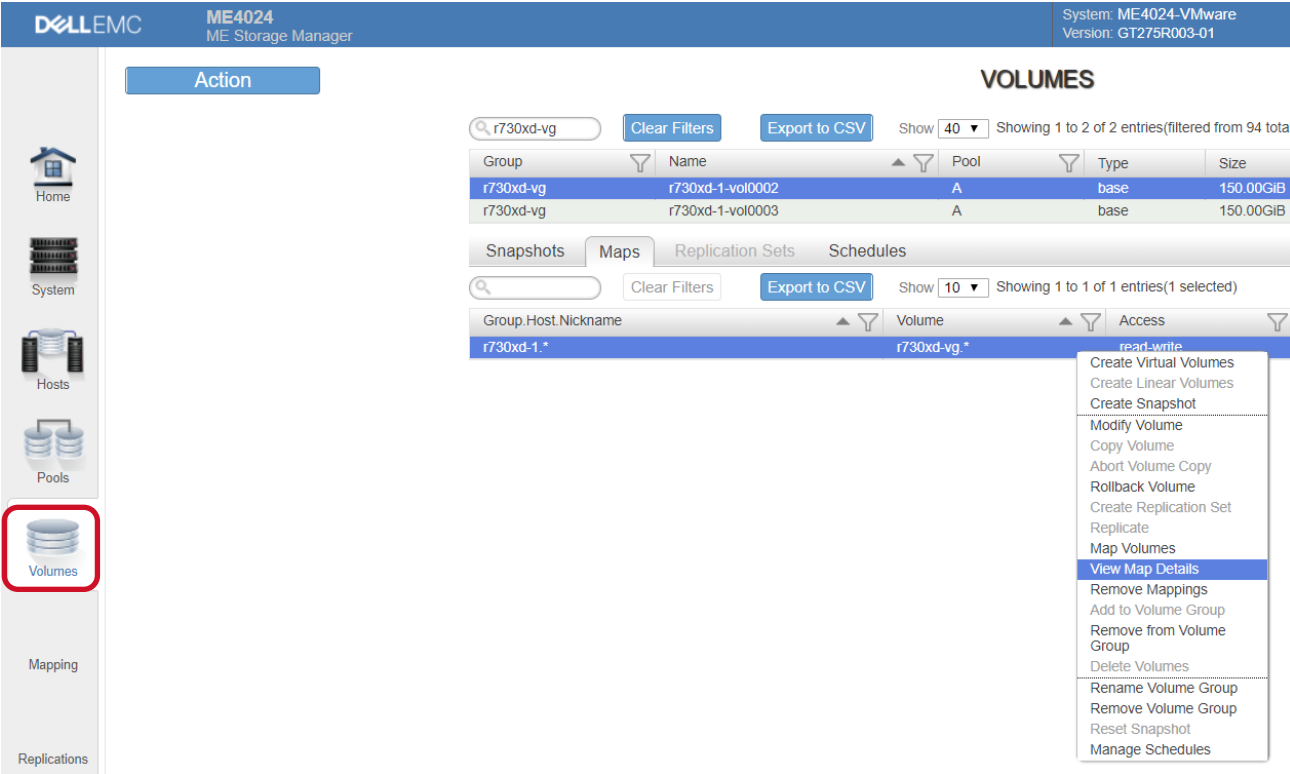

Figure 4   Accessing the LUN ID information in ME Storage Manager

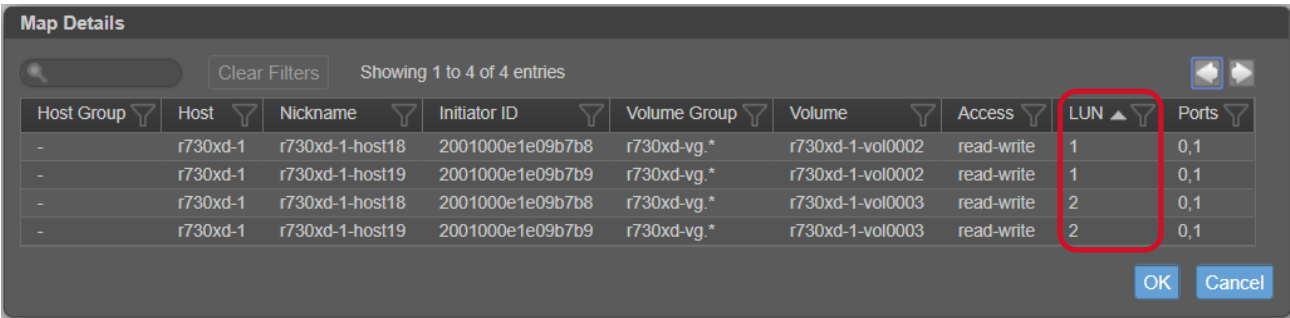The LUN ID information in the Map Details window is shown in Figure 5.

Figure 5   Volume map details

On the Linux host, run the **lsscsi** command to show the volume LUN IDs. The highlighted digits in red correspond to the LUN IDs of the volumes.

```
[root@r730xd-1 me4_fs]# lsscsi -is|grep ME4
[18:0:4:0]   disk    DellEMC   ME4              G275  /dev/sdl
3600c0ff0003c589f1e4b6b5b01000000   214GB
[18:0:4:1]   disk    DellEMC   ME4              G275  /dev/sdk
3600c0ff0003c589f97596b5b01000000   161GB
[18:0:4:2]   disk    DellEMC   ME4              G275  /dev/sdm
3600c0ff0003c589f98596b5b01000000   161GB
[18:0:5:0]   disk    DellEMC   ME4              G275  /dev/sdd
3600c0ff0003c589f1e4b6b5b01000000   214GB
[18:0:5:1]   disk    DellEMC   ME4              G275  /dev/sdn
3600c0ff0003c589f97596b5b01000000   161GB
[18:0:5:2]   disk    DellEMC   ME4              G275  /dev/sdo
3600c0ff0003c589f98596b5b01000000   161GB
[19:0:4:0]   disk    DellEMC   ME4              G275  /dev/sdi
3600c0ff0003c589f1e4b6b5b01000000   214GB
[19:0:4:1]   disk    DellEMC   ME4              G275  /dev/sdp
3600c0ff0003c589f97596b5b01000000   161GB
[19:0:4:2]   disk    DellEMC   ME4              G275  /dev/sdq
3600c0ff0003c589f98596b5b01000000   161GB
[19:0:5:0]   disk    DellEMC   ME4              G275  /dev/sdj
3600c0ff0003c589f1e4b6b5b01000000   214GB
[19:0:5:1]   disk    DellEMC   ME4              G275  /dev/sdr
3600c0ff0003c589f97596b5b01000000   161GB
[19:0:5:2]   disk    DellEMC   ME4              G275  /dev/sds
3600c0ff0003c589f98596b5b01000000   161GB
```

3.1.3.2   Identify volumes by SCSI IDs

Alternatively, using SCSI IDs to identify the volumes is more precise because SCSI IDs are unique on the storage system and on the host, and they will not change. This method reduces the chance of confusion if the LUN IDs are changed unintentionally. It is common to set up persistent device mapping to fixate the device file names, file ownership, and permission. This can be achieved by creating **udev** rules that search for specific SCSI IDs, rename device files, and make ownership and permission changes.

To view SCSI IDs, one option is to run the **lsscsi** command. See the prior example in section 3.1.3.1 which shows the SCSI IDs highlighted in blue.

Additional options to show SCSI IDs include using the **multipath** command (if device multipath is enabled), or **scsi_id** command. These options are detailed as follows:

Option 1: Show the SCSI ID with the **multipath** command.

```
# multipath -l mpathb
mpathb (3600c0ff0003c589f1e4b6b5b01000000) dm-6 DellEMC ,ME4
size=200G features='0' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=0 status=active
| |- 18:0:5:0 sdd 8:48  active undef unknown
| `- 19:0:5:0 sdj 8:144 active undef unknown
`-+- policy='service-time 0' prio=0 status=enabled
  |- 18:0:4:0 sdl 8:176 active undef unknown
  `- 19:0:4:0 sdi 8:128 active undef unknown
```

Option 2: Show the SCSI ID with the **scsi_id** command.

```
# ls -1 /dev/sd[a-z] /dev/dm*|while read path
do
    (echo $path:; /usr/lib/udev/scsi_id --page=0x83 --whitelisted --
device=$path)|xargs -n 2
done

/dev/dm-6: 3600c0ff0003c589f1e4b6b5b01000000
/dev/dm-7: 3600c0ff0003c589f97596b5b01000000
/dev/dm-8: 3600c0ff0003c589f98596b5b01000000
/dev/sdi: 3600c0ff0003c589f1e4b6b5b01000000
/dev/sdj: 3600c0ff0003c589f1e4b6b5b01000000
/dev/sdk: 3600c0ff0003c589f97596b5b01000000
/dev/sdl: 3600c0ff0003c589f1e4b6b5b01000000
/dev/sdm: 3600c0ff0003c589f98596b5b01000000
/dev/sdn: 3600c0ff0003c589f97596b5b01000000
/dev/sdo: 3600c0ff0003c589f98596b5b01000000
/dev/sdp: 3600c0ff0003c589f97596b5b01000000
/dev/sdq: 3600c0ff0003c589f98596b5b01000000
/dev/sdr: 3600c0ff0003c589f97596b5b01000000
/dev/sds: 3600c0ff0003c589f98596b5b01000000
```

3.1.3.3   Identify ME4 Series volume SCSI IDs

The SCSI IDs of ME4 Series volumes can be retrieved using one of the two following ways:

Option 1: Export the volume information in ME Storage Manager.

In ME Storage Manager in the **Volumes** topic, select the desired volume or volumes and click the **Export to CSV** button. Select the data and format in the **Export Data to CSV** dialog. Once the CSV file is created and downloaded to your computer, open the file with a word processor or spreadsheet application. The volume SCSI IDs are listed in the CSV file in the **WWN** column. See the following figures.

DELLEMC

Figure 6   Filter and select volumes in ME Storage Manager
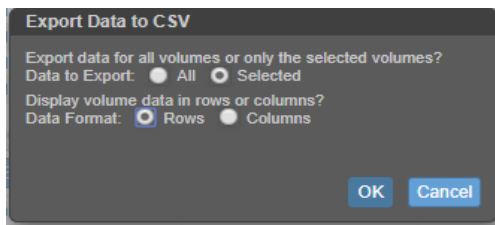
Figure 7   Export Data to CSV

Figure 8   Review the SCSI IDs (WWNs) in the CSV file

Option 2: Use the following interactive CLI command to identify ME4 Series volume SCSI IDs.

1. Log in to the ME4 controller using **ssh**.

   ```
   # ssh manage@IP-address
   ```

2. Execute the CLI command to show the volume information. The following example shows volumes that match the pattern **r730xd-1-*** and only display the non-snapshot volumes. The **details** argument includes extended attributes including the **WWN** of the volumes.

```
# show volumes details type base pattern r730xd-1-*
Pool Name              Total Size Alloc Size Class    Serial Number
WR Policy  Cache Opt Read Ahead Type Tier Affinity
Snap-Pool                        Desc WWN
  Large Virtual Extents  Role             Health Reason Action
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-----------------------------------------------------------------------------
A    r730xd-1-vol0001  200.0GiB   176.0MiB   Virtual
00c0ff3c589f00001e4b6b5b01000000  write-back standard  Adaptive   base No
Affinity   medium
600C0FF0003C589F1E4B6B5B01000000
  Disabled                            OK
A    r730xd-1-vol0002  150.0GiB   124.0MiB   Virtual
00c0ff3c589f000097596b5b01000000  write-back standard  Adaptive   base No
Affinity   medium
600C0FF0003C589F97596B5B01000000
  Disabled                            OK
A    r730xd-1-vol0003  150.0GiB   52.0MiB    Virtual
00c0ff3c589f000098596b5b01000000  write-back standard  Adaptive   base No
Affinity   medium
600C0FF0003C589F98596B5B01000000
  Disabled                            OK
-------------------------------------------------------------------------------
-------------------------------------------------------------------------------
-----------------------------------------------------------------------------
```

3.2 Multipathing

Multipathing is a software solution implemented at the host operating-system level. While multipathing is optional, it provides path redundancy, failover, and performance-enhancing capabilities. Therefore, it is highly recommended to deploy the solution in a production environment or any environments where availability and performance are critical.

The main benefits of using an MPIO solution include the following:

- Increase server availability by providing automatic path failover and failback
- Enhance I/O performance by providing automatic load balancing and capabilities for multiple parallel I/O paths
- Ease administration by providing persistent user-friendly names for the storage devices across cluster nodes

3.2.1 Native Linux multipath (device-mapper-multipath)

The native Linux multipath solution is supported and bundled with most popular Linux distributions in use today. Because the software is widely and readily available at no additional cost, many administrators prefer using it compared to other third-party solutions.

DELLEMC

> **Note:** Dell EMC recommends using the native Linux multipath solution with ME4 Series storage.

3.2.2 Configure multipath for ME4 Series storage

The native Linux multipath software comes with a set of default settings for an extensive list of storage models from different vendors. The default settings allow the software to work with the ME4 Series system right out of the box. However, these settings might not be optimal for all situations and should be reviewed and modified if necessary. It is highly recommended to create a specific ME4 Series device section in the configuration file. An example configuration is shown as follows:

```
defaults {
    find_multipaths yes
    user_friendly_names yes
}

blacklist {
}

devices {
device {
    vendor "DellEMC"
    product "ME4"
    path_grouping_policy "group_by_prio"
    path_checker "tur"
    hardware_handler "1 alua"
    prio "alua"
    failback immediate
    rr_weight "uniform"
    path_selector "service-time 0"
 }
}

multipaths {
    multipath {
            wwid "3600c0ff0003c589f1e4b6b5b01000000"
            alias app-data-01
    }
}
```

DELLEMC

Note the following in the prior example:

- The **device** section identifies the ME4 Series system by its vendor and product tags.
- It is recommended to enable **user_friendly_names**.
- It is recommended to enable Asymmetric Logical Unit Access (ALUA). See section 3.2.3.
- The **path_selector** can either be service-time or round-robin.
- When many volumes are mapped on a host, it might be beneficial to create device aliases to make it easier to find and organize volumes on the host. Use the volume WWN to create multipath device aliases. See the **multipaths** section in the prior example. The ME4 Series volume with the WWN of **3600c0ff0003c589f1e4b6b5b01000000** has a multipath device named **/dev/mapper/app-data-01** to reflect the purpose of the volume, instead of the generic name **/dev/mapper/mpathb**.

3.2.3   Asymmetric Logic Unit Access

ME4 Series storage uses Unified LUN Presentation (ULP), which can expose all LUNs through all host ports on both controllers. The storage system appears as an active-active system to the host. The host can choose any available path to access a LUN regardless of disk group ownership. When ULP is in use, the controllers' operating/redundancy mode is shown as active-active ULP. ULP uses the Asymmetric Logic Unit Access (ALUA) extensions to negotiate paths with the ALUA-aware hosts. If the host is not ALUA-aware, all paths are treated as equal by the host even though some paths might have lower latency than others.

To make the Linux host ALUA-aware, set the **path_grouping_policy**, **hardware_handler**, and **prio** parameters accordingly in the **/etc/multipath.conf** file. The prior example in section 3.2.2 shows the proper settings to enable ALUA awareness for ME4 Series storage.

Note: Dell EMC recommends enabling ALUA awareness in Linux multipath to optimize the I/O paths.

After modifying the multipath.conf file, reload the configuration.

```
# multipath -r
```

When the ALUA setting is enabled, the path information should look similar to the following example. Two path groups are formed. Each group consists of one or more paths. The optimized path group has a higher priority, **50** in the following example, than the non-optimized path group which has a lower priority of **10**. The optimized path group also has active status while the status of the non-optimized path group is enabled.

```
# multipath -ll mpathb
mpathb (3600c0ff0003c589f1e4b6b5b01000000) dm-6 DellEMC ,ME4
size=200G features='0' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| |- 18:0:5:0 sdd 8:48  active ready running
| `- 19:0:5:0 sdj 8:144 active ready running
`-+- policy='service-time 0' prio=10 status=enabled
  |- 18:0:4:0 sdl 8:176 active ready running
  `- 19:0:4:0 sdi 8:128 active ready running
```

3.3   LUN partition

In general, partitions (and partition tables) are not required for volumes other than the boot volume or for applications with very specific requirements. Dell EMC recommends using the whole volume instead of

DELLEMC

partitioning the volume wherever possible. This eases future volume expansion and reduces administration overhead for maintaining the partitions.

If partitioning is required, it is recommended to align the partition on the 1 MB boundary. ME4 Series storage uses internal disk pool chunk sizes ranging from 64k to 512k. The 1 MB boundary works well because it is a multiple of the chunk size. Misalignment can result in extra I/Os to the volumes and degrades overall performance. Either **fdisk** or **parted** can be used to create the partition. However, only **parted** can create partitions larger than 2 TB. The parted method is described as follows.

1. Create a partition label.

    ```
    # parted /dev/mapper/mpathb mklabel gpt
    Warning: The existing disk label on /dev/mapper/mpathb will be destroyed
    and all data on this disk will be lost. Do you want to continue?
    Yes/No? Yes
    Information: You may need to update /etc/fstab.
    ```

2. Create one partition with optimal alignment at 1 MB boundary.

    ```
    # parted -a opt /dev/mapper/mpathb mkpart primary 2048s 100%
    Information: You may need to update /etc/fstab.
    ```

3. Verify the partition alignment.

    ```
    # parted /dev/mapper/mpathb align-check opt 1
    1 aligned
    ```

3.4 Linux LVM

Linux Logical Volume Manager (LVM) is a common general-purpose storage manager included in all popular Linux distributions. When working with LVM and ME4 Series storage, LVM mirroring is not necessary because storage protection is already provided by the different RAID levels within the ME4 Series system. A LVM volume group is a collection of one or more LUNs created on the ME4 volumes. Logical volumes are created that span across the LUNs within the volume group.

3.4.1 LVM guidelines

Dell EMC recommends the following LVM guidelines:

- Use whole LUNs for logical volume groups.
- Create separate volume groups for different applications data. For example, if there are two databases, separate the databases into logical volume groups. This simplifies management and allows greater flexibility on array-based snapshots on individual applications.
- It is common to use two or more LUNs in a volume group to increase capacity and improve performance. All LUNs in a logical volume group should be of the same size. To realize the full performance benefit, evenly distribute the LUNs in different ME4 Series disk groups across the two controllers.
- When using ME4 Series volume snapshots to protect the data in a logical volume group, the snapshots must be taken on all LUNs in the same volume group at the same time to guarantee the consistency of the data. It is a best practice to pause/stop the applications that are accessing the LUNs before taking any snapshots.

- Snapshot schedules can be created on individual volumes in ME Storage Manager. Currently, it is not possible to create a schedule against a group of volumes. Therefore, it becomes a challenge to coordinate taking consistent snapshots for multiple volumes. One option is to write a script that incorporates the necessary steps to take a consistent snapshot. The following shows an example of the high-level tasks to be included in the script.

    a. Connect to the host.
    b. Pause/stop the applications.
    c. Connect to the ME4 Series system.
    d. Identify the volumes to create snapshots.
    e. Create snapshots on the volumes.
    f. Connect to the host.
    g. Resume/start the applications.

- Schedule the script to run periodically using an external scheduler such as Linux cron or other third-party commercial schedulers.

3.4.2 Physical volume data alignment

It might be necessary to specify data alignment on LVM physical volumes (PV) to avoid data misalignment. Since ME4 Series storage uses an internal chunk size ranging from 64k to 512k, it is recommended to initialize LVM PV with a 1 MB alignment. Use **pvcreate --dataalignment 1m** to indicate the alignment starts at 1 MB.

```
# pvcreate --dataalignment 1m /dev/mapper/mpathb
```

3.4.3 Expand storage for LVM

Adding capacity to an existing LVM volume group can be performed in one of the two ways:

- Create new ME4 Series volumes and initialize them into the LVM volume group. It is recommended that the new volumes are of equal capacity and have the same back-end storage type as the existing volumes in the volume group. Once the volume group is expanded, the logical volume can be resized to take up the new capacity. The advantage of this method is that existing volumes are not changed. Only new volumes are introduced in the environment and can be backed out easily.
- Dynamically increase the size of existing ME4 Series volumes in ME Storage Manager. ME4 Series storage can resize the virtual volumes without causing any application interruptions. On the Linux host, resize the physical and multipath devices on the OS, then resize the logical volume and file system. Depending on the file system type, the resize tasks can be done online without disrupting the applications. Section 3.5.3 provides an example of this procedure. This method involves fewer administrative tasks because there are no new volumes, new mappings, new snapshot schedules, or new LVM physical devices to manage and set up.

3.5 File systems

ME4 Series storage supports a wide range of Linux file systems. This section focuses on two popular and stable file systems: ext4 and xfs.

D¢LLEMC

3.5.1 File system layout

The file system can be created on top of a LUN, a LUN partition, or a logical volume within LVM. Dell EMC recommends using the whole LUN without a partition or logical volume in LVM for ease of management. If the file systems reside in LVM, follow the LVM guidelines when creating and grouping multiple file systems.

3.5.2 File system mount options

When mounting a file system, consider the following options and guidelines:

- Identify the file system by its UUID or LVM LV device in the **/etc/fstab** file. Avoid using any non-persistent device paths such as /dev/sd*.
- UUID can be queried by **blkid** command.

```
# blkid /dev/mapper/vgme4-lvme4
/dev/mapper/vgme4-lvme4: UUID="ad40e2c2-8ce0-42c6-94bb-9854b73f5336" TYPE="xfs"
```

- Include **nofail** in the mount option if the Linux operating system experiences mount issue during system boot. This prevents interruption during the boot process which requires manual intervention.
- For the xfs file system, the file system check (fsck option) should be disabled in /etc/fstab because it does not perform any check or repair automatically during boot time. The xfs journaling feature ensures the file system integrity and data is in a consistent state after abrupt shutdown. If a manual repair or check is necessary, use the **xfs_repair** utility to repair damaged file system.
- To disable fsck check, set a value of **0** in the sixth field for the file system entry in /etc/fstab. Here is an example of an xfs file system entry in /etc/fstab:

```
UUID="ad40e2c2-8ce0-42c6-94bb-9854b73f5336" /me4_fs xfs defaults,discard,nofail 0 0
```

3.5.3 Expand storage for the file system

Certain file system types, such as ext4 and xfs, support the online resize operation. The following procedure outlines the general steps to resize a file system online assuming non-partition LUNs are used.

1. Expand the size of existing LUNs in ME Storage Manager.
2. Perform a SCSI scan on the host systems, refresh the partition table on each LUN path, and reload multipath devices.

   ```
   # rescan-scsi-bus.sh --resize
   ```

3. Reload the multipath devices. Skip this step if multipath devices are disabled.

   ```
   # multipathd -k"resize map mpathb"
   ```

4. Expand the logical volume if the file system is on top of LVM. Skip this step if the file system does not reside in LVM.

   ```
   # lvresize –L $NEW_SIZE /dev/vgme4/lvme4
   ```

5. Extend the file system size to the maximum size, automatically and online

   ```
   # xfs_growfs –d /me4_fs (for xfs)
   ```

   or

   ```
   # resize2fs /dev/mapper/mpathb (for ext4)
   ```

DELLEMC

3.6   Using snapshots for data protection

ME4 Series storage can create multiple point-in-time snapshots of virtual volumes. This section discusses snapshots and some use cases.

Consider the following when deciding to use snapshots to protect data on virtual volumes.

- Do not enable snapshots for all volumes. Carefully consider what data requires this type of protection and plan accordingly. The maximum number of mappable snapshots per system is 1,024. When the limit is reached, no snapshots can be created until existing snapshots are deleted.
- If replication is enabled between two ME4 Series systems, the replication process creates internal snapshots which are not visible to the administrator. These snapshots do not count against the per-volume limit but do count toward the system limit.
- Snapshots can be created manually, or created by a pre-defined schedule on a per-volume basis.
- It is not possible to define a common snapshot schedule across multiple volumes. Therefore, if a consistent snapshot is required for a group of volumes, manual coordination must be done from applications and through the use of scripting to synchronize and automate the multi-step process.
- Snapshots are treated like any other volumes. They can be mapped to hosts and made read-only or read-write.
- Taking snapshots is fast and efficient. They consume no space initially until data is changed in the snapshots.
- Snapshots are logical copies of the source volumes. They are dependent on the source volumes, and I/Os to the snapshots might affect I/Os to the source volumes. Limit access to the number of copies of snapshots created from the same source volume and monitor performance carefully over time to determine the actual impact to the environment.

Refer to the ME4 Series *Administrator's Guide* for a detailed description of the features.

3.6.1   Mounting snapshots to recover data

Snapshots are useful to recover data that are deleted, corrupted, or changed unintentionally on the source volumes. Mount a snapshot created before the unintentional change happened on an alternate host and retrieve the data using the appropriate tools, such as cp, dd, or application-specific utilities. If data resides on multiple volumes, the corresponding snapshots of the volumes must be made available to the host.

If the snapshots are mapped to the same host as the source host, additional steps are necessary to mount the snapshots because the metadata on the snapshot volumes are exactly the same as the source volumes. These metadata such as disk label, UUID, and volume group metadata, conflict with existing mounted volumes. It is recommended to mount snapshots on an alternate host instead. However, if the same source host must be used, use the following procedures to modify the metadata on the snapshots. The snapshot volumes need to be mapped to the host with read-write access.

1. Use the **blkid** command to examine the UUIDs on both source and snapshot volumes.
2. If the snapshot volumes contain LVM, change the LVM VG UUID during the import.
3. After mapping the snapshot volumes and scanning the host, identify the snapshot volumes and their device files (see section 3.1). Use the **vgimportclone** command to properly change the UUIDs of the snapshot volume group and physical volumes.

```
# vgimportclone --basevgname vgme-snapshot /dev/mapper/mpathz
```

4. Change the snapshot volume's UUID. This step is not necessary if step 2 has been executed.

    a. Use the **uuidgen** command to generate a new unique UUID on the host.

    b. For xfs, use the **xfs_admin -U** command to change the filesystem UUID to the new UUID.

```
# uuidgen
09d56888-eb9f-4a8d-9e9e-bca028edf4ef

# xfs_admin -U 09d56888-eb9f-4a8d-9e9e-bca028edf4ef /dev/mapper/mpathe
Clearing log and setting UUID
writing all SBs
new UUID = 09d56888-eb9f-4a8d-9e9e-bca028edf4ef
```

    c. For ext4, RHEL 7.5 allows mounting the filesystem even though it has a duplicate UUID. It is recommended to change the UUID regardless to avoid future confusion and conflicts.

    Use the **uuidgen** and **tune2fs -U** command to change the filesystem UUID to the new UUID.

```
# uuidgen
6028cf0d-9a0a-4150-99ab-3ff35c6fd550

# tune2fs /dev/mapper/mpathg -U 6028cf0d-9a0a-4150-99ab-3ff35c6fd550
tune2fs 1.42.9 (28-Dec-2013)

# blkid /dev/mapper/mpathg
/dev/mapper/mpathg: UUID="6028cf0d-9a0a-4150-99ab-3ff35c6fd550"
TYPE="ext4"
```

5. Mount the file system using the normal process.

Note: Without changing the UUIDs, the snapshot volume group and file system cannot be mounted on the same system.

3.6.2 Using snapshots to roll back or reset a volume

The rollback feature replaces the data of a source volume/snapshot with the data of another snapshot that is created from it. This is useful to quickly restore a point-in-time copy of data on the original host without reconfiguring the applications on the host.

The reset feature replaces the data of a snapshot with the current data of the source volume. A common use case would be to resynchronize the test data with the most up-to-date data from the production source volume.

Since the rollback and reset features preserve the mapping and identity of the volume, there is no reconfiguration needed on the host. Although I/O access to the volume is restricted while the rollback or reset is in progress, it is highly recommended to dismount the volume on the host and restrict user access to the data. For filesystem, use the **umount** command to dismount. For LVM, use the **vgchange -a n** command to disable the volume group. If the volumes are controlled by applications, such as Oracle ASM, stop the applications using application-provided utilities. Once the process is complete, reactive LVM, file systems, and applications.

DELLEMC

# A	Technical support and resources

Dell.com/support is focused on meeting customer needs with proven services and support.

Storage solutions technical documents provide expertise that helps to ensure customer success on Dell EMC storage platforms.

## A.1	Related resources

The following ME4 Series publications and additional resources are available at Dell.com/support.

- Administrator's Guide
- Deployment Guide
- CLI Guide
- Owner's Manual
- Support Matrix

Also, see the following third-party referenced or recommended publications and articles:

- Red Hat Product Documentation Portal
- Oracle Product Documentation Portal

DELLEMC