Dell EMC Host Connectivity Guide for Linux

DELLEMC

CONTENTS

Contents

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.

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

Purpose

This guide describes the features and setup procedures for Linux host interfaces to Dell EMC PowerMax/VMAX All Flash/VMAX3 Series, Dell EMC VNX series, EMC VNXe series, Dell EMC Unity series, Dell EMC XtremIO, Dell EMC VPLEX, Dell EMC PowerStore and storage systems over Fibre Channel and (Symmetrix only) SCSI.

Audience

This guide is intended for use by storage administrators, system programmers, or operators who are involved in acquiring, managing, or operating Dell EMC PowerMax/VMAX All Flash/VMAX3 Series, Dell EMC VNX series, EMC VNXe series, Dell EMC Unity series, Dell EMC XtremIO, Dell EMC VPLEX, Dell EMC PowerStore, and host devices in Linux operating environment.

Related documentation

For the documentation referred in this guide, go to Dell EMC Online Support.

Special notice conventions used in this document

Dell EMC uses the following conventions for special notices:

⚠ **WARNING** 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.

(i) **Note:** Presents information that is important, but not hazard-related.

Typographical conventions

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

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 Service Center. 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.

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 techpubcomments@dell.com.

CHAPTER 1

Introduction to Linux Environment

This chapter includes the following topics:

Operating system limits and guidelines

This section provides operating system limits and restrictions that are imposed in a SAN environment. Factors such as number of supported Host Bus Adapters (HBAs), Logical Unit Numbers (LUNs), scalability of targets, file system, and volume management limits are discussed in this section.

Host initiators

On all Linux environments, Dell EMC supports up to 16 Fibre Channel (FC) initiator ports on a single host. The host initiators may be single or dual channel host bus adapters (HBAs). The number of host initiator ports on a server is also limited by the number of HBA slots available on the server and supported by the server vendor.

(i) **Note:** Dell EMC does not support the mixing of HBAs from different vendors.

(i) **Note:** Dell EMC PowerPath stipulates a maximum of 32-paths to a single LUN.

Logical unit

The number of logical units seen by a host system depends on the SCSI scan algorithm that is employed by the operating system and the LUN scan limits that are imposed by the host bus adapter.

The HBA initiator and host system limits are theoretical maximums. The following table provides these limits:

Table 1 Maximum SCSI devices

Operating system	Per initiator devices	Host system devices	Host system devices
Asianux 4.0	Emulex: 65536 QLogic: 65536	65536	8192
OL 5.0 [a]	Emulex: 65536 QLogic: 65536	65536	1024
OL 6.0	Emulex: 65536 QLogic: 65536	65536	8192
OL 7.0	Emulex: 65536 QLogic: 65536	65536	8192
OL 8.0	Emulex: 65536 QLogic: 65536	65536	8192
RHEL 5 [a]	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	1024
RHEL 6	Emulex: 65536 QLogic: 65536	65536	8192

Table 1 Maximum SCSI devices (continued)

Operating system	Per initiator devices	Host system devices	Host system devices
	Brocade: 256		
RHEL 7	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	16384
RHEL 8	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	16384
SLES 10 [b]	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	1024
SLES 11	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	8192
SLES 12 SLES 15 SLES 15 SP1	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	16384

a. Dell EMC supports up to 8192 Linux Native SCSI devices on RHEL 5.4 and later.
b. Dell EMC supports up to 8192 Linux Native SCSI devices on SLES 10 SP3 and later.

ⓘ **Note:** A related limitation is the highest LUN instance or device number that a host system can address. This number depends on the ability of the HBA to address high LUN numbers and the total number of SCSI devices that have been exposed to the host system.

The HBA addressable LUN ID is the theoretical maximum. The following table provides these limits, and the supported limits for Dell EMC storage attach.

Table 2 Highest addressable LUN ID

Operating system	HBA addressable LUN ID	Dell EMC supported
Asianux 4.0	256 - Default; 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
OL 5.0	256 - Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
OL 6.0	256 - Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
OL 7.0	256 - Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)

Table 2 Highest addressable LUN ID (continued)

Operating system	HBA addressable LUN ID	Dell EMC supported
OL 8.0	256 - Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
RHEL 5	256 - Default, 32768 - Max (Emulex) 65536 (QLogic) 256 (Brocade)	16384 (Emulex) 16384 (QLogic) 256 (Brocade)
RHEL 6	256 - Default, 32768 - Max (Emulex) 65536 (QLogic) 256 (Brocade)	16384 (Emulex) 16384 (QLogic) 256 (Brocade)
RHEL 7	256 - Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
RHEL 8	256 - Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
SLES 10	256 - Default, 32768 - Max (Emulex) 65536 (QLogic) 256 (Brocade)	16384 (Emulex) 16384 (QLogic) 256 (Brocade)
SLES 11	256 - Default, 32768 - Max (Emulex) 65536 (QLogic) 256 (Brocade)	16384 (Emulex) 16384 (QLogic) 256 (Brocade)
SLES 12 SLES 15 SLES 15 SP1	256 - Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)

Configuration example

The following hardware requirements are needed:

Server requirement:

- Must support RHEL 5.4, SLES 10 SP3, or SLES 11 minimum.
- Should have at least 16 GB of memory or more, dependent on the application.

Configuring multiple LUNs for RHEL 5.4

The following is an example of how to configure for 8192 Linux native SCSI devices on RHEL 5.4.

About this task

ⓘ Note:

- This same process can be used for either RHEL 5.4 or later, SLES 10 SP3 or later, or SLES 11 or later, except the number n would be 8192 for the other operating systems.

- For RHEL 6 and later, See Multiple LUNs for RHEL 6.0 and later.

For PowerPath, skip Step 1 and go to Step 2. There are no changes needed when using PowerPath.

Procedure

1. Modify the `max_fds` parameter in the `/etc/multipath.conf` file to accommodate 8,192 Linux native SCSI devices.

   The following screenshot is an example of the `/etc/multipath.conf` file that works with 8,192 Linux native SCSI devices.

2. Modify following options in `/etc/modprobe.conf` file to accommodate 8,192 Linux native SCSI devices:

   `n` is the desired number of maximum Linux native SCSI devices

   - SCSI module options:
     - max_report_luns = n
     - mod_max_luns = n-1

   options scsi_mod max_luns=8191 max_report_luns=8192

   - For Emulex: Emulex lpfc driver options:
     - lpfc_max_luns = n
     - options lpfc lpfc_max_luns= 8192
   - For QLogic:

   There are no parameters in the driver to change.

   - For Brocade:

   The Brocade driver does not support high device counts at this time.

The following screenshot shows an example of `/etc/modprobe.conf` file that can accommodate 8,192 Linux native SCSI devices.

3. Make the changes in `/etc/modprobe.conf` permanent by creating a new ramdisk image.

   For RHEL 5.4, use the following command:
   ```
   mkinitrd -f /boot/initrd-<kernel-version >.img <kernel-version>
   ```

4. Reboot the server for the new parameters to take effect.

Multiple LUNs for RHEL 6.0 and later

`scsi_mod` is now built into the kernel and is no longer a loadable module as in previous versions. Therefore, module options cannot be changed in RHEL 6 by adding a .conf file entry within the `/etc/modprode.d` directory. Settings should go on the kernel command line.

Procedure

1. Append the following to your grub.conf 'kernel' line (`/etc/default/grub`):

   ```
   scsi_mod.max_luns= n
   ```
   The default setting for `scsi_mod.max_luns` (SCSI mid layer) is 512. This can be checked with the following command.

   ```
   # cat /sys/module/scsi_mod/parameters/max_luns
   ```

   - For QLogic: 16384

   The default setting for `qla2xxx.ql2xmaxlun` is 65535. This can be checked with the following command:

   ```
   # cat /sys/module/qla2xxx/parameters/ql2xmaxlun
   ```

   - For Emulex:

   The default setting for `lpfc.lpfc_max_luns` (Emulex HBAs) is 255. This can be checked with the following command.

   ```
   #cat /sys/module/lpfc/parameters/lpfc_max_luns
   ```

2. Some new arrays also require the report LUNs entry value be set. In such case, also append it to your grub.conf kernel line:

   ```
   scsi_mod.max_report_luns= n
   ```

3. Reboot the system. After the reboot, the LUNs should appear.

Multiple LUNs for RHEL 7.0 and later

Procedure

1. Modify `/etc/default/grub` to add the highlighted text in the following figure:

```
[root@lin011148 ~]# more /etc/default/grub
GRUB_TIMEOUT=5
GRUB_DISTRIBUTOR="$(sed 's, release .*$,,g' /etc/system-release)"
GRUB_DEFAULT=saved
GRUB_DISABLE_SUBMENU=true
GRUB_TERMINAL_OUTPUT="console"
GRUB_CMDLINE_LINUX="vconsole.font=latarcyrheb-sun16 vconsole.keymap=us crashkernel=auto  selinu
x=0 rhgb quiet scsi_mod.max_luns=32768 scsi_mod.max_report_luns=32768"
GRUB_DISABLE_RECOVERY="true"
```

2. Modify `/etc/security/limits.conf` by adding the highlighted text:

```
#@student        -       maxlogins       4
*                       hard    nofile          20000
```

3. Modify `/etc/modprobe.d/lpfc.conf` by adding the highlighted text:

```
[root@lin011148 ~]# cat /etc/modprobe.d/lpfc.conf
options lpfc lpfc_max_luns=32768
```

(i) Note: For QLogic, no modification in modprobe is required.

4. To reduce logging noise, modify *inotify* configuration by adding the following text to `/etc/sysctl.conf`:

```
# For more information, see sysctl.c
fs.inotify.max_user_watches=16384
[root@lin011148 ~]#
```

5. The configuration changes in the previous steps must be compiled into the initrd by executing the following:

```
grub2-mkconfig -o /boot/grub2.cfg
dracut -f
reboot
```

Storage attach

This section describes fan-in and fan-out.

Fan-in

The Dell EMC VNX series supports scanning of a maximum of 4 VNX series and 32 VNX series SP ports (whichever is lesser) per host initiator port. The Dell EMC Unity series and VNXe series supports scanning of a maximum of 16 Unity series and VNXe series systems in replication or 32 Unity series and VNXe series ports per host initiator port.

While the Dell EMC PowerMAX and VMAX series do not impose such a restriction, currently a SCSI scan of up to 32 FA ports from a single initiator port has been qualified and is supported.

The following figure shows an example of fan-in:

Figure 1 Fan-in: 1 HBA port to *n* Dell EMC arrays

Fan-out

The host bus adapter also imposes limits on the number of distinct target ports (distinct WWPN) that the scanning algorithm attempts to scan. On Emulex, this limit is set to 256 and on QLogic this limit is set to 512. These limits are theoretical limits as exported by the host bus adapter. The following figure shows an example of fan-out:

Figure 2 Fan-out: *n* HBA ports to 1 Dell EMC array
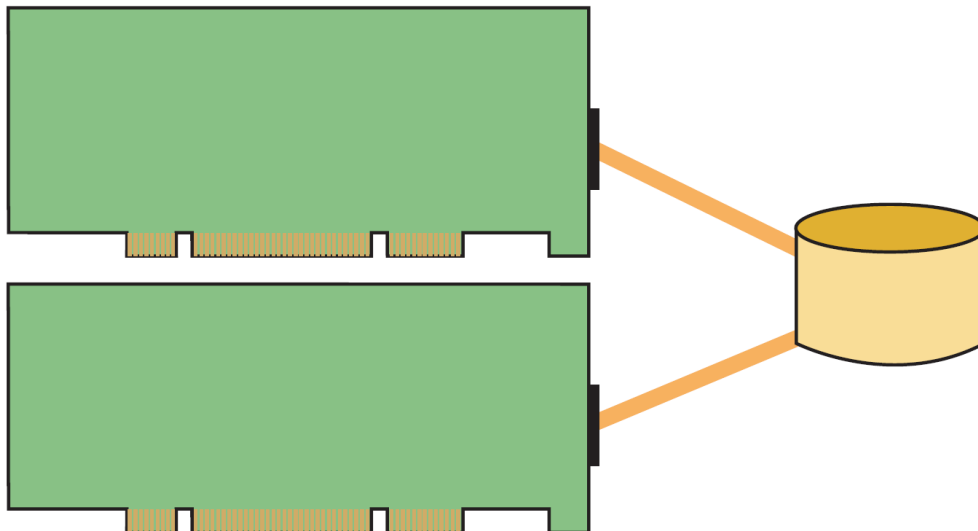
ⓘ **Note:** The time to boot the Linux operating system depends on the number of SCSI devices and targets exposed. With many SCSI devices, the boot process is noticeably longer.

Zoning recommendation

When using Linux hosts in a fabric environment, the recommended zoning methodology is single-initiator zoning. A single-initiator zone consists of only one Host Bus Adapter port. While multiple array target ports may be part of the same zone, it is recommended that a single zone should not contain target ports from multiple arrays.

When configuring zoning/subnetting from the host to the XtremIO cluster, the minimal zoning/ subnetting configuration for each host Initiator Group should be at least one path for two Storage Controllers belonging to the same X-Brick. A host port must be zoned to at least two Storage Controllers ports from the same X-Brick. For detailed information, see *Dell EMC XtremIO Storage Array Host Configuration Guide* on Dell EMC Online Support.

(i) **Note:** A single zone should not contain multiple initiator ports.

(i) **Note:** Multiple target ports from multiple arrays are supported in a single zone.

File systems and feature limitations

File system is the general name that is given to the host-based logical structures and software routines that are used to control storage, organization, manipulation, and retrieval of data. File systems map underlying disk sectors into logical data blocks, store the data, keep track of data location, and ensure that data is easy to find and access once needed.

The Linux file system is an ordered tree-like hierarchical structure that is composed of files and directories. The trunk of the tree structure starts at the root directory. Directories that are one level below are preceded by a slash, and they can further contain other subdirectories or files. Each file is described by an inode, which holds location and other important information of the file. A Linux file system is made available to users by mounting it to a specific mounting point.

Filesystem support

Dell EMC qualified and supported Linux file systems are listed in the following table:

Table 3 Supported file system

	RHEL	OL	SuSE
Ext3	Yes	Yes	Yes
Ext4	> RHEL 5.3	Yes	≥SLES 11 SP2
Btrfs	≥RHEL 7 ＜RHEL 8	≥OL 7 OL 7	≥SLES 12
GFS2	>RHEL 5.2	Yes	No
XFS	≥RHEL 5.5	>OL 6	Yes
OCFS2	No	Yes	≥SLES 10 SP2
VxFS	Yes	Yes	Yes

Red Hat Global File System (GFS) and Red Hat Cluster Suite (RHCS) are part of RHEL5 and Oracle Linux 5 and are supported by Oracle under the Linux Support Program. However, since GFS and RHCS are not included with RHEL4, Oracle Linux 4, and earlier versions, they are not supported by Oracle with RHEL4, OL4, and earlier versions. Beginning with Red Hat Enterprise Linux 6, several features were separated into add-ons, requiring a separate purchase, such as the High Availability Add-On for clustering and the Resilient Storage Add-On for GFS2. Oracle Linux Support does not include support for these add-ons.

The Red Hat Scalable File System Add-on is a solution which incorporates the Linux XFS file system and is available, for an additional cost per socket-pair, with the Red Hat Enterprise Linux Server subscription. Since RHEL8, the Brtfs file system is not supported. Oracle Linux customers with Premier Support subscriptions can receive support for XFS on Oracle Linux 6 at no additional charge. Beginning with Oracle Linux 7, XFS is the default file system and is included with Basic and Premier Support subscriptions at no additional charge. This support includes both the Unbreakable

Enterprise Kernel (UEK) and the Red Hat compatible kernel. For the Unbreakable Enterprise Kernel, you must use Release 2 or later.

Features and limitations

The following tables summarize file system features and limitations.

Table 4 Local file system features and limitations

Local file system	Compatibility	Capacity		Data structure	Journaling		Allocation techniques
		Max. file size	Max. volume size		Block	Metadata only	
Ext2		16 GB - 2 TB*	16 GB - 32 TB*	Block-mapping scheme	No	No	Sparse files
Ext3		16 GB - 2 TB*	16 GB - 32 TB*	Block-mapping scheme	Yes	Yes	Sparse files
Ext4	ext2, ext3	16 GB - 16 TB*	1 EB	Extent	Yes	Yes	Sparse files Persistent pre-allocation Delayed allocation
ReiserFS		8 TB	16 TB	B+ tree, tail packing	No	Yes	Sparse files
XFS	CIFS, NFS V3&V2	9 EB (64 bit)	9 EB (64 bit)	B+ tree, extent	No	Yes	Sparse files Striped allocation Delayed allocation
Btrfs		8 EB	16 EB	B tree	No	No	Integrated RAID

Table 5 Cluster file system features and limitations

Cluster file system	Compatibility	Capacity			Data structure	Journaling		Allocation techniques
		Max. file size	Max. volume size	Max. node number		Block	Metadata only	
OCFS2	Versions above OCFS2 1.4 are compatible with OCFS2 1.4	4 PB	4 PB	100+	Extent	Yes	Yes	Sparse files Pre-allocation
GFS	Upgrading of GFS2 from GFS is possible	2 TB - 8 EB	2 TB - 8 EB	100+	Block-mapping scheme	Yes	Yes	Sparse files
GFS2		2 TB - 8 EB	2 TB -8 EB*	Cluster/ Standalone	Block-mapping scheme	Yes	Yes	Sparse files Parallel-allocation

Table 5 Cluster file system features and limitations (continued)

Cluster file system	Compatibility	Capacity			Data structure	Journaling		Allocation techniques
		Max. file size	Max. volume size	Max. node number		Block	Metadata only	
VxFS		2^63 Byte (8 EiB)	2^77 Byte (128 ZiB)		Extent	Yes	No	Extent Sparse files

Linux volume managers

LVM

A logical-volume manager (LVM) is a utility that enables you to manage your disk space through user-defined logical volumes. Most LVMs can manage multiple GB of disk space. LVMs enable you to partition a disk drive into multiple volumes so that you can control where data is placed on a given disk.

LVM for the Linux operating system manages volume disk drives and similar mass-storage devices. It is suitable for managing large hard-disk farms by enabling you to add disks, replace disks, and copy and share contents from one disk to another without disrupting service, and you can resize your disk partitions easily as needed.

LVM allocates hard drive space into logical volumes that can be easily resized, instead of partitions. A volume manager can concatenate, stripe together, or otherwise combine partitions into larger virtual ones that can be resized or moved, while it is used.

The maximum LV size is 2 TB. For 32-bit CPUs on 2.6 kernels, the maximum LV size is 16 TB. For 64-bit CPUs on 2.6 kernels, the maximum LV size is 8 EB. Consult your distribution for more information.

Veritas VxVM and VxFS

Veritas Volume Manager (VxVM) and Veritas Filesystem (VxFS) are included as part of the Veritas Storage Foundation product. VxVM 4.x and 5.x are supported on RHEL and SLES Linux distributions.

VxVM is a storage management subsystem that enables you to manage hard drives as logical devices that are called volumes. A VxVM volume appears to applications and the operating system as a hard drive partition device on which file systems, databases, and other managed data objects can be configured. It also provides easy-to-use online disk storage management for computing environments and storage area network (SAN) environments. VxVM can be configured to protect against disk and hardware failure, and to increase I/O throughput as it supports the Redundant Array of Independent Disks (RAID) model. VxVM also provides features that enhance fault tolerance and fast recovery from disk failure.

For detailed documentation about Veritas VxVM and VxFS, see the Symantec website.

EVMS

Enterprise Volume Management System (EVMS) is available on SLES 10 and OES-Linux. EVMS provides a single, unified system for zoning. VMS provides a new model of volume management to Linux. EVMS integrates all aspects of volume management, such as disk partitioning, Linux logical volume manager (LVM), multi-disk (MD) management, and file system operations into a single cohesive package. With EVMS, various volume management technologies are accessible through one interface, and new technologies can be added as plug-ins as they are developed.

For detailed documentation about EVMS, see the EVMS website.

LUN limits

The number of logical units that are seen by a host system depends on the SCSI scan algorithm that is employed by the operating system and the LUN scan limits imposed by the host bus adapter.

LUNs are counted by their unique LUN ID. The SCSI middle layer by default supports 512 LUNs. The actual maximum LUN count is capped by the maximum supported LUNs of the HBA driver. The SCSI block device names can run from `/dev/sda` to `/dev/sdzzz`. This is a maximum of 18,278 devices. The detailed algorithm is as follows:

```
1. sda ~ sdz : 26

2. sdaa ~ sdzz : 26x26=676

3. sdaaa ~ sdzzz: 26x26x26=17576

4. total=26+26x26+26x26x26=18278
```

The HBA initiator and host system limits are theoretical maximums. Consult the HBA and operating system vendors for exact limitations.

PATH limits

Device Mapper Multipathing enables an operating system to configure multiple I/O paths between server nodes and storage arrays into a single device. These I/O paths are physical SAN connections that can include separate cables, switches, and controllers. There is path limit for the Linux inherent path management software. The device-mapper-multipath and the kernel support up to 1024 paths per path group and up to 1024 path groups.

Consult your vendor for other path management software limits.

Devices and operations

This section provides an overview of mechanisms that are provided by a Linux operating system for addressing and using SCSI devices.

- SCSI device addressing
- SCSI device operation interfaces

SCSI device addressing

Linux employs a four-attribute scheme to address SCSI devices:

- SCSI adapter number
- Channel number
- Target ID number
- Logical unit number (LUN)

This information is exported to the `/proc` file system and is available for viewing as follows: An example of a VMAX series:

 # cat /proc/scsi/scsi

```
Host:    scsi2            Channel:   00        Id:   00   Lun: 00
Vendor: EMC             Model: SYMMETRIX                    Rev:
5874
```

```
Type:    Direct-Access                                     ANSI SCSI
revision:   04
Host:    scsi2              Channel:    00          Id:    01   Lun: 01
Vendor: EMC                 Model: SYMMETRIX                         Rev:
5874
Type:    Direct-Access                                     ANSI SCSI
revision:   04
```

An example of a VNX series:

 # cat /proc/scsi/scsi

```
Host:    scsi2              Channel: 00               Id:    00   Lun:
00
Vendor: DGC          Model: RAID    5                                Rev:
0219
Type:    Direct-Access                                 ANSI    SCSI
revision:   04
Host:    scsi2              Channel: 00               Id:    01   Lun:
01
Vendor: DGC          Model: RAID    5                                Rev:
0219
Type:    Direct-Access                                 ANSI    SCSI
revision:   04
```

An example of XtremIO:

```
Host: scsi2        Channel: 00           Id: 00 Lun: 00
Vendor: XtremIO   Model: XtremApp                        Rev: 40f0
Type: RAID                              ANSI SCSI revision: 06
Host: scsi2 Channel: 00                 Id: 00 Lun: 01
Vendor: XtremIO   Model: XtremApp                        Rev: 40f0
Type: Direct-Access                     ANSI SCSI revision: 06
```

In the above output, two SCSI devices are seen by the host.

- **Host** implies that the LUNs are seen by SCSI host adapter instance 2 on the system.
- **Channel** refers to the SCSI bus. While the SCSI standards allow for multiple initiators to be present on a single bus, currently a single SCSI bus supports only one initiator on Linux.
- **ID** is the target ID number. This number is incremented for every new target or storage controller port seen by a host initiator.
- **LUN** is the actual logical unit instance assigned to the host.

Additionally, for each of the SCSI devices seen by the system, the above output provides the vendor and model information, the type of device, and the SCSI protocol version.

(i) Note: SCSI devices are identified by their major and minor device numbers. The instances are created in the /dev directory. The INQ utility can be used to correlate bus/target/LUN identifiers to sd device numbers, and thus to major/minor device numbers.

For a detailed review of the Linux SCSI implementation, see Linux 2.4 SCSI subsystem How To.

SCSI device operation interfaces

Linux provides various device operation interfaces. This includes block and character devices and raw device interfaces. With the 2.6 kernel, a new framework for device management, the device-mapper, was introduced. This section describes the concepts of the device-mapper.

Block and character devices

The four high-level device drivers in the SCSI subsystem are:

- sd - Direct access (disks)
- sg - SCSI generic interface
- sr - Data CD-ROMs
- st - Tapes

The sd, sr, and st drivers are block-based devices.

The sg driver is a character-based device that is used primarily for scanners, CD writers, and printers.

Block device

A native device filename for block devices takes the following form:

```
/dev/sdln
```

where l is a letter denoting the hard drive and n is a number denoting the partition on that hard drive

Usually, the partition number is not included when referring to the entire drive. Following this format, the filenames are as follows:

```
/dev/sd[a-z][a-z][1-15]
```

Character device

The corresponding character device filenames take the following form:

```
/dev/sg[n]
```

Where n begins with zero and increments by one.

The use of the alphabetic sg device filenames are now deprecated and are used as links to the sg numeric device filenames. Following this format, the filenames are as follows:

```
/dev/sg[a-z][a-z]
/dev/sg[n]
```

Raw device

Linux also presents a raw device interface for accessing devices. A raw device is a character device that is bound to a block device. With raw devices, the kernel's block buffer cache is entirely bypassed. The Linux utility, raw, provides the ability to access a block device as a raw device.

RHEL implementation

The raw interface is available on RHEL 5.

(i) Note: Although RHEL includes support for rawio, it is now a deprecated interface. It is recommended that any application that uses this interface be modified to open the block device with the O_DIRECT flag.

The raw device controller on RHEL is the `/dev/rawctl` and the raw devices are populated as `/dev/raw/raw<N>`, where <N> is the raw device instance. The man page for raw on Red Hat provides a description of this feature and steps for implementation.

SuSE SLES implementation

The raw interface is available on SLES 10 and SLES 11. The raw device controller on SLES is the /dev/raw/rawctl and the raw devices are populates as `/dev/raw/raw<N>`, where <N> is the raw device instance. The raw interface needs to be started using the initialization script /etc/

init.d/raw. The man page for raw on SLES provides a description of this feature and steps for implementation.

Device mapper

The device-mapper is a generic framework introduced by Linux distributions offering 2.6 kernel-based operating systems. The framework provides a mechanism to map a basic block device into a virtual block device with additional capabilities including striping, concatenation, mirroring, snapshots, and multipath.

Current operating system implementations for device-mapper include support for LVM2, EVMS, Software RAID (dmraid), and Linux native multipath (dm-mpio).

The device-mapper sources are included as part of the default kernel source and the functionality is available on 2.6 kernel-based operating systems, including RHEL 5, RHEL 6, SLES 10, and SLES 11.

Additional information is made available by the operating system distributor in the `/usr/ share/doc/device-mapper <version>` directory.

The device-mapper controller device is located at `/dev/device-mapper`. The device-mapper device instances are created as `/dev/dm-<N>`, where <N> is the instance of the device.

A userspace tool,**dmsetup**, enables the use of the device-mapper controller to create, remove, control, and query dm instances on the system. The man page for **dmsetup** provides detailed implementation guidelines and example use cases.

LUN scanning mechanism

Linux provides multiple mechanisms to rescan the SCSI bus and recognize SCSI devices that are exposed to the system. With the 2.6 kernel and later, significant improvements have been made and dynamic LUN scanning mechanisms are available.

The mechanisms for reconfiguring devices on a Linux host include:

- System reboot

- Unloading and reloading the modular HBA driver

- Echoing the SCSI device list in `/proc`

- Executing a SCSI scan function through attributes exposed to `/sys`

- Executing a SCSI scan function through HBA vendor scripts Each mechanism is discussed further in this section.

ⓘ Note: It is recommended that all I/O on the SCSI devices should be quiesced prior to attempting to rescan the SCSI bus.

System reboot

Rebooting the host allows reliable detection of newly added devices. The host may be rebooted after all I/O has stopped, whether the driver is modular or statically linked.

HBA driver reload

By default, the HBA drivers are loaded in the system as modules. This allows for the module to be unloaded and reloaded, causing a SCSI scan function in the process. In general, before removing the driver, all I/O on the SCSI devices should be quiesced, file systems should be unmounted, and multipath services need to be stopped. If there are agents or HBA application helper modules, they

should also be stopped on the system. The Linux utility **modprobe** provides a mechanism to unload and load the driver module.

SCSI scan function in /proc

Since 2.4 kernel, the `/proc` file system provides a listing of available SCSI devices. If SCSI devices that are exposed to the system are reconfigured, then these changes can be reflected on the SCSI device list by echoing the **/proc** interface.

To add a device, the host, channel, target ID, and LUN numbers for the device to be added to `/proc/scsi/`, scsi must be identified.

The command to be run follows this format:

```
# echo "scsi add-single-device 0 1 2 3" > /proc/scsi/scsi
```

Where:

0 is the host ID

1 is the channel ID

2 is the target ID

3 is the LUN

This command will add the new device to the `/proc/scsi/scsi` file. If one does not already exist, a device filename might need to be created for this newly added device in the `/dev` directory.

To remove a device, use the appropriate host, channel, target ID, and LUN numbers and issue a command similar to the following:

```
# echo "scsi remove-single-device 0 1 2 3" > /proc/scsi/scsi
```

where:

0 is the host ID

1 is the channel ID

2 is the target ID

3 is the LUN

ⓘ Note: This mechanism is deprecated and should not be used in 2.6-based, or later kernels.

ⓘ Note: HBA driver vendors provide scripts that automate the scanning of the SCSI interface. Dell EMC does not provide support for these scripts. Support resides solely with the HBA vendor.

SCSI scan function in /sys

The Host Bus Adapter driver in the 2.6 kernel and later exports the scan function to the /sys directory which can be used to rescan the SCSI devices on that interface. The scan function is available as follows:

```
# cd /sys/class/scsi_host/host4/
# ls -al scan
# echo '- - -' > scan
```

The three dash marks refer to channel, target, and LUN numbers. The above action causes a scan of every channel, target, and LUN visible through host-bus adapter instance '4'.

ⓘ Note: This functionality is available on specific driver versions/operating system combinations only. Contact your Linux distributor for guidance and support of using this technique.

SCSI scan through HBA vendor scripts

QLogic

Use QLogic script to dynamically scan the devices. QLogic has the *QLogic FC HBA LUN Scan Utility* which is available in the Dell EMC-approved section on the QLogic website.

Usage examples

- To rescan all the HBAs, type one of the following commands:
  ```
  # ./ql-dynamic-tgt-lun-disc.sh
  ```
  ```
  # ./ql-dynamic-tgt-lun-disc.sh -s
  ```
  ```
  # ./ql-dynamic-tgt-lun-disc.sh --scan
  ```

- To rescan and remove any lost LUNs, type one of the following commands:
  ```
  # ./ql-dynamic-tgt-lun-disc.sh -s -r
  ```
  ```
  # ./ql-dynamic-tgt-lun-disc.sh --scan --refresh
  ```

- To invoke the menu, type one of the following commands:
  ```
  # ./ql-dynamic-tgt-lun-disc.sh -i
  ```
  ```
  # ./ql-dynamic-tgt-lun-disc.sh --interactive
  ```

Emulex

Use Emulex script to dynamically scan the devices. Emulex has the *LUN Scan Utility* which is available in the Dell EMC-approved section on the Emulex (now Broadcom) website.

Usage examples

```
# gunzip lun_scan.sh.gz
```

```
# chmod a+x lun_scan
```

- To scan all lpfc HBAs:
  ```
  # lun_scan all
  ```

- To scan the lpfc HBA with scsi host number 2:
  ```
  # lun_scan 2
  ```

- To scan the lpfc HBAs with scsi host number 2 and 4:
  ```
  # lun_scan 2 4
  ```

(i) Note: HBA driver vendors provide scripts that automate the scanning of the SCSI interface. Dell EMC does not provide support for these scripts. Support resides solely with the HBA vendor.

SCSI scan through Linux distributor provided scripts

SuSE Linux Enterprise Server (SLES) provides a script named `/bin/rescan-scsi-bus.sh`. It can be found as part of the SCSI utilities package.

```
l82bi094:~ # rpm -qa | grep scsi
yast2-iscsi-server-2.13.26-0.3
yast2-iscsi-client-2.14.42-0.3
open-iscsi-2.0.707-0.44
scsi-1.7_2.36_1.19_0.17_0.97-12.21
xscsi-1.7_2.36_1.19_0.17_0.97-12.21
```

The following is an example from SLES 10 SP2:

```
l82bi094:~ # /bin/rescan-scsi-bus.sh -h
Usage: rescan-scsi-bus.sh [options] [host [host ...]]
Options:
 -l       activates scanning for LUNs 0-7              [default: 0]
 -L NUM   activates scanning for LUNs 0--NUM       [default: 0]
 -w       scan for target device IDs 0 .. 15      [default: 0-7]
 -c       enables scanning of channels 0 1            [default: 0]
 -r       enables removing of devices           [default: disabled]
 -i       issue a FibreChannel LIP reset        [default: disabled]
--remove:           same as -r
--issue-lip:        same as -i
--forceremove:      Remove and readd every device (DANGEROUS)
--nooptscan:        don't stop looking for LUNs is 0 is not found
--color:            use coloured prefixes OLD/NEW/DEL
--hosts=LIST:       Scan only host(s) in LIST
--channels=LIST:   Scan only channel(s) in LIST
--ids=LIST:         Scan only target ID(s) in LIST
--luns=LIST:        Scan only lun(s) in LIST
Host numbers may thus be specified either directly on cmd line (deprecated)
or
  or with the --hosts=LIST parameter (recommended).
LIST: A[-B][,C[-D]]... is a comma separated list of single values and ranges
 (No spaces allowed.)
l82bi094:~ #
```

(i) Note: HBA driver vendors provide scripts that automate the scanning of the SCSI interface. Dell EMC does not provide support for these scripts. Support resides solely with the HBA vendor.

Persistent binding

In a SAN environment with many storage connections, device additions/removals, topology changes, and other events may cause device references to change. Linux device assignments (sd, st, sr, and so forth) are dynamically determined at boot time, and therefore mountpoints based on those devices may or may not be consistent across reboots. For example, the device referred to as /dev/sdc may, or may not, contain the same data when the host is rebooted or the SCSI bus rescanned. In order to ensure that the correct device is referenced at a given mountpoint, persistent binding techniques must be used.

Persistent binding can either be *target-based* or *device-based* (for instance, LUN).

Target-based persistent binding causes the host to scan the available SAN targets in a fixed order, but does not provide persistence for the LUNs under those targets. Therefore, it does not solve the issue of different devices being mounted on a particular mountpoint across reboots.

Device-based persistent binding provides a mechanism to uniquely identify the LUN itself, and therefore references based on device-based identifiers will not change across reboots or reconfigurations.

This section describes the following persistent binding features available on Linux:

- HBA persistent binding

- Udev

- Native MPIO

- PowerPath pseudo-names

- Logical volumes

HBA persistent binding

In Emulex and QLogic drivers, target-based persistent binding feature was available in the driver implementation. Therefore, the host bus adapter would scan for targets in a predefined order defined in a configuration file which would be read at driver load time. This does not provide LUN persistence or stop sd device numbers from changing. See the current Dell EMC HBA documentation provided on the Dell EMC-approved web page of Emulex (now Broadcom) or QLogic for information about how to configure the appropriate drivers.

- *Dell EMC Host Connectivity with QLogic Fibre Channel and iSCSI Host Bus Adapters (HBAs) and Converged Network Adapters (CNAs) for the Linux Environment*

Udev

Udev is a Linux base subsystem feature that is introduced in distributions based on the 2.6 Linux kernel.

Udev(8) provides a dynamic device directory containing only the files for actually present devices. It creates or removes device node files that are usually located in the `/dev` directory. It is part of the hotplug subsystem. Unlike its predecessor *devfs (8)*, *udev (8)* is a user space interface and not a kernel space interface. It is executed if a kernel device is added or removed from the system.

Its configuration file may be found in `/etc/udev/udev.conf`. A list of rules are used, `/etc/udev/rules.d/`, to match against specific device attributes. On device addition, udev (8) matches its configured rules against the available device attributes to uniquely name the device. udev (8) maintains its own database for devices present on the system in `/dev/udevdb`. This database can be queried for the relationship of the kernel device path and the name of the device file via udevinfo (8).

On device removal, udev queries its database for the name of the device file to be deleted. After the device node handling, a list of collected programs specific to this device are executed.

Native MPIO

DM-MPIO, native multipathing, provides a mechanism to address device names persistently through the use of udev and scsi-id. The names used to address multipath names rely on the properties of the physical device, and are thus both unique and consistent across reboots.

PowerPath pseudo-names

The PowerPath pseudo-names are persistent device names that are mapped based on the physical attributes of the storage attach and are thus both unique and consistent across reboots.

Logical volumes

Logical volumes are another mechanism to provide persistent addressing from a host. When a Logical volume is created, a unique signature is constructed and deposited in the metadata region of the physical device. This information is then mapped on subsequent scans of the device. Logical volumes are not suitable for all partitions or volumes (for example, /boot) as the information is not available in the boot-loader phase.

ⓘ Note: A combination of LVM and PowerPath pseudo-names is currently used to provide persistent binding of boot devices in a multipath environment for 2.6 kernel-based environments.

Mitigating the effects of storage array migration for Linux hosts

Generally, the Linux device tree is not static but built each time upon system reboot in current releases of Linux kernel 2.6-based systems, such as RHEL 5/6/7 and SLES 10/11. All devices present on the system, such as a Fiber Channel HBA, should generate a kernel hotplug event which in turn loads the appropriate driver. Device information from the kernel is exported to sysfs under the `/sys` directory. A user space program, udev (8), will notice and create the appropriate device node devices.

Linux file system can be mounted by different methods. Possibilities for mount() include:

- By a block device name (/dev/sda1, /dev/sdb, /dev/mapper/mpath0, etc)
- By label (LABEL=MYLABEL)
- By id (Use of the scsi-id of a given LUN)

LUNs that contain file systems using mount by label should not be adversely affected by having migrated to a new storage array. The file system label is copied to the new target LUN and mount can identify the corresponding device without user intervention.

Basic block device names, such as /dev/sda1 or /dev/sdb, are created during boot or dynamically upon a LUN rescan event by the kernel. These names are assigned in the order by which the devices are scanned. Such block device names are not considered persistent names.

The /dev/sd block device names are created by the system and cannot be renamed. Fabric changes, such as the addition or deletion of LUNs from a storage target, would likely change the block device names upon a subsequent reboot.

If a systems administrator is using device nodes names such as /dev/sda for mounting file systems or accessing data storage on the array, the devices may not mount or be accessible by the previously used device name.

In a complicated SAN environment, where fabric changes such as the addition or removal of LUNs may occur, it is not recommended to use these non-persistent device names for accessing data.

It is recommended accessing data on devices using persistent names such as /dev/mapper/ mpath0 (for native MPIO devices in RHEL 5/6/7), scsi-id (under /dev/disk/by-id in SLES 10/11), or by-label as previously described.

The Udev(8) tool in current releases of Linux provides a mechanism for creating and maintaining device filenames. Udev (8) is a rules-based approach to device naming. The main configuration file is `/etc/udev/udev.conf`. This configuration file contains specifics for udev_root, permissions, udev_rules, and logging. The default location for udev (8) rules is located in `/etc/udev/rules.d`. Read the distribution-specific information about udev(8) as there are slight variations between SLES and RHEL.

PowerPath has the capability to rename pseudo devices. This approach can also be used to rename devices that are enumerated differently once a host has migrated to a new storage array.

In conjunction with PowerPath is the PowerPath Migration Enabler (PPME), another useful tool to enable migration from one array to another while maintaining data availability and accessibility. Currently, PPME is available for use with Open Replicator (OR) along with PowerPath 5.x for Linux. For additional information regarding implementation and usage, see the PowerPath Migration Enabler (PPME) and Open Replicator (OR) documentation available on Dell EMC Online Support.

Useful utilities

The following tables provides a list of useful system utilities on Linux. Consult the respective man pages for detailed information and usage. Some of the following commands requires the installation of optional packages. Consult your Linux distributor for the appropriate packages.

Table 6 Useful system utilities on Linux

Command name	Purpose [From the respective 'man' pages]
Create partitions, file systems, mount file system, and monitor IO status	
fdisk	Command used to create and manipulate partition tables.
parted	a partition manipulation program.
mkfs	Command used to create a Linux file system on a device partition.
fsck	Command used to check and repair a Linux files ystem.
mount	Command used to attach the file system on a device to the file tree.
umount	Command used to detach a file system.
iostat	The iostat command is used for monitoring system input/output device loading by observing the time the devices are active in relation to their average transfer rates.
LVM command	
lvm	lvm provides the command-line tools for LVM2.
pvcreate	Initialize a disk or partition for use by LVM.
pvdisplay	Display attributes of a physical volume.
vgcreate	Create a volume group.
vgdisplay	Display attributes of volume groups.
vgextend	Add physical volumes to a volume group.
vgreduce	Reduce a volume group.
lvcreate	Create a Logical Volume in an existing volume group.
lvdisplay	Display attributes of a Logical Volume.
lvextend	Extend the size of a Logical Volume.
lvreduce	Reduce the size of a Logical Volume.
Multipath command	
multipath	Multipath is used to detect multiple paths to devices for fail-over or performance reasons and coalesces them.
kpartx	Create device maps from partition tables.
dmsetup	dmsetup manages logical devices that use the device-mapper driver.
devmap_name	devmap_name queries the device-mapper for the name for the device specified by major and minor number.
scsi_id	Retrieve and generate a unique SCSI identifier.

Table 6 Useful system utilities on Linux (continued)

Command name	Purpose [From the respective 'man' pages]
Driver mode utility	
modprobe	Utility used to load or remove a set of modules that can be either a single module or a stack of dependent modules.
lsmod	Utility used to list the currently loaded modules.
insmod	Utility used to dynamically load a single module into a running kernel.
rmmod	Utility used to unload modules from the running kernel if they are not in use.
udev utility	
udev	udev creates or removes device node files usually located in the /dev directory. It provides a dynamic device directory containing only the files for actually present devices.
udevinfo	Query device information from the udev database.
udevmonitor	Print the kernel and udev event sequence to the console.
iSCSI utiltiy command	
iscsiadm	Open-iscsi administration utility.
iscsi-ls	List iscsi device utility.
Other utility	
lspci	Utility used to display information about all of the PCI buses in the system and all of the devices connected to those buses.
lsscsi	Utility used to display information about all of the SCSI devices in the system.
hotplug	Hotplug is a program which is used by the kernel to notify user mode software when some significant (usually hardware related) events take place.

Disk partition adjustment for PowerMAX series, VMAX series, VNX series, VNXe series, Unity series, or XtremIO

This section provides the information alignment and how aligning the data with the physical layout of Dell EMC storage may benefit overall system performance. For operating system-specific information about methods of how alignment may be performed on different file systems and volume managers, contact the operating system vendor.

To maximize disk performance, any I/O to a PowerMAX series, VMAX series, VNX series, VNXe series, Unity series, or XtremIO system needs to be structured to prevent any single I/O operation straddling (crossing) any significant boundaries in the Dell EMC storage. If an I/O does straddle a boundary, this can consume extra resources or cause additional work in the storage array leading to performance loss. There are significant boundaries for the VMAX/VMAX2 that is discussed briefly in this section. For PowerMAX, VMAX All Flash and VMAX3, the track size has increased from 128 blocks (64K) to 256 blocks (128K). So all the boundaries will be doubled and no metastripe boundaries.

- Cache Slot Boundaries (one Track [128 Blocks] - 64 KB)

- RAID 5 Boundaries (four Tracks [512 Blocks] - 256 KB)

- Metastripe Boundaries (two Cylinders [7680 Blocks] - 3840 KB)

Try to minimize the possibility of any single I/O causing a write to both sides of any of the above boundaries.

Windows disks over 7.8 GB are usually deemed to have a geometry of 63 sectors/track and 255 heads/cylinder. Note that sector numbers always start with one (not zero).This means that the first sector on a track is sector one, not sector zero. Additionally, the next track begins with sector one again. However, it is convenient to think of a disk as a sequence of blocks starting from address zero and incrementing until the end of the disk. Because of this, it is important to think in terms of blocks rather than sectors.

By default, partitions created on disks are normally aligned on a cylinder (as defined by Windows) boundary, with one exception. The first partition after the MBR (Master Boot Record) is actually track-aligned, presumably since it was determined that it was too wasteful to just have one block (the MBR) in an empty cylinder. This is a legacy issue.

fdisk allows the creation of a primary partition at any desired block address rather than the default 63 blocks. This means that a partition can be created to minimize the boundary crossings mentioned earlier.

Block 0 on the disk contains the MBR, which defines the disk layout. Since partitions are created on cylinder boundaries, the first partition cannot be created on top of the MBR; therefore, the partition is created at the next track boundary. This is at block address 63. Remember that you start counting blocks at zero, not one, so the first partition starts at the 64th block (and stating the obvious, there are 63 blocks before this first partition).

To align partitions on XtremIO volumes presented to Linux hosts, create a partition using the fdisk command to ensure that the file system is aligned and that the starting sector number is a multiple of 16 (16 sectors, at 512 bytes each, is 8KB).

However, VMAX storage defines tracks differently. On a VMAX series array, a track is considered to be 128 blocks and a VMAX cache memory slot is based on this track size and offset. On VNX series, the unit of allocation is an element, which is (by default) 128 KB.

Track boundaries

If you use the Windows default partition location(63), an I/O of 4 KB (eight blocks) starting at the beginning of the partition will write one block to the last block of the first VMAX track and seven blocks to the start of the second VMAX track. This means the I/O has straddled the first and second VMAX tracks. This requires the array to reserve two cache slots for the data and also requires two flush I/O operations to the VMAX disk, which impacts performance.

For I/O to this partition:

- Any I/O of 32 KB or larger will always cause a boundary crossing.

- Any random I/O of 16 KB will cause a boundary crossing 50 percent of the time.

- Any random I/O of 8 KB will cause a boundary crossing 25 percent of the time.

- Any random I/O of 4 KB will cause a boundary crossing 12.5 percent of the time.

Figure 3 Misaligned partition (not to scale)

| M B R | Reserved (62 blocks) | 4 KB I/O | Disk Partition 1 |

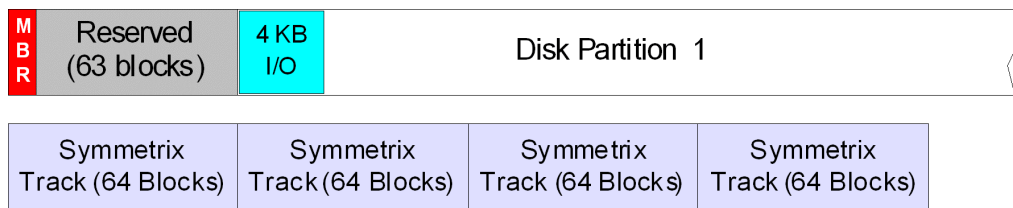| Symmetrix Track (64 Blocks) | Symmetrix Track (64 Blocks) | Symmetrix Track (64 Blocks) | Symmetrix Track (64 Blocks) |
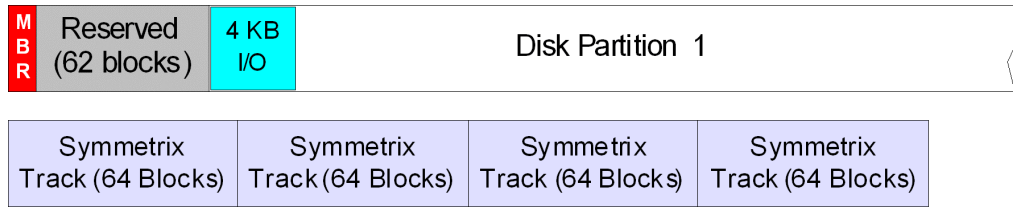
As the previous figure shows, by default the first partition starts on block address 63, whereas to be aligned with a VMAX track, it should start at block address 64. A 4 KB I/O at the start of the disk partition will cause two cache memory slots to be reserved (one for each track).

If the partition started at block 64 (zero based), then no I/O (of 32 KB or less) would cause any boundary crossings.

Figure 4 Aligned partition (not to scale)

| M B R | Reserved (63 blocks) | 4 KB I/O | Disk Partition 1 |

| Symmetrix Track (64 Blocks) | Symmetrix Track (64 Blocks) | Symmetrix Track (64 Blocks) | Symmetrix Track (64 Blocks) |

As the previous shows, starting the first partition on block address 64 will align the I/Os and will not cause boundary crossings.
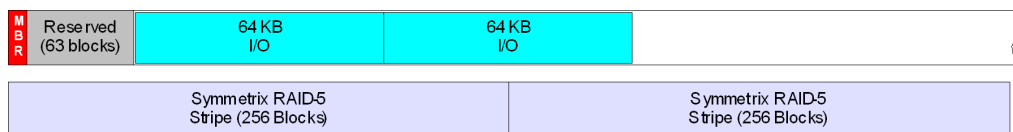
RAID 5 boundaries

In RAID 5 volumes, there are two differences to consider:

- In Linux kernels prior to 2.6.18, the maximum I/O size that Linux will issue is 64 KB. Larger I/Os are broken up into 64 KB chunks. In Linux kernels 2.6.18 and higher, this is handled by the /sys/block/<devicesettingfor>/queue/max_sectors_kb parameter (which is less than the max_hw_sectors_kb the driver/ hardware maximum). The current default value is 512 (KB).

- The RAID 5 stripe size is four VMAX tracks (256 blocks).

If the partition is aligned at 64 blocks, we will still have a potential performance problem. Assume that an I/O write of 64 KB (132 blocks) is issued to the start of the partition. This will not be a problem. Two cache memory slots are required but they are the minimum required for this I/O size. If another I/O (sequential write) of 64 KB is issued, then there is a problem. This second I/O straddles two RAID 5 stripe elements and requires the two stripes to be updated. This requires twice the resources at the back end of the VMAX array compared with an I/O that does not cross a stripe boundary.

Figure 5 Misaligned RAID 5 stripe (not to scale)

| M B R | Reserved (63 blocks) | 64 KB I/O | 64 KB I/O | |

| Symmetrix RAID-5 Stripe (256 Blocks) | Symmetrix RAID-5 Stripe (256 Blocks) |

As the previous figure shows, starting the first partition on block address 64 will cause stripe boundary crossings for 64 KB I/Os. To correct this, the partition needs to be aligned to 128 blocks.

Figure 6 Aligned RAID 5 stripe (not to scale)
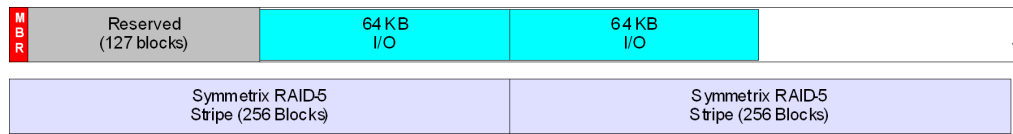
As previous figure shows, starting the first partition on block address 64 will not cause stripe boundary crossings for 64 KB I/Os.

Metastripe boundaries

Metastripes used in Symmetrix, VMAX and VMAX2 are two VMAX cylinders (30 VMAX tracks) in size. Metastripe boundaries have the same back-end performance hit as RAID 5 stripe but have one further complication when Dell EMC SRDF is used. If an I/O crosses a metastripe boundary, it is broken up into two I/O operations for RDF purposes(one for either side of the boundary). In RDF Journal Mode 0, this means that the acknowledgment for the first I/O must be received before the second I/O can be sent. In a similar fashion, if the data is being fetched from the R2 side while performing a RDF restore, the I/O will be broken up into two RDF I/O reads.

Given that the Linux maximum I/O transfer size is 64 KB (128 blocks) and the Metastripe size is 30 VMAX cylinders (1920 blocks), we can see that an alignment of 128 will work since 128 is a multiple of 1920: (1920 / 128 = 15).

VNX series, VNXe series or Unity series

On a VNX system, the default element size is 128 blocks, but it can be from four to 256 blocks.

On a Unity series or VNXe system, there is no element size or LUN offset for user modification. Perform the alignment using a host-based method which align with a 1MB offset.

Determining the correct offset to partition

Before you begin

You must first determine where to start the partition:

- For a VMAX array, the start address is at block 128. For PowerMAX, VMAX All Flash, VMAX3, the start address is at block 256.

- For a VNX system, you must determine the element size. To do this, you can use Dell EMC Unisphere/Navisphere Array Manager.

- For an XtremIO array, the starting sector number is a multiple of 16 (16 sectors, at 512 bytes each, is 8 KB).

To determine the element size for VNX system, perform the following steps:

Procedure

1. Start Unisphere/Navisphere.

2. Go to the appropriate storage group and LUN to be used.

   For example, if want to create an aligned partition on LUN 2 on the lab-w2k host, the LUN number you see in this display is not the LUN number that the host sees. The LUN number in the display is the array LUN number. To get the host LUN number, you must look at the Storage tab in the host properties.

3. Right-click on the LUN, and select **Properties**. **Properties** dialog box is displayed.
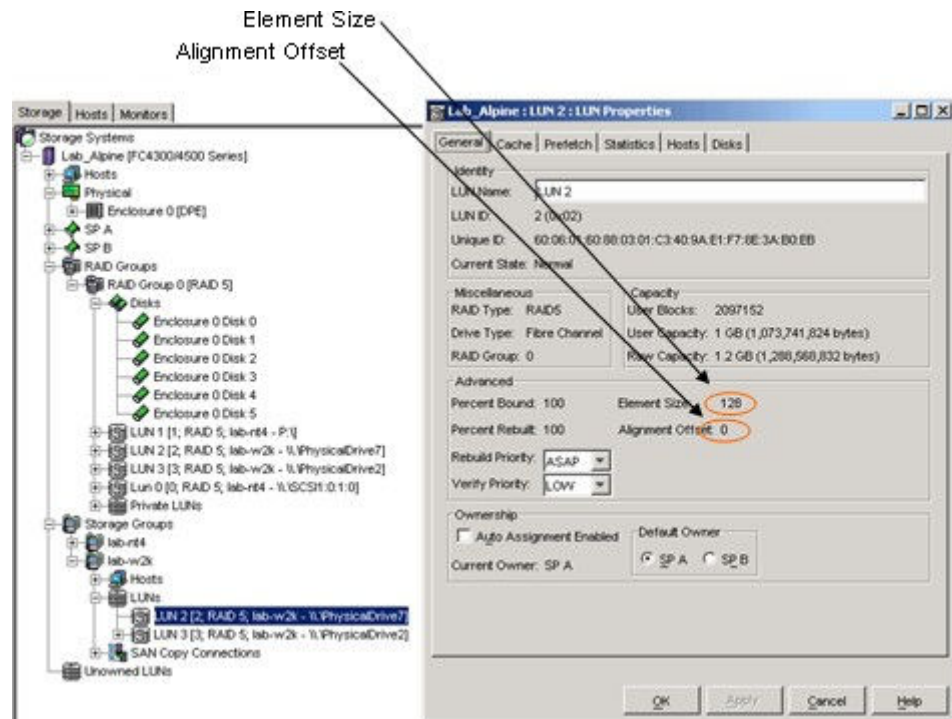
4. Ensure that the Alignment Offset value is 0. If not, the VNX series LUN has been deliberately misaligned, possibly to correct the partition misalignment that you intend to correct. If so, the Alignment Offset value would most likely be 63. If a value other than 0 or 63 appears in the field, then further investigation is required to determine the reason for having a nonstandard value.

Aligning the partition

In Linux, align the partition table before data is written to the LUN as the partition map will be rewritten and all data on the LUN destroyed.

To avoid the performance problems created by misalignment, it is necessary to create and align partitions on Linux using the `fdisk` or `parted` commands.

The `fdisk` will not create partitions larger than 2 TB. To solve this problem, use the GNU `parted` command with GPT. It supports Intel EFI/GPT partition tables. The GUID Partition Table (GPT) is a standard for the layout of the partition table on a physical hard disk. It is a part of the Extensible Firmware Interface (EFI) standard that is proposed by Intel as a replacement for the outdated personal computer BIOS, one of the few remaining relics of the original IBM personal computer.

EFI uses GPT where BIOS uses a Master Boot Record (MBR). EFI GUID Partition support works on both 32 bit and 64 bit platforms.

(i) Note: You must include GPT support in kernel to use GPT. If you do not include GPT support in the Linux kernel, after rebooting the server the file system will no longer be mountable or the GPT table will get corrupted.

By default Red Hat Enterprise Linux/CentOS comes with GPT kernel support. However, if you are using Debian or Ubuntu Linux, you need to recompile the kernel.

Proper alignment examples:

The following show three examples of creating a partition and aligning the partition on Linux. The partition is aligned to start at a 1 MB (= 1,048,576 bytes) mark by default, which covers all common SSD page and block size scenarios, also covers PowerMAX series, VMAX series, Unity series, VNX series, VNXe series and XtremIO, as it is divisible by all commonly used sizes - 1 MB, 512 KB, 128 KB, 4 KB, and 512 bytes.

Example 1: Using older version of fdisk

```
# fdisk -S 32 -H 64 /dev/xxx
```

```
[root@lin105100 Hellya]# fdisk -S 32 -H 64 /dev/emcpowerh
Welcome to fdisk (util-linux 2.23.2).

Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.


Command (m for help): p

Disk /dev/emcpowerh: 21.5 GB, 21474836480 bytes, 41943040 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x00000000

          Device Boot      Start          End      Blocks   Id  System

Command (m for help): n
Partition type:
   p   primary (0 primary, 0 extended, 4 free)
   e   extended
Select (default p):
Using default response p
Partition number (1-4, default 1):
First sector (2048-41943039, default 2048):
Using default value 2048
Last sector, +sectors or +size{K,M,G} (2048-41943039, default 41943039)
:
Using default value 41943039
Partition 1 of type Linux and of size 20 GiB is set
```

```
Command (m for help): p

Disk /dev/emcpowerh: 21.5 GB, 21474836480 bytes, 41943040 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x00000000

          Device Boot      Start          End      Blocks   Id  System
/dev/emcpowerh1               2048     41943039    20970496   83  Linux
```

Example 2:Using fdisk version 2.17.1 or later

```
# fdisk -c -u /dev/xxx
```

```
[root@lin105100 ~]# fdisk -c -u /dev/emcpowerh
Welcome to fdisk (util-linux 2.23.2).

Changes will remain in memory only, until you decide to write them.
Be careful before using the write command.

Device does not contain a recognized partition table
Building a new DOS disklabel with disk identifier 0x7527907c.

Command (m for help): p

Disk /dev/emcpowerh: 21.5 GB, 21474836480 bytes, 41943040 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x7527907c

         Device Boot      Start         End      Blocks   Id  System

Command (m for help): n
Partition type:
   p   primary (0 primary, 0 extended, 4 free)
   e   extended
Select (default p):
Using default response p
Partition number (1-4, default 1):
First sector (2048-41943039, default 2048):
Using default value 2048
Last sector, +sectors or +size{K,M,G} (2048-41943039, default 41943039):
Using default value 41943039
Partition 1 of type Linux and of size 20 GiB is set
```

```
Command (m for help): p

Disk /dev/emcpowerh: 21.5 GB, 21474836480 bytes, 41943040 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x7527907c

         Device Boot      Start         End      Blocks   Id  System
/dev/emcpowerh1             2048      41943039    20970496   83  Linux
```

Example 3: Larger than 2 TB LUN

```
# parted /dev/xxx
(parted) mklabel gpt
(parted) mkpart primary 0% 100% (parted) print
(parted) quit
```

```
[root@e2e-4-10042 ~]# parted /dev/emcpowercz
GNU Parted 3.1
Using /dev/emcpowercz
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) mklabel gpt
(parted) mkpart primary 0% 100%
(parted) print
Model: Unknown (unknown)
Disk /dev/emcpowercz: 15.0GB
Sector size (logical/physical): 4096B/4096B
Partition Table: gpt
Disk Flags:

Number  Start   End     Size    File system  Name      Flags
 1      1049kB  15.0GB  15.0GB                primary

(parted) quit
Information: You may need to update /etc/fstab.

[root@e2e-4-10042 ~]#
```

Operating systems

Dell EMC has a long history of supporting a wide variety of operating systems for our customers. For Dell EMC, Linux is another operating system that is moving more and more into the enterprise, so we are extending our offerings - platforms, software, and services - to fully support Linux environments for enterprise deployment. Unlike the other operating systems, no one vendor owns Linux. There are numerous Linux distributions available. Dell EMC has chosen to support an identified set of enterprise-focused Linux versions and then, over time, ensure all relevant Dell EMC products are available on those distributions

Dell EMC currently supports:

- Red Hat Enterprise Linux (RHEL)

- SuSE Linux Enterprise Server (SLES)

- Asianux (A combination of RedFlag, Miracle Linux, and Haansoft- the leading Linux versions in China, Japan, and Korea respectively)

- Oracle Enterprise Linux

- Xen

Xen is an open-source virtualization technology from XenSource. Dell EMC currently supports Xen server implementations as provided by SuSE Linux and RedHat Linux. Dell EMC does not support Xen implementations from XenSource directly.

See the Dell EMC Simple Support Matrix Linux operating system footnotes in the base connectivity section for supported configurations.

For more information about Xen, see the SuSE and Redhat websites.

Host software

Dell EMC provides a wide range of products at the host end to manage a host in a SAN environment. This section provides details and documentation to reference regarding:

- Dell EMC Solutions Enabler for Linux

- Navisphere CLI

- Unisphere CLI
- Dell EMC replication software

Dell EMC Solutions Enabler for Linux

The Dell EMC Solutions Enabler kit is the software that provides the host with the Symmetrix Command Line Interface (SYMCLI), including the SYMAPI and CLARAPI shared libraries. SYMCLI is a comprehensive command set for managing the storage environment. SYMCLI can be used in many platforms as defined in the Dell EMC Simple Support Matrix.

Documentation

Dell EMC Solutions Enabler documentation is available at Dell EMC Online Support and can be found using the words *Linux* and *Solutions Enabler* in the title search.

Navisphere CLI

The Navisphere CLI/agent is a host-based package that consists of the following two components:

- Navisphere Host Agent
  This is server-based software that communicates with Navisphere client applications, such as the Navisphere Command Line Interface (CLI) and Manager, to manage storage systems. A Host Agent automatically registers hosts and host bus adapters (HBAs) and also provides drive mapping information for the UI and CLI. The Host Agent has no direct user interface.

- NaviCLI
  This is a client application that allows simple operations on a VNX storage system. The CLI issues commands to a Host or SP Agent, requests storage-system status, and displays the resulting output as a tool for problem determination. The Secure CLI is a client application that allows simple operations on a Dell EMC VNX storage system. The Secure CLI issues commands to an SP Agent, requests storage-system status, and displays the resulting output as a tool for problem determination. It is implemented using the Unisphere/Navisphere 6.x security model, which includes role-based management, auditing of all user change requests, management data protected via SSL, and centralized user account management.

Example

Format The naviseccli command is used as follows:

```
naviseccli -help
or
naviseccli

[-address IPAddress | NetworkName | -h  IPAddress | NetworkName]
[-AddUserSecurity]
[-f filename]
[-m]
[-nopoll | -np]
[-parse | -p]
[-password password]
[-port port]
[-q]
[-RemoveUserSecurity]
[-scope 0 | 1]
[-timeout | -t timeout]
[-user username]
[-v]
[-xml]
CMD [optional_command_switches]
```

Documentation

Unisphere and Navisphere documentation is available at Dell EMC Online Support and can be found using the words *Linux*, *Unisphere* or *Navisphere* in the title search.

Unisphere CLI

You can use Unisphere CLI to run commands on a system through a prompt from a Microsoft Windows or UNIX/Linux host. Use Unisphere for managing a system. The Unisphere CLI is intended for advanced users who want to use commands in scripts for automating routine tasks.

Use the Unisphere CLI to manage a Unity or VNXe system. Tasks include:

- Configuring and monitoring the system
- Managing users
- Provisioning storage
- Protecting data
- Controlling host access to storage

For more information, see the Unisphere CLI documentation on Dell EMC Online Support.

Dell EMC replication software

This section discusses Dell EMC RecoverPoint replication software.

RecoverPoint for Linux

The RecoverPoint system enables the reliable replication of data over any distance (within the same site or to another site halfway around the globe). It supports replication of data that your applications are writing over Fibre Channel to local SAN-attached storage. It uses your existing Fibre Channel infrastructure to integrate seamlessly with your existing host applications and data storage subsystems. For long-distance replication, it uses existing IP to send the replicated data

over a WAN. It provides successful failover of your operations to a secondary site in the event of a disaster at the primary site.

The current version of RecoverPoint for Linux supports RHEL, SLES, and VMware. See the Dell EMC Simple Support Matrix for the most current support information for the operating system version and RecoverPoint.

Documentation

RecoverPoint document is available at Dell EMC Online Support and can be found using the word *RecoverPoint* in the title search.

Server vendor

Dell EMC supports 64-bit Intel, AMD, and PowerPC architecture-based servers from a range of server vendors. Stand-alone rack, tower and blade servers are supported. For more detailed information, see the operating system vendor documentation.

(i) Note: Dell EMC provides support for the IBM PowerPC architecture. The IBM P6 and P7 series standard LPAR and VIO clients are supported for PowerMax series, VMAX series, XtremIO, VNX series, VNXe series and Unity series. Linux native DM-MPIO is supported on Linux LPARs. Support for booting Linux LPARS from the SAN is available. See the Dell EMC Simple Support Matrix for specific configurations.

The Dell EMC Simple Support Matrix provides a detailed listing of supported servers, HBA models, drivers, and operating system revisions. For further information regarding Linux support on the server models, see the vendor's website.

Host bus adapters

Dell EMC storage arrays support QLogic, Emulex, and Brocade Fibre Channel host bus adapters. See the Linux *Base Connectivity* section of the Dell EMC Simple Support Matrix for supported HBA models and see the appropriate installation guide, available at Dell EMC Online Support, for configuring the host adapter and driver for the system.

Dell EMC iSCSI storage arrays support both the QLogic iSCSI hardware initiator and the generic NIC iSCSI software initiator. See the Linux *iSCSI Connectivity* section of the Dell EMC Simple Support Matrix for supported configurations and required driver revision levels. See iSCSI Connectivity for more details.

Dell EMC published *QLogic iSCSI guide* is available on the QLogic website.

Dell EMC-published *HBA driver configuration guides* are available in the Dell EMC-approved sections of the Broadcom/QLogic/Brocade websites.

Converged network adapters

Dell EMC supports Emulex, QLogic, and Brocade Fibre Channel over Ethernet (FCoE) Converged Network Adapters (CNAs). FCoE adapters represent a method to converge both Fibre Channel and Ethernet traffic over a single physical link to a switch infrastructure that manages both storage (SAN) and network (IP) connectivity within a single unit.

The benefits of FCoE technology become apparent in large data centers:

* Where dense, rack-mounted and blade server chassis exist.

* Where physical cable topology simplification is a priority.

- In virtualization environments, where several physical storage and network links are commonly required.

The installation of an FCoE CNA provides the host with an Intel-based 10 Gb Ethernet interface (using the existing in-box drivers), and an Emulex, QLogic, or Brocade Fibre Channel adapter interface. Upon installation of the proper driver for the FCoE CNA, the Fibre Channel interface functions identically to that of a standard Emulex, QLogic, or Brocade Fibre Channel HBA. The FCoE CNA simply encapsulates Fibre Channel traffic within Ethernet frames. As such, FC-based content within this guide also applies directly to Emulex, QLogic, or Brocade FCoE CNAs.

To install one or more Dell EMC-qualified Emulex, QLogic, or Brocade CNAs into a Linux host and configure the host for connection to Dell EMC storage arrays over FCoE, follow the procedures available in the Dell EMC OEM section of the Broadcom, QLogic, or Brocade websites, or at Dell EMC Online Support.

FCoE technology is being phased-out by some Operating System vendors. For further information regarding FCoE support, see the Operating System vendor's website.

Dell EMC storage

This section provides basic information about following Dell EMC storage:

- PowerMAX series or VMAX series
- Unity series
- VNX series
- VPLEX
- XtremIO
- VXFlex
- XtremCache
- PowerStore

PowerMAX series or VMAX series

This section contains information about PowerMAX and VMAX applications, offerings, array-specific settings, and documentation.

Applications

Information about PowerMAX and VMAX applications can be found at the Dell EMC VMAX All Flash page Storage on DellEMC.com.

Offerings

The following table lists the PowerMAX and VMAX arrays supported on the Linux platform, along with the minimum Dell EMC Enginuity microcode revisions. The particular code levels supported within an Enginuity microcode family are listed in the Path Management Software table in the Dell EMC Simple Support Matrix .

Table 7 Supported PowerMAX and VMAX arrays

Storage arrays	Array code requirements
PowerMAX 8000	PowerMax OS 5978
PowerMAX 2000	PowerMax OS 5978
VMAX 950F/FX	HYPERMAX OS 5977

Table 7 Supported PowerMAX and VMAX arrays (continued)

Storage arrays	Array code requirements
VMAX 850F/FX	HYPERMAX OS 5977
VMAX 450F/FX	HYPERMAX OS 5977
VMAX 400K	HYPERMAX OS 5977
VMAX 250F/FX	HYPERMAX OS 5977
VMAX 200K	HYPERMAX OS 5977
VMAX 100K	HYPERMAX OS 5977
VMAX 40K	Enginuity 5876
VMAX 20K	Enginuity 5876
VMAX [Fibre Channel and iSCSI]	Enginuity 5874/5875/5876
VMAX 10K (Systems with SN xxx987xxxx)	Enginuity 5876
VMAX 10K (Systems with SN xxx959xxxx)	Enginuity 5876
VMAXe [Fibre Channel and iSCSI]	Enginuity 5875 /5876
Symmetrix DMX-4 [Fibre Channel and iSCSI]	Enginuity 5772 /5773
Symmetrix DMX-3 [Fibre Channel and iSCSI]	Enginuity 5771/5772

See the Base Connectivity and the iSCSI Connectivity tables in the Dell EMC Simple Support Matrix for supported HBAs, operating system revisions, and servers.

PowerMAX and VMAX series array

Specific settings-When attaching a Linux host to a PowerMAX or VMAX system, use the Linux default FA settings referred to in the Director Bit Settings Simple Support Matrices at Dell EMC E-Lab Interoperability Navigator.

Documentation

PowerMAX and VMAX documentation is available at Dell EMC Online Support.

Unity series

This section contains information about Unity series applications, offerings, and documentation.

- See information about Unity applications on the Unity All Flash Storage page on DellEMC.com and on Dell EMC Online Support.
- The Unity series includes All-Flash storage arrays (Unity 300F, 380F, 400F, 480F, 500F, 600F, 680F, 880F) or Hybrid Flash storage arrays (Unity 300, 380, 400, 480, 500, 600, 680, 880)

VNX series

This section contains information about VNX series, offerings, array-specific settings, and documentation.

Applications

Information about VNX application can be found on the EMC VNX Hybrid Flash Storage Family page on DellEMC.com.

VNX series

specific settings-When attaching a Linux host to a VNX system, follow the appropriate guidelines, depending upon the environment.

- A Linux host using PowerPath or the Linux native DM-MPIO supports both the default failover mode 1 and the optional failover mode 4 (ALUA). Consult the PowerPath release notes at Dell EMC Online Support for supported versions of PowerPath and failover mode 4. For the Linux native DM-MPIO, see ALUA supported Linux operating systems.

- If the Linux host is using Veritas VxVM/DMP, the failover mode must be set to *2* for Veritas Storage Foundation v4.0 and v4.1 and either *1* or *4* for Veritas Storage Foundation v5.0.

- VNX series supports ALUA

Documentation

VNX documentation can be found at Dell EMC Online Support.

VPLEX

Dell EMC VPLEX is a platform that delivers Local and Distributed Federation.

Documentation

See the following documents for configuration and administration operations:

- *EMC VPLEX with GeoSynchrony 5.0 Product Guide*

- *EMC VPLEX with GeoSynchrony 5.0 CLI Guide*

- *EMC VPLEX with GeoSynchrony 5.0 Configuration Guide*

- *EMC VPLEX Hardware Installation Guide*

- *EMC VPLEX Release Notes*

- *Implementation and Planning Best Practices for EMC VPLEX Technical Notes*

- *VPLEX online help*, available on the Management Console GUI

- *VPLEX Procedure Generator*, available at Dell EMC Online Support

- *Dell EMC Simple Support Matrix, Dell EMC VPLEX, and GeoSynchrony*, available on Dell EMC E-Lab Navigator

For the most up-to-date support information, always see the *Dell EMC Simple Support Matrix.*, available on Dell EMC E-Lab Navigator.

XtremIO

Dell EMC XtremIO is an all-flash storage array that has been designed from the ground-up to unlock flash's full performance potential and deliver array-based capabilities that leverage the unique characteristics of SSDs, based on flash media.

XtremIO uses industry standard components and proprietary intelligent software to deliver unparalleled levels of performance. Achievable performance ranges from hundreds of thousands to millions of IOPS, and consistent low latency of under one millisecond.

The system is also designed to provide minimal planning, with a user-friendly interface that makes provisioning and managing the array easy.

XtremIO leverages flash to deliver value across the following main dimensions:

- Performance - Regardless of how busy the system is, and regardless of storage capacity utilization, latency, and through- put remain consistently predictable and constant. Latency within the array for an I/O request is typically far less than one millisecond.

- Scalability - The XtremIO storage system is based on a scale-out architecture. The system begins with a single building block, called an X-Brick. When additional performance and capacity are required, the system scales out by adding X-Brick blocks. Performance scales

linearly, ensuring that two X-Brick blocks supply twice the IOPS and four X-Bricks supply four times the IOPS of the single X-Brick configuration. Latency remains consistently low as the system scales out.

- Efficiency - The core engine implements content-based Inline Data Reduction. The XtremIO storage array automatically reduces (deduplicates) data on the fly, as it enters the system. This reduces the amount of data written to flash, improving longevity of the media and driving down cost. XtremIO arrays allocate capacity to volumes on-demand in granular 4KB chunks. Volumes are always thin-provisioned without any loss of performance, over-provisioning of capacity, or fragmentation.

- Data Protection - XtremIO leverages a proprietary flash-optimized data protection algorithm (XtremIO Data Protection or XDP) which provides RAID-6 protection for data, while enabling performance that is superior to any existing RAID algorithms. Optimizations in XDP also result in fewer writes to flash media for data protection purposes.

- Functionality - XtremIO supports high performance and space-efficient snapshots, Inline Data Reduction, and thin provisioning, as well as support for Fibre Channel and iSCSI protocols.

Multipathing support

XtremIO storage is supported with PowerPath, Linux native DM-MPIO, Veritas DMP. Consult the versions of Dell EMC PowerPath and Veritas DMP supported in the Dell EMC Simple Support Matrix.

VxFlex

This section contains information about Dell EMC VxFlex applications, offerings, installation/ management /configuration processes, support for Linux, and documentation.

Applications

Find information about VxFlex applications on the Dell EMC ScaleIO.

Implementation

Implementing a VxFlex system is, in general, a two-step process: first build the physical storage layer, then configure the virtual SAN layer on top of it.

Physical layer

The physical layer consists of the hardware (servers with storage devices and the network between them) and the VxFlex software installed on them. To implement the physical layer, perform the following steps:

Procedure

1. Install the MDM component on the MDM nodes in one of the following configurations:

   - Single node (one master MDM)

   - Three-node redundant cluster (one Master MDM, one Slave MDM, and one Tie-Breaker)

   - Starting with VxFlex 2.0.x - Five-node redundant cluster (one Master MDM, two Slave MDMs, and two Tie-Breakers).

2. Install the SDS component on all nodes that contribute some or all their physical storage:

   - Starting with VxFlex v2.0.x, up to 4 SDSs can be installed on a single host server.

   - Divide the SDS nodes into Protection Domains. Each SDS can be a member of only one Protection Domain.

   - Per Protection Domain, divide the physical storage units into Storage Pools, and optionally, into Fault Sets.

3. Install the SDC component on all nodes on which the application accesses the data exposed by the VxFlex volumes.

   Communication is done over the existing LAN using standard TCP/IP. The MDM and SDS nodes can be assigned up to eight IP addresses, enabling wider bandwidth and better I/O performance and redundancy.

   You can perform physical layer setup using the following methods:

   • VxFlex Installation Manager

   • VxFlex VMware plug-in

   • Manual installation

   After completing this installation, the physical layer is ready, and it exposes a virtual storage layer.

SAN virtualization layer

The MDM cluster manages the entire system. It aggregates the entire storage exposed to it by all the SDSs to generate virtual layer - virtual SAN storage. Volumes can now be defined over the Storage Pools and can be exposed to the applications as a local storage device using the SDCs.

To expose the virtual SAN devices to your servers (the ones on which you installed and configured SDCs), perform the following:

• Define volumes
  Each volume defined over a Storage Pool is evenly distributed over all members using a RAID protection scheme. By having all SDS members of the Storage Pool participate, VxFlex ensures:

  ▪ Highest and most stable and consistent performance possible

  ▪ Rapid recovery and redistribution of data

  ▪ Massive IOPS and throughput

  You can define volumes as thick, where the entire capacity is provisioned for storage, or thin, where only the capacity currently needed is provisioned.

• Map volumes
  Designate which SDCs can access the given volumes. This gives rise to the following:

  ▪ Access control per volume exposed

  ▪ Shared nothing or shared everything volumes

Once an SDC is mapped to a volume, it immediately gets access to the volume and exposes it locally to the applications as a standard block device. These block devices appear as /dev/sciniX, where X is a letter, starting from 'a'.

For example:

```
/dev/scinia
/dev/scinib
```

Support for Linux

See the *Dell EMC Simple Support Matrix - VxFlex Node* and *EMC Simple Support Matrix - VxFlex Software*, located at Dell EMC E-Lab Interoperability Navigator .

Documentation

Documentation can be found at Dell EMC Online Support.

XtremCache

This section contains information on Dell EMC XtremCache/SF applications, offerings, installation/management /configuration processes, supports for Linux, and documentation.

Applications

For more information about XtremCache applications, see Dell EMC VxFlex page.

Implementation

1. Install the caching software on every machine that provides caching services.

2. Install the management utilities on those workstations to be used to manage XtremCache and flash cards.

   - Command Line Interface (CLI)

   - VSI plug-in for VMware

   - Management Center

   - Lite Client

3. Install the license.

4. Configure using CLI for the first time.

Considerations

- To use a partitioned device, create the partitions before enabling the cache device.

- Do not use a system disk as a source disk.

- Enable and start a cache device:
  ```
  vfcmt add -cache_dev <device>
  ```

- Enable and start a source device:
  ```
  vfcmt add -source_dev <device>
  ```

- You can review the configuration by typing the following command:
  ```
  vfcmt display -all
  ```

The XtremCache installation and management files are included on the installation media, or you can download them from Dell EMC Online Support.

Support for Linux

See *Dell EMC Simple Support Matrix - EMC XtremCache*, located at Dell EMC E-Lab Interoperability Navigator .

Documentation

XtremCache documentation are available onDell EMC Online Support.

PowerStore

The PowerStore storage array provides operational simplicity and agility, using a container-based microservices architecture, advanced storage technologies, and integrated machine learning to unlock the power of your data. PowerStore is a versatile platform with a performance-centric design that delivers multidimensional scale, always on data reduction, and support for next generation media. It brings the simplicity of public cloud to on-premises infrastructure, streamlining operations with an integrated machine learning engine and seamless automation, while offering predictive analytics to easily monitor, analyze, and troubleshoot the environment. PowerStore is also highly adaptable, providing the flexibility to host specialized workloads directly on the appliance and modernize infrastructure without disruption, all while offering investment protection through flexible payment solutions and data-in-place upgrades.

It is designed to take mid-range IT to the next level of simplicity, performance and economics using a new I/O stack designed from the ground-up for the next wave of media, a scale-up-and-out architecture, flexible consumption models, and extreme simplicity:

Predictable

Optimizes its use of ultra fast NVMe media, including Flash and high endurance non-volatile media, combined with share-based QOS, and workload analytics to automate the balance of compute, memory, and storage resources to meet the performance demands of each application. It has the ability to detect performance trends and seasonality of workloads, and forecast future demands for each application. It uses this trending and forecasting to balance resources and workloads to ensure that future demand can be met .

Inline Efficiencies

Optimizes storage efficiency by providing inline compression, inline deduplication, pattern reduction. Trident offers software-based compression for software-defined and virtual deployments. It also offers hardware-based compression and hashing for appliance and server deployments that include Intel Quick-Assist Technology. Another feature offered is Log Structure writes and mapped RAID which allows for variable sized compression and reduced parity overhead. to improve storage virtualization overhead

High Availability

Base on a highly available dual node active/active controller architecture with shared storage. It implements a fully active/active data path which allows for transparent fail-over when masking hardware and software failures, FRU replacements, and software upgrades.

Scale Up Scale out

PowerStore appliances can be scaled out as compute and capacity demand increases beyond the capacity of a single appliance. It leverages the compute and storage resources across the cluster to serve SAN targets, NAS shares, and host customer applications. The storage resources and applications (compute resources) are balanced across HA appliances within the cluster to best meet compute, performance, and capacity demands.

Flexible Consumption

Takes unification to the next level by offering the ability to serve SAN, NAS, and Virtual Machines. An appliance can be deployed as a traditional Unified Storage Array - serving SAN, NAS, and Virtual Volumes. An alternate deployment is a Mixed-Mode appliance that is a Unified Appliance that can serve Virtual Machines.

Integrated Data Protection

Includes a rich set of Integrated Data Protection features which include Application consistent, VM consistent, and crash consistent snapshots and clones. Local and remote backup to PowerStore Appliances, Data Domain and Cloud. Synchronous and Asynchronous Disaster Recovery and VASA 3.0 replication

Ultra Simple

PowerStore infrastructure can be managed through a single highly available management application. The management application can be discovered on the local network using any Avahi/Bonjour enabled browser or utility. It integrates with CloudIQ for health monitoring and copy data management. It also provides seamless migrations tools that are used to migrate storage from existing Dell EMC storage products including Unity, SC, PS, and VNX. It also has the ability to seamlessly migrate storage from Non-Dell EMC products. For Serviceability Cyclone leverages EMC Secure Remote Services.

PowerStore storage array consist of two models:

- PowerStore X models - A bare metal storage appliance that serves Block (iSCSI, FC, VVols) and File (NFS and CIFS) protocols to external clients (servers).
- PowerStore T models - A unified storage appliance which allows a customer to deploy applications, in the form of Virtual Machines, that run directly on the Storage Appliance meanwhile it serves Block (iSCSI, FC, and VVols) and File (NFS and CIFS) protocols to external clients (servers)

Documentation

The PowerStore storage documentation is available at Dell EMC Online Support.

CHAPTER 2

Connectivity

This chapter includes the following topics:

Fibre Channel Connectivity

Introduction

Fibre Channel captures some of the benefits of both channels and networks. A Fibre Channel fabric is a switched network, providing a generic, low-level services onto which host channel architectures and network architectures can be mapped. Networking and I/O protocols (such as SCSI commands) are mapped to Fibre Channel constructs and then encapsulated and transported within Fibre Channel frames. This process allows high-speed transfer of multiple protocols over the same physical interface.

The phrase Fibre Channel is often used as an abbreviation of SCSI over Fibre Channel. Fibre Channel is a transport protocol that allows mapping other service-oriented or device-oriented protocols within its transport frames. SCSI over Fibre Channel allows us to overcome the distance, dynamic flexibility, and accessibility limitations associated with traditional direct-attach SCSI.

As with direct-attach SCSI, Fibre Channel provides block level access to the devices that allows the host system to identify the device as a native device. The true power of native device identification is seen in our ability to use all of our current applications (for example: backup software, volume management, and raw disk management) without modification.

Fibre Channel is a technology for transmitting data between computer devices at data rates of up to 8 GBs at this time. Fibre Channel is flexible; devices can be as far as ten kilometers (about six miles) apart if optical fiber is used as the physical medium.

Fibre Channel supports connectivity over fiber optic cabling or copper wiring. Fibre Channel devices using fiber optic cabling use two unidirectional fiber optic cables for each connection. One fiber optic cable is used for transmitting; the other for receiving. Fibre channel over fiber optic cable supports cable distances of up to 10 km.

Configuring the HBAs for a Linux host

This section describes the procedures for installing a Dell EMC-approved Broadcom (Emulex), Marvell (QLogic), and Marvell (Brocade) adapter into a Linux host environment and configuring the host for connection to a Dell EMC storage array over Fibre Channel.

This section contains the following information:

- Prerequisites for first-time installation
- Broadcom (Emulex) Fibre Channel HBA
- Marvell (QLogic) Fibre Channel HBA
- Marvell (Brocade) Fibre Channel HBA

Prerequisites for first-time installation

This section lists the requirements before you install the HBA for the first time.

(i) **Note:** Dell EMC does not support mixing different types of Fibre Channel adapter (including different types from the same vendor) in a server.

- See Dell EMC Simple Support Matrix or contact your Dell EMC representative for the latest information on qualified adapters, drivers, and Linux distributions.
- See the vendor's Fibre Channel Host Adapter (HBA) product documentation to properly install an HBA in your server.

- See the vendor's product documentation to verify and update HBA firmware and boot BIOS to Dell EMC-qualified versions.

Broadcom (Emulex) Fibre Channel HBA

Installing the driver using the Emulex Fibre Channel HBA adapter with the Linux operating system requires adapter driver software. The driver functions at a layer below the Linux SCSI driver to present Fibre Channel devices to the operating system as if they were standard SCSI devices.

See the latest Dell EMC Simple Support Matrix for specific qualified kernel and driver versions and Linux distributions.

Dell EMC supports the Emulex in-kernel and out-of-kernel drivers.

(i) Note: The installation of the in-kernel driver occurs when you install your Linux distribution of choice.

If your installation requires an out-of-kernel driver, download it from the Dell EMC-approved section of the Broadcom website. Follow the links to your adapter for the appropriate operating system and version.

(i) Note: The support that is stated in the Dell EMC Simple Support Matrix supersedes versions that are listed in this document.

Dell EMC-supported Emulex driver versions

The following table lists the Emulex driver versions that are supported with the corresponding operating system updates. These driver versions are included by default in the kernel and do not require any installation.

Table 8 Dell EMC supported Emulex Linux in-kernel drivers

Operating system	In-kernel Driver Version	Supported adapter				
		1/2 Gb	4 Gb	8 Gb	16 Gb	32 Gb
SLES 10 GA	8.1.6	Yes	Yes			
RHEL 5.0 Oracle OL 5.0 SLES 10 SP1 (errata kernels 2.6.16.53-0.8, 2.6.16.53-0.16)	8.1.10.3	Yes	Yes			
RHEL 5.1 Oracle OL 5.1 SLES 10 SP1 (errata kernels 2.6.16.53-0.8, 2.6.16.53-0.16)	8.1.10.9	Yes	Yes			
SLES 10 SP1 (errata kernels equal to or greater than 2.6.16.54-0.2.3)	8.1.10.12-update	Yes	Yes			
RHEL 5.2 SLES 10 SP2 Oracle OL 5.2	8.2.0.22	Yes	Yes	Yes		
RHEL 5.4 SLES 10 SP3	8.2.0.48.2p	Yes	Yes	Yes		
RHEL 5.3	8.2.0.33.3p	Yes	Yes	Yes		

Table 8 Dell EMC supported Emulex Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		1/2 Gb	4 Gb	8 Gb	16 Gb	32 Gb
OL 5.3						
RHEL 5.4 OL 5.4	8.2.0.48.2p 8.2.0.48.3p (errata kernels 2.6.18-164.11.1. 0. 1.el5 and higher)	Yes	Yes	Yes		
RHEL 5.5 OL 5.5	8.2.0.63.3p	Yes	Yes	Yes		
RHEL 5.6	8.2.0.87.1p	Yes	Yes	Yes		
RHEL 5.7	8.2.0.96.2p	Yes	Yes	Yes		
RHEL 5.8	8.2.0.108.4p	Yes	Yes	Yes		
RHEL 5.9 RHEL 5.10 RHEL 5.11	8.2.0.128.3p	Yes	Yes	Yes	Yes	
	8.2.0.128.3p	Yes	Yes	Yes	Yes	
RHEL 6.0	8.3.5.17	Yes	Yes	Yes		
RHEL 6.1	8.3.5.30.1p	Yes	Yes	Yes		
RHEL 6.2	8.3.5.45.4p	Yes	Yes	Yes		
RHEL 6.3	8.3.5.68.5p	Yes	Yes	Yes	Yes	
RHEL 6.4	8.3.5.86.1p	Yes	Yes	Yes	Yes	
RHEL 6.5	8.3.7.21.4p	Yes	Yes	Yes	Yes	
RHEL 6.6	10.2.8020.1	Yes	Yes	Yes	Yes	
RHEL 6.7	10.6.0.20	Yes	Yes	Yes	Yes	
RHEL 6.8	11.0.0.4	Yes	Yes	Yes	Yes	
RHEL 6.9	11.0.0.5 11.0.0.6 (Errata kernel 2.6.32-696.3.1. el6 and later)	Yes	Yes	Yes	Yes	
RHEL 6.10	11.0.1.6	Yes	Yes	Yes	Yes	Yes
RHEL 7.0	8.3.7.34.3p	Yes	Yes	Yes	Yes	
RHEL 7.1	10.2.8021.1	Yes	Yes	Yes	Yes	
RHEL 7.2	10.7.0.1	Yes	Yes	Yes	Yes	
RHEL 7.3	11.1.0.2	Yes	Yes	Yes	Yes	

Table 8 Dell EMC supported Emulex Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		1/2 Gb	4 Gb	8 Gb	16 Gb	32 Gb
RHEL 7.4	11.2.0.6	Yes	Yes	Yes	Yes	Yes
RHEL 7.5	11.4.0.4	Yes	Yes	Yes	Yes	Yes
RHEL 7.6	12.0.0.5	Yes	Yes	Yes	Yes	Yes
RHEL 7.7	12.0.0.10	Yes	Yes	Yes	Yes	Yes
RHEL 7.8	12.0.0.13	Yes	Yes	Yes	Yes	Yes
RHEL 8.0	12.0.0.6			Yes	Yes	Yes
RHEL8.1	12.2.10.3			Yes	Yes	Yes
SLES 10 SP4	8.2.0.92.1p	Yes	Yes	Yes	Yes	
SLES 11 GA	8.2.8.14	Yes	Yes	Yes		
SLES 11 SP1	8.3.5.8.1p 8.3.5.8.2p	Yes	Yes	Yes	Yes	
SLES 11 SP2	8.3.5.48.2p 8.3.5.48.3p	Yes	Yes	Yes	Yes	
SLES 11 SP3	8.3.7.10.6p 8.3.7.10.7p	Yes	Yes	Yes	Yes	
SLES 11 SP4	10.4.8000.0	Yes	Yes	Yes	Yes	Yes
SLES 12	10.2.8040.1	Yes	Yes	Yes	Yes	
SLES 12 SP1	10.5.0.2	Yes	Yes	Yes	Yes	Yes
SLES 12 SP2	11.1.0.1	Yes	Yes	Yes	Yes	Yes
SLES 12 SP3	11.4.0.1	Yes	Yes	Yes	Yes	Yes
SLES 12 SP4	12.0.0.6	Yes	Yes	Yes	Yes	Yes
SLES 12 SP5	12.4.0.0 12.6.0.2	Yes	Yes	Yes	Yes	Yes
SLES 15	12.0.0.1	Yes	Yes	Yes	Yes	Yes
SLES 15 SP1	12.2.0.0	Yes	Yes	Yes	Yes	Yes
Oracle Linux 5. x UEK R1 [2.6.32-100]	8.3.18	Yes	Yes	Yes		
Oracle Linux 6. x UEK R1 [2.6.32-100]	8.3.5.30.1p	Yes	Yes	Yes		
Oracle Linux 5. x UEK R1 U1 [2.6.32-200] Oracle Linux 6. x UEK R1 U1 [2.6.32-200]	8.3.5.44	Yes	Yes	Yes		
Oracle Linux 5. x UEK R1 U2 [2.6.32-300]	8.3.5.45.4p	Yes	Yes	Yes	Yes	

Table 8 Dell EMC supported Emulex Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		1/2 Gb	4 Gb	8 Gb	16 Gb	32 Gb
Oracle Linux 6.x UEK R1 U2 [2.6.32-300] Oracle Linux 5.x UEK R1 U3 [2.6.32-400] Oracle Linux 6.x UEK R1 U3 [2.6.32-400]						
Oracle Linux 5.x UEK R2 [2.6.39-100] Oracle Linux 6.x UEK R2 [2.6.39-100]	8.3.5.58.2p	Yes	Yes	Yes	Yes	
Oracle Linux 5.x UEK R2 U1 [2.6.39-200] Oracle Linux 6.x UEK R2 U1 [2.6.39-200]	8.3.5.68.6p	Yes	Yes	Yes	Yes	
Oracle Linux 5.x UEK R2 U2 [2.6.39-300] Oracle Linux 6.x UEK R2 U2 [2.6.39-300]	8.3.5.82.2p	Yes	Yes	Yes	Yes	
Oracle Linux 5.x UEK R2 U3 [2.6.39-400] Oracle Linux 6.x UEK R2 U3 [2.6.39-400]	8.3.5.86.2p 8.3.7.10.4p	Yes	Yes	Yes	Yes	
Oracle Linux 5.x UEK R2 U4 [2.6.39-400.109] Oracle Linux 6.x UEK R2 U4 [2.6.39-400.109]	8.3.7.10.4p	Yes	Yes	Yes	Yes	
Oracle Linux 5.x UEK R2 U5 [2.6.39-400.209] Oracle Linux 6.x UEK R2 U5 [2.6.39-400.209]	8.3.7.26.3p	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 [3.8.13-16]	8.3.7.26.2p	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U1 [3.8.13-26] Oracle Linux 6.x UEK R3 U2 [3.8.13-35] Oracle Linux 7.x UEK R3 U2 [3.8.13-35] Oracle Linux 6.x UEK R3 U3 [3.8.13-44] Oracle Linux 7.x UEK R3 U3 [3.8.13-44]	8.3.7.34.4p	Yes	Yes	Yes	Yes	

Table 8 Dell EMC supported Emulex Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		1/2 Gb	4 Gb	8 Gb	16 Gb	32 Gb
Oracle Linux 6.x UEK R3 U4 [3.8.13-55] Oracle Linux 7.x UEK R3 U4 [3.8.13-55]	10.2.8061.0	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U5 [3.8.13-68] Oracle Linux 7.x UEK R3 U5 [3.8.13-68]	10.6.61.0	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U6 [3.8.13-98] Oracle Linux 7.x UEK R3 U6 [3.8.13-98]	10.6.61.1	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U7 [3.8.13-118] Oracle Linux 7.x UEK R3 U7 [3.8.13-118]	11.0.0.1	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R4 [4.1.12] Oracle Linux 7.x UEK R4 [4.1.12]	11.0.0.3	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U1 [4.1.12-37] Oracle Linux 7.x UEK R4 U1 [4.1.12-37]	11.0.0.13	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U2 [4.1.12-61] Oracle Linux 7.x UEK R4 U2 [4.1.12-61]	11.1.0.4	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U4 [4.1.12-94] Oracle Linux 7.x UEK R4 U4 [4.1.12-94]	11.2.0.5	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U5 [4.1.12-103] Oracle Linux 7.x UEK R4 U5 [4.1.12-103]	11.4.0.2	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U6 [4.1.12-112] Oracle Linux 7.x UEK R4 U6 [4.1.12-112]	11.4.0.2	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U7 [4.1.12-124]	11.4.0.7	Yes	Yes	Yes	Yes	Yes

Table 8 Dell EMC supported Emulex Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		1/2 Gb	4 Gb	8 Gb	16 Gb	32 Gb
Oracle Linux 7.x UEK R4 U7 [4.1.12-124]						
Oracle Linux 7.x UEK R5 [4.1.35-1818]	12.0.0.2 12.0.0.5	Yes	Yes	Yes	Yes	Yes
Oracle Linux 7.x UEK R5 U1 [4.1.35-1844]	12.0.0.7	Yes	Yes	Yes	Yes	Yes
Oracle Linux 7.x UEK R5 U2 [4.1.35-1902]	12.0.0.10	Yes	Yes	Yes	Yes	Yes
XenServer [x86_64] 6.2	10.0.803.24	Yes	Yes	Yes		
XenServer [x86_64] 6.5	10.2.340.0	Yes	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.0	11.0.0.12	Yes	Yes	Yes	Yes	Yes [a]
XenServer [x86_64] 7.1	11.1.210.1	Yes	Yes	Yes	Yes	Yes[a]
XenServer [x86_64] 7.2	11.1.210.1	Yes	Yes	Yes	Yes	Yes [a]
XenServer [x86_64] 7.3	11.1.210.1	Yes	Yes	Yes	Yes	Yes [a]
XenServer [x86_64] 7.4	11.1.210.1	Yes	Yes	Yes	Yes	Yes [a]

a. Customers who use Broadcom Limited's lpfc driver with XenServer 7.0 are recommended to use Broadcom Limited's certified optics with the card.

Table 9 Dell EMC-supported Emulex Linux out-of-kernel drivers

Operating system	Out-of-kernel Driver Version	Supported adapter				
		1/2 Gb	4 Gb	8 Gb	16 Gb	32 Gb
RHEL 6.5	11.1.38.64					Yes [a]
RHEL 6.6	11.1.38.64					Yes[a]
RHEL 6.7	11.1.38.64					Yes[a]
RHEL 6.9	11.1.38.64					Yes[a]
RHEL 6.10	11.1.38.64					Yes[a]
RHEL 7.0	11.1.38.64					Yes[a]
RHEL 7.0	8.7.7.34.3p		Yes [b]			
RHEL 7.1	11.1.38.64					Yes[a]
RHEL 7.1	8.7.7.34.3p		Yes [b]			
RHEL 7.2	11.1.38.64					Yes[a]
RHEL 7.3	11.1.38.64					Yes[a]
SLES 11 SP3	11.1.38.64					Yes[a]
SLES 11 SP4	11.1.38.64					Yes[a]

Table 9 Dell EMC-supported Emulex Linux out-of-kernel drivers (continued)

Operating system	Out-of-kernel Driver Version	Supported adapter				
		1/2 Gb	4 Gb	8 Gb	16 Gb	32 Gb
SLES 12	11.1.38.64					Yes[a]
SLES 12 SP2	11.1.38.64	Yes		Yes	Yes	Yes[a]
SLES 12 SP3	12.0.261.26					Yes
SLES 12 SP4	12.0.261.26					Yes
SLES 15	12.0.261.26					Yes

a. For models LPe31000 and LPe32002 only
b. For models LPe1105-HP 4 Gb FC HBA for HP c-Class Blade system only

Marvell (QLogic) Fibre Channel HBA

Installing the driver

Using the QLogic Fibre Channel HBA adapter with the Linux operating system requires adapter driver software. The driver functions at a layer below the Linux SCSI driver to present Fibre Channel devices to the operating system as if they were standard SCSI devices. See the latest Dell EMC Simple Support Matrix for specific qualified kernel and driver versions, and Linux distributions.

Dell EMC supports both in-kernel and out-of-kernel drivers.

(i) **Note:** The installation of the in-kernel driver occurs when you install your Linux distribution of choice.

If your installation requires an out-of-kernel driver, download it from the Dell EMC-approved section of the QLogic website. Follow the links to your adapter for the appropriate operating system and version.

(i) **Note:** The support that is stated in the Dell EMC Simple Support Matrix supersedes versions that are listed in this document.

Dell EMC-supported Marvell (QLogic) driver versions

Following table lists the QLogic driver versions that are supported with the corresponding operating system updates. These driver versions are included by default in the kernel and do not require any installation.

Table 10 Dell EMC supported Marvell (QLogic) Linux in-kernel drivers

Operating system	In-kernel Driver Version	Supported adapter				
		½ Gb	4 Gb	8 Gb	16 Gb	32 Gb
SLES 10 GA	8.01.04-k	Yes	Yes			
RHEL 5.0 OL 5.0	8.01.07-k1	Yes	Yes			
SLES 10 SP1	8.01.07-k3	Yes	Yes			
RHEL 5.1 OEL 5.1	8.01.07-k7	Yes	Yes	Yes		

Table 10 Dell EMC supported Marvell (QLogic) Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		½ Gb	4 Gb	8 Gb	16 Gb	32 Gb
RHEL 5.2 OL 5.2	8.02.00-k5-rhel5.2- 03	Yes	Yes	Yes		
RHEL 5.2 (errata kernels equal to or greater than 2.6.18-92.1.6.el5) OL 5.2 (errata kernels equal to or greater than 2.6.18-92.1.6.0.1.el5)	8.02.00-k5-rhel5.2- 04	Yes	Yes	Yes		
SLES10 SP2	8.02.00-k6-SLES10 -05	Yes	Yes	Yes		
RHEL 5.3 OL 5.3	8.02.00.06.05.03-k	Yes	Yes	Yes		
SuSE SLES 11 GA	8.02.01.03.11.0-k9	Yes	Yes	Yes		
RHEL 5.4 OL 5.4	8.03.00.10.05.04-k	Yes	Yes	Yes		
RHEL 5.4 (errata kernels equal to or greater than 2.6.18-164.2.1.el5) OL 5.4 (errata kernels equal to or greater than 2.6.18-164.2.1.0.1.el5)	8.03.00.1.05.05-k	Yes	Yes	Yes		
SuSE SLES 10 SP3	8.03.00.06.10.3-k4	Yes	Yes	Yes		
RHEL 5.5 OL 5.5	8.03.01.04.05.05-k	Yes	Yes	Yes		
RHEL 5.6 OL 5.6	8.03.01.05.05.06-k	Yes	Yes	Yes		
SLES 10 SP4	8.03.07.03.06.1-k	Yes	Yes	Yes		
SLES 11 SP1 (kernel < 2.6.32.13-0.4.1)	8.03.01.06.11.1-k8	Yes	Yes	Yes		
SLES 11 SP1 (kernel > 2.6.32.13-0.4.1< 2.6.32.27-0.2.2)	8.03.01.07.11.1-k8	Yes	Yes	Yes		
SLES 11 SP1 (kernel > 2.6.32.27-0.2.2)	8.03.01.08.11.1-k8	Yes	Yes	Yes		
SLES 11 SP2	8.03.07.07-k	Yes	Yes	Yes		
SLES 11 SP3	8.04.00.13.11.3-k	Yes	Yes	Yes	Yes	
SLES 11 SP4	8.07.00.18-k	Yes	Yes	Yes	Yes	Yes
SLES 12	8.07.00.08.12.0-k	Yes	Yes	Yes	Yes	
SLES 12 SP1	8.07.00.18-k	Yes	Yes	Yes	Yes	Yes

Table 10 Dell EMC supported Marvell (QLogic) Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		½ Gb	4 Gb	8 Gb	16 Gb	32 Gb
SLES 12 SP2	8.07.00.33-k	Yes	Yes	Yes	Yes	Yes
SLES 12 SP3	9.00.00.00-k	Yes	Yes	Yes	Yes	Yes
SLES 12 SP4	10.00.00.11-k	Yes	Yes	Yes	Yes	Yes
SLES 12 SP5	10.01.00.18-k 10.01.00.21-k 10.01.00.22-k	Yes	Yes	Yes	Yes	Yes
SLES 15	10.00.00.06-k	Yes	Yes	Yes	Yes	Yes
SLES 15 SP1	10.00.00.13-k	Yes	Yes	Yes	Yes	Yes
RHEL 5.7	8.03.07.00.05.07-k	Yes	Yes	Yes		
RHEL 5.8	8.03.07.09.05.08-k	Yes	Yes	Yes		
RHEL 5.9 RHEL 5.10 RHEL 5.11	8.03.07.15.05.09-k	Yes	Yes	Yes		
RHEL 6.0	8.03.05.01.06.1-k0	Yes	Yes	Yes		
RHEL 6.1	8.03.07.03.06.1-k	Yes	Yes	Yes		
RHEL 6.2	8.03.07.05.06.2-k	Yes	Yes	Yes		
RHEL 6.3	8.04.00.04.06.3-k	Yes	Yes	Yes	Yes	
RHEL 6.4	8.04.00.08.06.4-k	Yes	Yes	Yes	Yes	
RHEL 6.5	8.05.00.03.06.5-k2	Yes	Yes	Yes	Yes	
RHEL 6.6	8.07.00.08.06.6-k1	Yes	Yes	Yes	Yes	
RHEL 6.7	8.07.00.16.06.7-k	Yes	Yes	Yes	Yes	
RHEL 6.8	8.07.00.26.06.8-k	Yes	Yes	Yes	Yes	
RHEL 6.9	8.07.00.26.06.8-k	Yes	Yes	Yes	Yes	
RHEL 6.10	8.07.00.26.06.8-k	Yes	Yes	Yes	Yes	Yes
RHEL 7.0	8.06.00.08.07.0-k2	Yes	Yes	Yes	Yes	
RHEL 7.1	8.07.00.08.07.1-k2	Yes	Yes	Yes	Yes	
RHEL 7.2	8.07.00.18.07.2-k	Yes	Yes	Yes	Yes	
RHEL 7.3	8.07.00.33.07.3-k 8.07.00.38.07.4-k1 (Errata kernel 3.10.0-514.26.2.el7 and above)	Yes	Yes	Yes	Yes	Yes
RHEL 7.4	8.07.00.38.07.4-k1	Yes	Yes	Yes	Yes	Yes

Table 10 Dell EMC supported Marvell (QLogic) Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		½ Gb	4 Gb	8 Gb	16 Gb	32 Gb
RHEL 7.5	9.00.00.00.07.5-k1	Yes	Yes	Yes	Yes	Yes
RHEL 7.6	10.00.00.06.07.6-k	Yes	Yes	Yes	Yes	Yes
RHEL 7.7	10.00.00.12.07.7-k	Yes	Yes	Yes	Yes	Yes
RHEL 7.8	10.01.00.20.07.8-k	Yes	Yes	Yes	Yes	Yes
RHEL 8.0	10.00.00.07.08.0-k1			Yes	Yes	Yes
RHEL 8.1	10.01.00.15.08.1-k1			Yes	Yes	Yes
Oracle Linux 5. x UEK R1 [2.6.32-100]	8.03.01.02.32.1-k9	Yes	Yes	Yes		
Oracle Linux 6. x UEK R1 [2.6.32-100]	8.03.07.03.32.1-k	Yes	Yes	Yes		
Oracle Linux 5. x UEK R1 U1 [2.6.32-200] Oracle Linux 6. x UEK R1 U1 [2.6.32-200]	8.03.07.04.32.1-k	Yes	Yes	Yes		
Oracle Linux 5. x UEK R1 U2 [2.6.32-300] Oracle Linux 6.x UEK R1 U2 [2.6.32-300] Oracle Linux 5.x UEK R1 U3 [2.6.32-400] Oracle Linux 6.x UEK R1 U3 [2.6.32-400]	8.03.07.08.32.1-k	Yes	Yes	Yes		
Oracle Linux 5.x UEK R2 [2.6.39-100] Oracle Linux 6.x UEK R2 [2.6.39-100]	8.03.07.12.39.0-k	Yes	Yes	Yes		
Oracle Linux 5.x UEK R2 U1 [2.6.39-200] Oracle Linux 6.x UEK R2 U1 [2.6.39-200]	8.04.00.03.39.0-k	Yes	Yes	Yes		
Oracle Linux 5.x UEK R2 U2 [2.6.39-300] Oracle Linux 6.x UEK R2 U2 [2.6.39-300]	8.04.00.08.39.0-k	Yes	Yes	Yes		
Oracle Linux 5.x UEK R2 U3 [2.6.39-400]	8.04.00.11.39.0-k, 8.05.00.03.39.0-k	Yes	Yes	Yes		

Table 10 Dell EMC supported Marvell (QLogic) Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		½ Gb	4 Gb	8 Gb	16 Gb	32 Gb
Oracle Linux 6.x UEK R2 U3 [2.6.39-400]						
Oracle Linux 5.x UEK R2 U4 [2.6.39-400.109] Oracle Linux 6.x UEK R2 U4 [2.6.39-400.109]	8.05.00.03.39.0-k	Yes	Yes	Yes		
Oracle Linux 5.x UEK R2 U5 [2.6.39-400.209] Oracle Linux 6.x UEK R2 U5 [2.6.39-400.209]	8.05.00.03.39.0-k	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 [3.8.13-16]	8.05.00.03.39.0-k	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U1 [3.8.13-26]	8.06.00.14.39.0-k	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U2 [3.8.13-35] Oracle Linux 6.x UEK R3 U3 [3.8.13-44] Oracle Linux 7.x UEK R3 U2 [3.8.13-35] Oracle Linux 7.x UEK R3 U3 [3.8.13-44]	8.07.00.08.39.0-k1	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U4 [3.8.13-55] Oracle Linux 7.x UEK R3 U4 [3.8.13-55] Oracle Linux 6.x UEK R3 U5 [3.8.13-68] Oracle Linux 7.x UEK R3 U5 [3.8.13-68]	8.07.00.16.39.0-k	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U6 [3.8.13-98] Oracle Linux 7.x UEK R3 U6 [3.8.13-98]	8.07.00.18.39.0-k	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R3 U7 [3.8.13-118] Oracle Linux 7.x UEK R3 U7 [3.8.13-118]	8.07.00.26.39.0-k	Yes	Yes	Yes	Yes	
Oracle Linux 6.x UEK R4 [4.1.12]	8.07.00.26.39.0-k	Yes	Yes	Yes	Yes	Yes

Table 10 Dell EMC supported Marvell (QLogic) Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		½ Gb	4 Gb	8 Gb	16 Gb	32 Gb
Oracle Linux 7.x UEK R4 [4.1.12]						
Oracle Linux 6.x UEK R4 U1 [4.1.12-37] Oracle Linux 7.x UEK R4 U1 [4.1.12-37]	8.07.00.33.40.0-k	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U2 [4.1.12-61] Oracle Linux 7.x UEK R4 U2 [4.1.12-61]	8.07.00.38.40.0-k	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U4 [4.1.12-94] Oracle Linux 7.x UEK R4 U4 [4.1.12-94]	8.07.00.38.40.0-k	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U5 [4.1.12-103] Oracle Linux 7.x UEK R4 U5 [4.1.12-103]	8.07.00.38.40.0-k	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U6 [4.1.12-112] Oracle Linux 7.x UEK R4 U6 [4.1.12-112]	9.00.00.00.40.0-k	Yes	Yes	Yes	Yes	Yes
Oracle Linux 6.x UEK R4 U7 [4.1.12-124] Oracle Linux 7.x UEK R4 U7 [4.1.12-124]	9.00.00.00.40.0-k 9.00.00.00.41.0.-k1	Yes	Yes	Yes	Yes	Yes
Oracle Linux 7.x UEK R5 [4.1.35-1818]	10.00.00.06-k1 10.00.00.07-k1	Yes	Yes	Yes	Yes	Yes
Oracle Linux 7.x UEK R5 U1 [4.1.35-1844]	10.00.00.07-k1	Yes	Yes	Yes	Yes	Yes
Oracle Linux 7.x UEK R5 U2 [4.1.35-1902]	10.00.00.07-k1 10.00.00.13-k	Yes	Yes	Yes	Yes	Yes
XenServer [x86_64] 6.2	8.05.00.03.55.6-k	Yes	Yes	Yes		
XenServer [x86_64] 6.5	8.07.00.09.66.5-k	Yes	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.0	8.07.00.33.77.0-k	Yes	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.1	8.07.00.41.77.1-k	Yes	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.2	8.07.00.41.77.1-k	Yes	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.3	8.07.00.56.71.0-k	Yes	Yes	Yes	Yes	Yes

Table 10 Dell EMC supported Marvell (QLogic) Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter				
		½ Gb	4 Gb	8 Gb	16 Gb	32 Gb
XenServer [x86_64] 7.4	8.07.00.56.71.0-k	Yes	Yes	Yes	Yes	Yes

Table 11 Dell EMC supported QLogic Linux out-of-kernel drivers

Operating system	Out-of-kernel Driver Version	Supported adapter				
		½ Gb	4 Gb	8 Gb	16 Gb	32 Gb
RHEL 6.7	8.07.00.39.06.0-k					Yes[a]
RHEL 6.8	8.07.00.39.06.0-k					Yes[a]
RHEL 6.9	8.07.00.39.06.0-k					Yes[a]
RHEL 7.0	8.07.00.39.07.0-K					Yes[a]
RHEL 7.1	8.07.00.39.07.0-K					Yes[a]
RHEL 7.2	8.07.00.39.07.0-K					Yes[a]
SLES 11 SP3	8.07.00.39.11.3-k					Yes[a]
SLES 11 SP4	8.07.00.39.11.3-k					Yes[a]
SLES 12	8.07.00.39.11.3-k					Yes[a]
SLES 12 SP1	8.07.00.39.11.3-k					Yes[a]
SLES 12 SP2	8.07.00.39.11.3-k					Yes[a]

a. For model QLE2740 and QEL2742 only

Marvell (Brocade) Fibre Channel HBA

The following Brocade HBAs/CNAs are now provided by QLogic under the same model numbers:

- Brocade 1860 Fabric Adapters
- Brocade 815/825 and 415/425 Fibre Channel Host Bus Adapters (HBAs)
- Brocade 1010/1020 Converged Network Adapters (CNAs)
- OEM HBA and CNA mezzanine adapters (1007, 1867, 1869 & BR1741M-k)

Installing the driver

The driver versions that are listed in table below are included by default in the kernel and do not require installation.

Dell EMC-supported Brocade driver versions

The following table lists the Brocade driver versions that are supported with the corresponding operating system updates.

Table 12 Dell EMC-supported Brocade Linux in-kernel drivers

Operating system	In-kernel Driver Version	Supported adapter			
		4 Gb	8 Gb	16 Gb	32 Gb
RHEL 5.5	2.1.2.0	Yes	Yes		
RHEL 5.6	2.1.2.2	Yes	Yes		
RHEL 5.7	2.3.2.3	Yes	Yes		
RHEL 5.8	3.0.2.2	Yes	Yes		
RHEL 5.9 RHEL 5.10 RHEL 5.11	3.0.23.0	Yes	Yes	Yes	
RHEL 6.0	2.1.2.1	Yes	Yes		
RHEL 6.1	2.3.2.3	Yes	Yes		
RHEL 6.2	3.0.2.2 (BFA) 3.0.2.2r (BNA)	Yes	Yes		
RHEL 6.3	3.0.2.2	Yes	Yes	Yes	
RHEL 6.4	3.0.23.0	Yes	Yes	Yes	
RHEL 6.5	3.2.21.1	Yes	Yes	Yes	
RHEL 6.6 RHEL 6.7	3.2.23.0	Yes	Yes	Yes	Yes
RHEL 6.8 RHEL 6.9 RHEL 6.10	3.2.25.0	Yes	Yes	Yes	Yes
RHEL 7.0 RHEL 7.1 RHEL 7.2	3.2.23.0	Yes	Yes	Yes	Yes
RHEL 7.3	3.2.25.0	Yes	Yes	Yes	Yes
RHEL 7.4 RHEL 7.5 RHEL 7.6 RHEL 7.7	3.2.25.1	Yes	Yes	Yes	Yes
SLES 10 SP3	2.0.0.0	Yes	Yes		
SLES 10 SP4	2.3.2.1	Yes	Yes		
SLES 11	1.1.0.2	Yes	Yes		
SLES 11 SP1	2.1.2.1	Yes	Yes		
SLES 11 SP2	3.0.2.2	Yes	Yes		

Table 12 Dell EMC-supported Brocade Linux in-kernel drivers (continued)

Operating system	In-kernel Driver Version	Supported adapter			
		4 Gb	8 Gb	16 Gb	32 Gb
SLES 11 SP3	3.1.2.1	Yes	Yes	Yes	
SLES 11 SP4 SLES 12 SLES 12 SP1	3.2.23.0	Yes	Yes	Yes	
SLES 12 SP2 Kernel less than 4.4.49-92.11.1	3.2.25.0	Yes	Yes	Yes	Yes
SLES 12 SP2 Kernel equal to or greater than 4.4.49-92.11.1	3.2.25.1	Yes	Yes	Yes	Yes
SLES 12 SP3 SLES 15	3.2.25.1	Yes	Yes	Yes	Yes
XenServer [x86_64] 6.2	3.2.1.1	Yes	Yes		
XenServer [x86_64] 6.5	3.2.1.1	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.0	3.2.1.1	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.1	3.2.23.0	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.2	3.2.23.0	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.3	3.2.23.0	Yes	Yes	Yes	Yes
XenServer [x86_64] 7.4	3.2.23.0	Yes	Yes	Yes	Yes

SNIA API for third-party software (Solution Enabler)

The SNIA Common HBA API is an industry standard, programming interface for accessing management information in Host Bus Adapters (HBA). Developed through the Storage Networking Industry Association (SNIA), the HBA API has been adopted by Storage Area Network vendors to help manage, monitor, and deploy storage area networks in an interoperable way.

There are certain SNIA libraries for Broadcom (also referred to as Emulex) and Cavium (also referred to as QLogic) which need to be installed so that Solutions Enabler CLI can obtain host HBA information. By default, SNIA libraries are not pre-installed on the host. Without the SNIA API Library installed, HBA information will not be able to be gathered from the host.

Here is an example of a failed command:

```
[root@Server1 bin]# ./syminq HBA Host Name : Server1
Could not find any HBAs to list.
```

If you see the previous output, you need to install SNIA API libraries.

Installing SNIA API libraries for Qlogic HBA

Perform the following steps to install SNIA API libraries for Qlogic HBA:

Procedure

1. Run the following command to find the vendor information and model:

   ```
   Server1:~ # cat /sys/class/fc_host/host1/symbolic_name QLE2742
   FW:v8.03.04 DVR:v8.07.00.33-k
   ```

2. Use the following URL, and open the Cavium download page:

   http://driverdownloads.qlogic.com/QLogicDriverDownloads_UI/DefaultNewSearch.aspx

3. On the Cavium download page, select **Fibre Channel Adapters**. Select the specific model identified in step 1 (**QLE2742**), and right-click **Operating System** and click **Go**.

4. Select the **API Libraries** tab and download the **FC-FCoE API for Linux.zip** file (qlapi-v6.04build8-rel.tgz).

5. Extract the file and use the following commands to install the SNIA API libraries:

   ```
   Server1:~/qlapi # ./libinstall
   Setting up QLogic HBA API library...
   Please make sure the /usr/lib/libqlsdm.so file is not in use.
   Installing 32bit api binary for x86_64.
   Installing 64bit api binary for x86_64.
   Done.
   ```

6. After the successful installation of the SNIA API libraries , /etc/hba.conf will be created (if the file does not exist) and will have the following entries:

   ```
   Server1:~/qlapi # cat /etc/hba.conf
   qla2xxx            /usr/lib/libqlsdm.so
   qla2xxx64        /usr/lib64/libqlsdm.so
   ```

   (i) Note: Linux SuperInstaller package under the Drivers category on Cavium website also includes SNIA APIs and can be applied.

Installing SNIA API libraries for Emulex HBA

Perform the following steps to install the SNIA API libraries for Emulex HBA:

Procedure

1. Use the following command to find the vendor information and model:

   ```
   Server1:~ # cat /sys/class/fc_host/host3/symbolic_name

   Emulex LPe12002 FV2.00A4 DV11.4.142.21

   Server1:~ #
   ```

2. Go to the following URL, and open the Broadcom HBA OneCommand Manager download page:

   https://www.broadcom.com/products/storage/fibre-channel-host-bus-adapters/onecommand-manager-centralized#downloads

3. Select the **Management Software and Tools** tab and download the **OneCommand Manager Core Application Kit (CLI) for Linux** for your operating system (elxocmcore-rhel5-rhel6-rhel7-11.0.243.13-1.tgz).

4. Extract the file and use the following commands to install the SNIA API libraries:

```
Server1:~/elxocmcore# ./install.sh

Beginning OneCommand Manager Core Kit Installation...

Installing ./x86_64/rhel-7/elxocmcorelibs-11.2.156.23-1.x86_64.rpm

Installing ./x86_64/rhel-7/
hbaapiwrapper-32bit-11.2.156.23-1.x86_64.rpm

Installing ./x86_64/rhel-7/hbaapiwrapper-11.2.156.23-1.x86_64.rpm

Installing ./x86_64/rhel-7/elxocmcore-11.2.156.23-1.x86_64.rpm
```

5. After the successful installation of the SNIA API libraries, `/etc/hba.conf` will be created (if the file does not exist) and will have the following entry:

```
Server1:~/qlapi # cat /etc/hba.conf

com.emulex.emulexapilibrary64 /usr/lib64/libemulexhbaapi.so

com.emulex.emulexapilibrary32 /usr/lib/libemulexhbaapi.so
```

Hitachi Virtage

Hitachi Virtage can partition physical server resources by constructing multiple logical partitions (LPARs) that are isolated. Each of these environments can run independently. A different operating system (called a Guest OS) can run on each LPAR on a single physical server.

Hitachi Virtage is a integrated (firmware) feature on Hitachi server products, including the BladeSymphony 2000, the BladeSymphony 500, and BladeSymphony 320. This feature requires no separate operating system layer or third-party virtualization software and can easily be activated or deactivated based on customer needs using the Baseboard Management Controller (BMC). This is supported by Hitachi, like their system BIOS, BMC, service processor, and other hardware functions.

Besides dedicating a Fibre Channel card to a logical partition, Hitachi also offers Fibre Channel I/O virtualization for Hitachi Virtage. This allows multiple logical partitions to access a storage device through a single Fibre Channel card, allowing fewer physical connections between server and storage and increasing the utilization rates of the storage connections. For additional information, see Hitachi website.

The Fibre Channel drivers used for logical partitions are optimized for the Hitachi Virtage feature. Such optimizations involve different behaviors, which make the default settings not suitable for working with PowerPath.

Complete the following step to check if default settings are used:

Check the dev_loss_tmo value.

 # grep .

```
/sys/class/fc_remote_ports/rport-*/*tmo
/sys/class/fc_remote_ports/rport-0:0-0/dev_loss_tmo:1
/sys/class/fc_remote_ports/rport-0:0-0/fast_io_fail_tmo:off
/sys/class/fc_remote_ports/rport-0:0-1/dev_loss_tmo:1
/sys/class/fc_remote_ports/rport-0:0-1/fast_io_fail_tmo:off
/sys/class/fc_remote_ports/rport-1:0-0/dev_loss_tmo:1
/sys/class/fc_remote_ports/rport-1:0-0/fast_io_fail_tmo:off
/sys/class/fc_remote_ports/rport-1:0-1/dev_loss_tmo:1
/sys/class/fc_remote_ports/rport-1:0-1/fast_io_fail_tmo:off
```

......

If dev_loss_tmo equals 1, then the default settings are used.

Once it is confirmed that the default settings are used for a Fibre Channel driver on a logical partition , the following tunings should be applied before installing PowerPath:

1. Enter the following commands:
   ```
   # cd /opt/hitachi/drivers/hba

   # ./hfcmgr -E hfc_rport_lu_scan 1

   # ./hfcmgr -p all ld 6

   # ./hfcmgr -E hfc_dev_loss_tmo 10

   # cd /boot

   # /sbin/mkinitrd -f <image-file-name>.img `uname -r`
   ```

2. Create a new boot entry in `/boot/grub/menu.lst` with the initrd image generated in the previous step.

3. Reboot the operating system with the new initrd.

ⓘ **Note:** Hitachi Virtage is used across the world. So, some customers may set hardware clock to values other than UTC. This is not usually a problem, but when multipath software, such as PowerPath, is leveraged for a boot from SAN (BFS) setup, PowerPath may fail to mark path status accordingly when a link down events happens. This can be solved by changing the hardware clock to UTC. See http://support.emc.com/kb/209658 for details.

Fibre Channel over Ethernet connectivity

Introduction

This section provides a basic introduction to Fibre Channel over Ethernet (FCoE), which is a new approach to I/O consolidation that has been defined in the FC-BB-5 T11 work group.

The information in this chapter was derived from the following sources, which also provide more details about FCoE, including encapsulation, frame format, address mapping, lossless Ethernet, and sample topologies:

- *Fibre Channel over Ethernet - A review of FCoE Today*

- *Silvano, Gai, Data Center Network and Fibre Channel over Ethernet, Nuova Systems Inc., 2008*

I/O consolidation, simply defined, is the ability to carry different types of traffic, having different traffic characteristics and handling requirements, over the same physical media. I/O consolidation's most difficult challenge is to satisfy the requirements of different traffic classes within a single network. Since Fibre Channel is the dominant storage protocol in the data center, any viable I/O consolidation solution for storage must allow for the FC model to be seamlessly integrated. FCoE meets this requirement in part by encapsulating each Fibre Channel frame inside an Ethernet frame.

The goal of FCoE is to provide I/O consolidation over Ethernet, allowing Fibre Channel and Ethernet networks to share a single, integrated infrastructure, thereby reducing network complexities in the data center. An example is shown in the following figure.

Figure 7 Typical topology versus FCoE example using NEX-5020
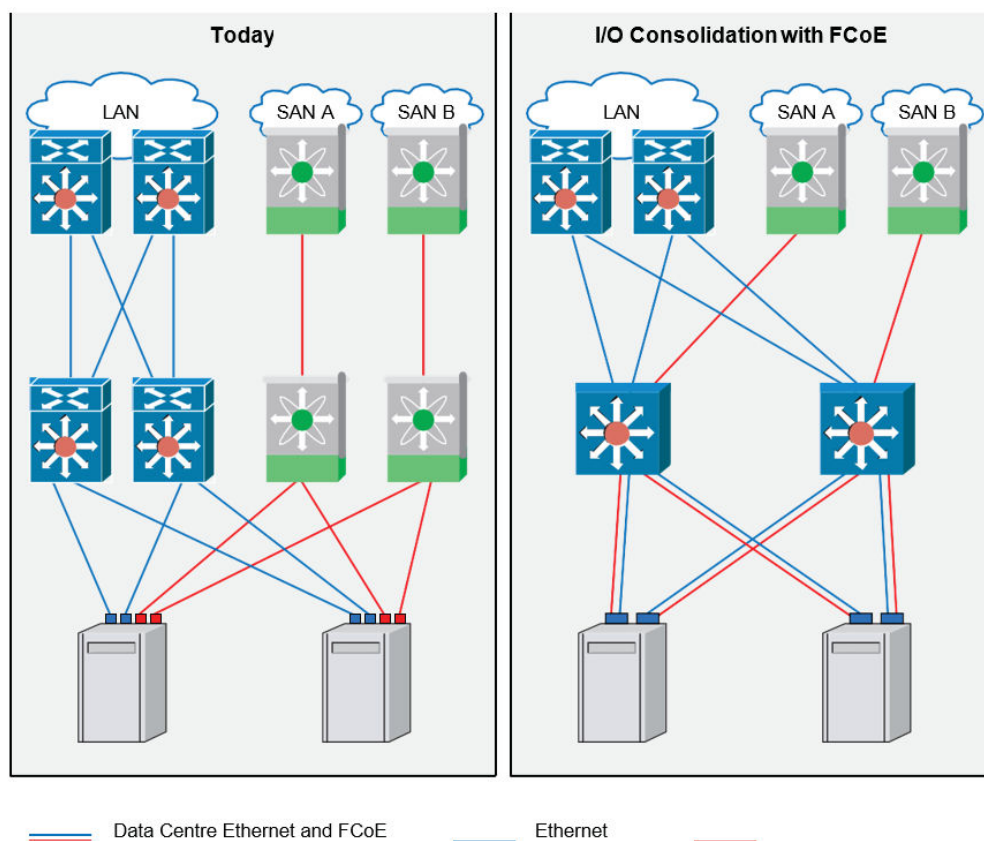
FCoE consolidates both SANs and Ethernet traffic onto one Converged Network Adapter (CNA), eliminating the need for using separate host bus adapters (HBAs) and network interface cards (NICs).

For more information about Fibre Channel over Ethernet, see the *Fibre Channel over Ethernet (FCoE) Data Center Bridging (DCB) Concepts and Protocol TechBook* and the *Fibre Channel over Ethernet (FCoE) Data Center Bridging (DCB) Case Studies TechBook*, available at Dell EMC E-Lab Interoperability Navigator , under the **Topology Resource Center** tab.

Configuring the Linux host

From a logical connectivity point of view, an all-FCoE fabric is identical to an FC fabric and supports all of the same functionality including zoning, a distributed name server, and RSCNs. As a result, the same types of scalability limits apply to both FC and FCoE fabrics, such as the maximum number of hops, VN_Ports, and domains.

Zoning best practices

Dual or multiple paths between the hosts and the storage system are required. This includes redundant HBAs, a robust implementation, strictly following management policies and procedures, and dual attachment to storage systems. These requirements are the same as for Fibre Channel HBAs.

The common guidelines include:

- Redundant paths from host to storage
- Use of multipath software and use of failover modes

- Dual Fabrics
- Single initiator zoning

CNAs

Before the CNA can access the remote storage, the CNA must be installed from a specific vendor in order for it to operate normally. The following steps summarize this process:

1. Install the CNA.
2. Verify the boot code and firmware version.
3. Update the boot code and firmware version if needed.
4. Install the driver.
5. Install the management application kit.
6. Connect to storage.

See the vendor-specific user manual for the detailed installation process on Linux systems.

Emulex

Using the Emulex Coverged Network Adapter with the Linux operating system requires adapter driver software. The driver functions at a layer below the Linux SCSI driver to present Fibre Channel devices to the operating system as if they were standard SCSI devices.

Dell EMC supports both in-kernel and out-of-kernel drivers.

In-kernel driver versions are included by default in the kernel and do not require any installation. Table 13 lists the Emulex driver versions that are supported with the corresponding operating system updates. Table 14 lists the Emulex out-of-kernel driver versions that are supported with the corresponding operating system updates.

See the latest Dell EMC Simple Support Matrix for your specific Linux distribution, kernel version, and driver to determine if you need to proceed with the following out-of-kernel instructions.

If your installation requires an out-of kernel driver, download the latest from the Dell EMC-approved section of the Broadcom Support Documents and Downloads page.

Table 13 Supported FCOE in-kernel drivers

Operating system	Driver Version	Supported Adapters
		CNA
RHEL 5.4	8.2.0.48.2p 8.2.0.48.3p (errata kernels 2.6.18-164.11.1.0.1.el5 and later)	Yes[a]
RHEL 5.5	8.2.0.63.3p	Yes[a]
RHEL 5.6	8.2.0.87.1p	Yes[a]
RHEL 5.7	8.2.0.96.2p	Yes[a]
RHEL 5.8	8.2.0.108.4p	Yes[b]
RHEL 5.9 RHEL 5.10	8.2.0.128.3p	Yes[b]

Table 13 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
RHEL 5.11		
RHEL 6.0	8.3.5.17	Yes[b]
RHEL 6.1	8.3.5.30.1p	Yes[b]
RHEL 6.2	8.3.5.45.4p	Yes[b]
RHEL 6.3	8.3.5.68.5p	Yes[b]
RHEL 6.4	8.3.5.86.1p	Yes[b]
RHEL 6.5	8.3.7.21.4p	Yes[b]
RHEL 6.6	10.2.8020.1	Yes[b]
RHEL 6.7	10.6.0.20	Yes[b]
RHEL 6.8	11.0.0.4 11.0.181.0 (Errata kernel 2.6.32-642.3.1.el6 and 2.6.32-642.4.2.el6)	Yes[b]
RHEL 6.9	10.6.0.3r	Yes[b]
RHEL 7.0	8.3.7.34.3p	Yes[b]
RHEL 7.1	10.2.8021.1	Yes[b]
RHEL 7.2	10.7.0.1	Yes[b]
RHEL 7.3	11.1.0.2	Yes[b]
RHEL 7.4	11.2.0.2	Yes[b]
RHEL 7.5	11.4.0.0r	Yes[b]
RHEL 7.6	11.4.0.0r	Yes[b]
RHEL 7.7	11.4.0.0r	Yes[b]
RHEL 7.8	12.0.0.0r	Yes[b]
RHEL 8.0	12.0.0.0	Yes[c]
Linux OL 6.x [x86_64] UEK R1 [2.6.23-100]	8.3.5.30.1p	Yes[a]
Linux OL 5.x [x86_64] UEK R1 [2.6.32-100]	8.3.18	Yes[a]
Linux OL 5.x [x86_64] UEK R1 U1 [2.6.32-200] Linux OL 6.x [x86_64] UEK R1 U1 [2.6.32-200]	8.3.5.44	Yes[a]
Linux 6.x [x86_64] UEK R1 U2 [2.6.32-300] Linux OL 5.x [x86_64] UEK R1 U2 [2.6.32-300] Linux OL 6.x [x86_64] UEK R1 U2 [2.6.32-300]	8.3.5.45.4p	Yes[b]

Table 13 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
Linux OL 5.x [x86_64] UEK R1 U3 [2.6.32-400] Linux OL 6.x [x86_64] UEK R1 U3 [2.6.32-400]		
Linux OL 5.x [x86_64] UEK R2 [2.6.39-100] Linux OL 6.x [x86_64] UEK R2 [2.6.39-100]	8.3.5.58.2p	Yes[a]
Linux OL 5.x [x86_64] UEK R2 U1 [2.6.39-200] Linux OL 6.x [x86_64] UEK R2 U1 [2.6.39-200]	8.3.5.68.6p	Yes[a]
Linux OL 5.x [x86_64] UEK R2 U2 [2.6.39-300] Linux OL 6.x [x86_64] UEK R2 U2 [2.6.39-300]	8.3.5.82.2p	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U3 [2.6.39-400] Linux OL 6.x [x86_64] UEK R2 U3 [2.6.39-400]	8.3.5.82.2p 8.3.7.10.4p (This is in-kernel driver is only available on errata kernels equal to greater than 2.6.39-400.109.1 el6uek)	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U4 [2.6.39-400.109] Linux OL 6.x [x86_64] UEK R2 U4 [2.6.39-400.109]	8.3.7.10.4p	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U5 [2.6.39-400.209] Linux OL 6.x [x86_64] UEK R2 U5 [2.6.39-400.209]	8.3.7.26.3p	Yes[b]
Linux OL 6.x [x86_64] UEK R3 [3.8.13-16]	8.3.7.26.2p	Yes[b]
Linux OL 6.x [x86_64] UEK R3 U1 [3.8.13-26] Linux OL 6.x [x86_64] UEK R3 U2 [3.8.13-35] Linux OL 6.x [x86_64] UEK R3 U3 [3.8.13-44]	8.3.7.34.4p	Yes[b]
Linux OL 6.x [x86_64] UEK R3 U4 [3.8.13-55]	10.2.8061.0	Yes[b]
Linux OL 7.x [x86_64] UEK R3 U2 [3.8.13-35] Linux OL 7.x [x86_64] UEK R3 U3 [3.8.13-44]	8.3.7.34.4p	Yes[b]
	8.3.7.34.4p	Yes[b]
Linux OL 7.x [x86_64] UEK R3 U4 [3.8.13-55]	10.2.8061.0	Yes[b]
Linux OL 7.x [x86_64] UEK R3 U5 [3.8.13-68]	10.6.61.0	Yes[b]
Oracle Linux 6.x UEK R3 U6 [3.8.13-98]	10.4.200.0o	Yes[b]

Table 13 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
Oracle Linux 7.x UEK R3 U6 [3.8.13-98]		
Oracle Linux 6.x UEK R3 U7 [3.8.13-118] Oracle Linux 7.x UEK R3 U7 [3.8.13-118]	10.6.0.1o	Yes[b]
Oracle Linux 6.x UEK R4 [4.1.12] Oracle Linux 7.x UEK R4 [4.1.12]	10.6.0.1o	Yes[b]
Oracle Linux 6.x UEK R4 U1 [4.1.12-37] Oracle Linux 7.x UEK R4 U1 [4.1.12-37]	10.6.0.2	Yes[b]
Oracle Linux 6.x UEK R4 U2 [4.1.12-61] Oracle Linux 7.x UEK R4 U2 [4.1.12-61]	11.0.0.0	Yes[b]
Oracle Linux 6.x UEK R4 U4 [4.1.12-94] Oracle Linux 7.x UEK R4 U4 [4.1.12-94]	11.2.1.0	Yes[b]
Oracle Linux 6.x UEK R4 U5 [4.1.12-103] Oracle Linux 7.x UEK R4 U5 [4.1.12-103]	11.2.1.0	Yes[b]
Oracle Linux 6.x UEK R4 U6 [4.1.12-112] Oracle Linux 7.x UEK R4 U6 [4.1.12-112]	11.2.1.0	Yes[b]
Oracle Linux 6.x UEK R4 U7 [4.1.12-124] Oracle Linux 7.x UEK R4 U7 [4.1.12-124]	11.2.1.0	Yes[b]
Oracle Linux 7.x UEK R5 [4.1.35-1818]	11.4.0.0	Yes[b]
Oracle Linux 7.x UEK R5 U1 [4.1.35-1844]	11.4.0.0	Yes[b]
Oracle Linux 7.x UEK R5 U2 [4.1.35-1902]	11.4.0.0	Yes[b]
SLES 10 SP4	8.2.0.92.1p	Yes[a]
SuSE SLES 11 GA	8.2.8.14	Yes[a]
SuSE SLES 11 SP1	8.3.5.8.1p 8.3.5.8.2p	Yes[a]
SuSE SLES 11 SP2	8.3.5.48.2p 8.3.5.48.3p	Yes[b]
SuSE SLES 11 SP3	8.3.7.10.6p	Yes[b]
SuSE SLES 11 SP4	10.4.8000.0.	Yes[b]
SuSE SLES 12	10.2.8040.1	Yes[b]
SLES 12 SP1	10.5.0.2	Yes[b]
SLES 12 SP2	11.1.0.1	Yes[b]
SLES 12 SP3	11.4.0.1	Yes[b]

Table 13 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
SLES 12 SP4	12.0.0.6	Yes[b]
SLES 15	12.0.0.1	Yes[b]
SLES 15 SP1	12.2.0.0	Yes[b]
XenServer [x86_64] 6.0.2	8.3.5.39.1p	Yes[a]
XenServer [x86_64] 6.1	8.3.5.77.1p	Yes[a]
XenServer [x86_64] 6.2	8.3.7.14	Yes[b]
XenServer [x86_64] 6.5	10.2.340.0	Yes[b]
XenServer [x86_64] 7.0	11.0.0.12	Yes[b]
XenServer [x86_64] 7.1	11.1.210.1	Yes[b]
XenServer [x86_64] 7.2	11.1.210.1	Yes[b]
XenServer [x86_64] 7.3	11.1.210.1	Yes[b]
XenServer [x86_64] 7.4	11.1.210.1	Yes[b]

a. For models LP21xxx, OCe10xxx, and OCe11xxx CNAs.
b. For models LP21xxx, OCe10xxx, OCe11xxx and OCe14xxx CNAs.
c. some Emulex CNA is not support since RHEL8.0, see RHEL8.0 release doc (https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/8/html-single/considerations_in_adopting_rhel_8/index#removed-adapters_hardware-enablement)

Table 14 Supported FCOE out-of-kernel drivers

Operating system	Driver Version	Supported Adapters
		CNA
RHEL 5.5 - 5.8	10.4.255.16	Yes[a]
RHEL 6.2	10.2.261.0	Yes[b]
RHEL 6.3	10.2.261.0	Yes[b]
RHEL 6.4	10.2.261.0	Yes[b]
RHEL 6.5	10.2.261.0	Yes[b]
RHEL 7.0	10.2.261.0	Yes[b]
RHEL 7.1	10.2.261.0	Yes[b]
SLES 10 SP3-SP4	8.2.2.10.1p-1	Yes[a]
SLES 11 SP2-SP3	10.4.255.16	Yes[a]
SLES 12	10.4.255.16	Yes[a]
SLES 12 SP1	11.2.1153.13	Yes[b]
SLES 12 SP2-SP3	11.4.1186.5-1	Yes[b]

Table 14 Supported FCOE out-of-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
XenServer 6.1-6.2	10.0.794.0	Yes[a]

a. For models OCe11xxx and OCe14xxx CNAs only
b. For model IBM Flex SystemCN4054R 2-port 10Gb Virtual Fabric Adapter (00Y3306) only.

Utility

OneCommand Manager, the successor to Emulex market proven HBAnyware management application, provides centralized management of Emulex fabric (Fibre Channel HBA) and network (iSCSI UCNA, FCoE CNA and 10 Gb/s Ethernet Adapter) connectivity solutions in physical and virtual server deployments through a graphical user interface (GUI), as well as a fully scriptable command line user interface (CLI). OneCommand Manager provides powerful adapter provisioning and diagnostic capabilities helping to increase administration efficiency and business agility.

To download the application kit, go to Broadcom Support Documents and Downloads page.

Update firmware and boot code

Emulex CNA firmware and boot code are provided in a single flash image called a .UFI file. The .UFI file contains all of the files that support all OneConnect adapters. The boot code allows you to designate a device that is attached to the adapter as a boot device. Emulex supports the following types of boot code for CNA:

*   Preboot eXecution Environment (PXE) boot for NIC adapters in x86 and x64 systems
*   x86 BootBIOS for FCoE adapters in x86 and x64 systems
*   iSCSI boot for iSCSI adapters in x86 and x64 systems
*   UEFIBoot for NIC, iSCSI, and FCoE adaptersin x64 systems. This provides system boot capability through the use of the UEFI (Unified Extensible Firmware Interface) Shell. It also functions on UEFI 2.x-based platforms through the HII (Human Interface Infrastructure) interface.
*   OpenBoot for FCoE adapters in Sun SPARC systems (OpenBoot is also called FCode)

To download the latest firmware and boot code, go to the Broadcom Support Documents and Downloads page.

The OneCommand Manager application enables you to update firmware for a single adapter or simultaneously for multiple adapters. See the *User Manual*, which can be downloaded from the Broadcom Support Documents and Downloads page.

You can also download the offline bootable ISO image to update the firmware and boot code.

QLogic 8xxx Series

Using the QLogic Converged Network Adapter with the Linux operating system requires adapter driver software. The driver functions at a layer below the Linux SCSI driver to present Fibre Channel devices to the operating system as if they were standard SCSI devices.

Dell EMC supports both in-kernel and out-of-kernel drivers.

In-kernel driver versions are included by default in the kernel and do not require any installation. Out-of-kernel driver versions from the vendor need manual installation. See Table 15 for supported in-kernel driver versions. Table 16 lists the out-of-kernel driver versions supported with the corresponding operating system updates.

See the latest Dell EMC Simple Support Matrix for your specific Linux distribution, kernel version, and driver to determine if you need to proceed with the following out-of-kernel instructions.

If your installation requires an out-of-kernel driver, download it from the Dell EMC-approved section of the QLogic Driver Downloads/Documentation.

Table 15 Supported FCOE in-kernel drivers

Operating system	Driver Version	Supported Adapters
		CNA
RHEL 5.4	8.03.00.10.05.04-k 8.03.00.1.05.05-k	Yes[a]
RHEL 5.5	8.03.01.04.05.05-k	Yes[a]
RHEL 5.6	8.03.01.05.05.06-k	Yes[b]
RHEL 5.7	8.03.07.00.05.07-k	Yes[b]
RHEL 5.8	8.03.07.09.05.08-k	Yes[b]
RHEL 5.9 RHEL 5.10 RHEL 5.11	8.03.07.15.05.09-k	Yes[b]
RHEL 6.0	8.03.05.01.06.1-k0	Yes[b]
RHEL 6.1	8.03.07.03.06.1-k	Yes[b]
RHEL 6.2	8.03.07.05.06.2-k	Yes[b]
RHEL 6.3	8.04.00.04.06.3-k	Yes[b]
RHEL 6.4	8.04.00.08.06.4-k	Yes[b]
RHEL 6.5	8.05.00.03.06.5-k2	Yes[b]
RHEL 6.6	8.07.00.08.06.6-k1	Yes[b]
RHEL 6.7	8.07.00.16.06.7-k	Yes[b]
RHEL 6.8	8.07.00.26.06.8-k	Yes[b]
RHEL 6.9	8.07.00.26.06.8-k	Yes[b]
RHEL 6.10	8.07.00.26.06.8-k	Yes[b]
RHEL 7.0	8.06.00.08.07.0-k2 8.06.00.08.07.0-k3 (Errata kernel 3.10.0-123.6.3.el7 or higher)	Yes[b]
RHEL 7.1	8.06.00.08.07.1-k2	Yes[b]
RHEL 7.2	8.07.00.18.07.2-k	Yes[b]
RHEL 7.3	8.07.00.33.07.3-k	Yes[b]
RHEL 7.4	8.07.00.38.07.4-k1	Yes[b]
RHEL 7.5	9.00.00.00.07.5-k1	Yes[b]
RHEL 7.6	10.00.00.06.07.6-k	Yes[b]

Table 15 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
RHEL 7.7	10.00.00.12.07.7-k	Yes[b]
RHEL 7.8	10.01.00.20.07.8-k	Yes[b]
RHEL 8.0	10.00.00.07.08.0-k1	Yes[c]
RHEL 8.1	10.01.00.15.08.1-k1	Yes[c]
Linux OL 6.x [x86_64] UEK R1 [2.6.23-100]	8.03.07.03.32.1-k	Yes[a]
Linux OL 5.x [x86_64] UEK R1 [2.6.32-100]	8.03.01.02.32.1-k9	Yes[a]
Linux OL 5.x [x86_64] UEK R1 U1 [2.6.32-200] Linux OL 6.x [x86_64] UEK R1 U1 [2.6.32-200]	8.03.07.04.32.1-k	Yes[a]
Linux OL 5.x [x86_64] UEK R1 U2 [2.6.32-300] Linux OL 6.x [x86_64] UEK R1 U2 [2.6.32-300] Linux OL 5.x [x86_64] UEK R1 U3 [2.6.32-400] Linux OL 6.x [x86_64] UEK R1 U3 [2.6.32-400]	8.03.07.08.32.1-k	Yes[b]
Linux OL 5.x [x86_64] UEK R2 [2.6.39-100] Linux OL 6.x [x86_64] UEK R2 [2.6.39-100]	8.03.07.12.39.0-k	Yes[a]
Linux OL 5.x [x86_64] UEK R2 U1 [2.6.39-200] Linux OL 6.x [x86_64] UEK R2 U1 [2.6.39-200]	8.04.00.03.39.0-k	Yes[a]
Linux OL 5.x [x86_64] UEK R2 U2 [2.6.39-300] Linux OL 6.x [x86_64] UEK R2 U2 [2.6.39-300]	8.04.00.08.39.0-k	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U3 [2.6.39-400] Linux OL 6.x [x86_64] UEK R2 U3 [2.6.39-400]	8.04.00.11.39.0-k 8.05.00.03.39.0-k (This in kernel driver is only available on errata kernels equal to or greater than 2.6.39-400.109.1.el6uek)	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U4 [2.6.39-400.109] Linux OL 6.x [x86_64] UEK R2 U4 [2.6.39-400.109]	8.05.00.03.39.0-k	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U5 [2.6.39-400.209] Linux OL 6.x [x86_64] UEK R2 U5 [2.6.39-400.209]	8.05.00.03.39.0-k	Yes[b]
Linux OL 6.x [x86_64] UEK R3 [3.8.13-16]	8.05.00.03.39.0-k	Yes[b]

Table 15 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
Linux OL 6.x [x86_64] UEK R3 U1 [3.8.13-26]	8.06.00.14.39.0-k	Yes[b]
Linux OL 6.x [x86_64] UEK R3 U2 [3.8.13-35]	8.07.00.08.39.0-k1	Yes[b]
Linux OL 6.x [x86_64] UEK R3 U3 [3.8.13-44]	8.07.00.08.39.0-k1	Yes[b]
Linux OL 6.x [x86_64] UEK R3 U4 [3.8.13-55]	8.07.00.16.39.0-k	Yes[b]
Linux OL 7.x [x86_64] UEK R3 U2 [3.8.13-35] Linux OL 7.x [x86_64] UEK R3 U3 [3.8.13-44]	8.07.00.08.39.0-k1	Yes[b]
Linux OL 7.x [x86_64] UEK R3 U4 [3.8.13-55]	8.07.00.16.39.0-k	Yes[b]
Linux OL 7.x [x86_64] UEK R3 U5 [3.8.13-68]	8.07.00.16.39.0-k	Yes[b]
Oracle Linux 6.x UEK R3 U6 [3.8.13-98] Oracle Linux 7.x UEK R3 U6 [3.8.13-98]	5.04.00.07.06.02-uek3	Yes[b]
Oracle Linux 6.x UEK R3 U7 [3.8.13-118] Oracle Linux 7.x UEK R3 U7 [3.8.13-118]	5.04.00.07.06.02-uek3	Yes[b]
Oracle Linux 6.x UEK R4 [4.1.12] Oracle Linux 7.x UEK R4 [4.1.12]	5.04.00-k6	Yes[b]
Oracle Linux 6.x UEK R4 U1 [4.1.12-37] Oracle Linux 7.x UEK R4 U1 [4.1.12-37]	5.04.00-k6	Yes[b]
Oracle Linux 6.x UEK R4 U2 [4.1.12-61] Oracle Linux 7.x UEK R4 U2 [4.1.12-61]	5.04.00-k6	Yes[b]
Oracle Linux 6.x UEK R4 U4 [4.1.12-94] Oracle Linux 7.x UEK R4 U4 [4.1.12-94]	5.04.00-k6	Yes[b]
Oracle Linux 6.x UEK R4 U5 [4.1.12-103] Oracle Linux 7.x UEK R4 U5 [4.1.12-103]	5.04.00-k6	Yes[b]
Oracle Linux 6.x UEK R4 U6 [4.1.12-112] Oracle Linux 7.x UEK R4 U6 [4.1.12-112]	5.04.00-k6	Yes[b]
Oracle Linux 6.x UEK R4 U7 [4.1.12-124] Oracle Linux 7.x UEK R4 U7 [4.1.12-124]	5.04.00-k6	Yes[b]
Oracle Linux 7.x UEK R5 [4.1.35-1818]	5.04.00-k6	Yes[b]
Oracle Linux 7.x UEK R5 U1 [4.1.35-1844]	5.04.00-k6	Yes[b]
Oracle Linux 7.x UEK R5 U2 [4.1.35-1902]	5.04.00-k6	Yes[b]
SLES 10 SP4	8.03.01.12.10.3-k4	Yes[b]
SuSE SLES 11 GA	8.02.01.03.11.0-k9	Yes[a]

Table 15 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters	
		CNA	
SuSE SLES 11 SP1	8.03.01.06.11.1-k8	Yes[b]	
	8.03.01.07.11.1-k8		
	8.03.01.08.11.1-k8		
SuSE SLES 11 SP2	8.03.07.07-k	Yes[b]	
SuSE SLES 11 SP3	8.04.00.13.11.3-k	Yes[b]	
SuSE SLES 11 SP4	8.07.00.18-k	Yes[b]	
SuSE SLES 12	8.07.00.08.12.0-k	Yes[b]	
SLES 12 SP1	8.07.00.18-k	Yes[b]	
SLES 12 SP2	8.07.00.33-k	Yes[b]	
SLES 12 SP3	9.00.00.00-k	Yes[b]	
SLES 12 SP4	10.00.00.11-k	Yes[b]	
SLES 15	10.00.00.06-k	Yes[b]	
SLES 15 SP1	10.00.00.13-k	Yes[b]	
XenServer [x86_64] 6.0.2	8.03.07.03.55.6-k2	Yes[b]	
XenServer [x86_64] 6.1	8.04.00.02.55.6-k	Yes[b]	
XenServer [x86_64] 6.2	8.05.00.03.55.6-k	Yes[b]	
XenServer [x86_64] 6.5	8.07.00.09.66.5-k	Yes[b]	
XenServer [x86_64] 7.0	8.07.00.33.77.0-k	Yes[b]	
XenServer [x86_64] 7.1	8.07.00.41.77.1-k	Yes[b]	
XenServer [x86_64] 7.2	8.07.00.41.77.1-k	Yes[b]	
XenServer [x86_64] 7.3	8.07.00.56.71.0-k	Yes[b]	
XenServer [x86_64] 7.4	8.07.00.56.71.0-k	Yes[b]	

a. For models LP21xxx, OCe10xxx and OCe11xxx CNAs.
b. For models LP21xxx, OCe10xxx, OCe11xxx and OCe14xxx CNAs.
c. Some Emulex CNA is not support since RHEL8.0, refer RHEL8.0 release doc (https://access.redhat.com/documentation/en-us/red_hat_enterprise_linux/8/html-single/considerations_in_adopting_rhel_8/index#removed-adapters_hardware-enablement)

Table 16 Supported FCOE out-of-kernel drivers

Operating system	Driver Version	Supported Adapters	
		CNA	
SLES 10 SP4	8.04.00.15.10.3-k	Yes[a]	
	8.06.00.10.10.3-k		

Table 16 Supported FCOE out-of-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
SLES 11 SP2	8.05.00.03.11.1-k 8.06.00.10.11.1-k	Yes[a]
SLES 11 SP3	8.06.00.10.11.1-k	Yes[a]

a. For models QLE81xx and QLE82xx CNAs only.

Utility

QLogic provides a full suite of tools for storage and networking I/O manageability. Deploy Fibre Channel, Converged Networking, Intelligent ethernet, FabricCache, or virtualized I/O solutions using QConvergeConsole (QCC). The comprehensive graphical and command line user interfaces centralize I/O management of multiple generations of QLogic storage and network adapters across operating systems and protocols.

QConvergeConsole management suite includes a browser-based graphical user interface (GUI) and a lightweight command line interface (CLI) for Linux.

To download the application kit, go to the QLogic download page.

Update firmware and boot code

QLogic CNA firmware and boot code are all bundled in a .zip file. The files that are contained in the Flash image package are zipped into a file that will expand to provide the various versions for NIC PXE, FCOE BIOS, ISCSI BIOS, firmware, etc. See the package read1st.txt for package details.

The firmware and bootcode can be updated by using QConvergeConsole GUI/CLI Management tools. See the QConvergeConsole *User Guide* available at QLogic website.

You can also download the multi-boot firmware liveCD to update firmware and boot code.

QLogic 1000/1800 series

Using the QLogic BR Converged Network Adapter with the Linux operating system requires adapter driver software. The driver functions at a layer below the Linux SCSI driver to present Fibre Channel devices to the operating system as if they were standard SCSI devices.

Dell EMC supports both in-kernel and out-of-kernel drivers.

In-kernel driver versions are included by default in the kernel and do not require any installation. Out-of-kernel driver versions from the vendor need manual installation. See Table 17 for supported in-kernel driver versions. Table 18 lists the out-of-kernel driver versions that are supported with the corresponding operating system updates.

See the latest Dell EMC Simple Support Matrix for your specific Linux distribution, kernel version, and driver to determine if you need to proceed with the following out-of-kernel instructions.

If your installation requires an out-of-kernel driver, download it from the Dell EMC-approved section of the QLogic website.

Table 17 Supported FCOE in-kernel drivers

Operating system	Driver Version	Supported Adapters
		CNA
RHEL 5.5	2.1.2.0	Yes[a]
RHEL 5.6	2.1.2.0	Yes[b]
RHEL 5.7	2.3.2.5	Yes[b]
RHEL 5.8	3.0.2.2	Yes[b]
RHEL 5.9 RHEL 5.10 RHEL 5.11	3.0.23.0	Yes[b]
RHEL 6.0	2.1.2.1	Yes[b]
RHEL 6.1	2.3.2.3	Yes[b]
RHEL 6.2	3.0.2.2	Yes[b]
RHEL 6.3	3.0.2.2	Yes[b]
RHEL 6.4	3.0.23.0	Yes[b]
RHEL 6.5	3.2.21.1	Yes[b]
RHEL 6.6	3.2.23.0	Yes[b]
RHEL 6.7	3.2.23.0	Yes[b]
RHEL 7.0	3.2.23.0	Yes[b]
RHEL 7.1, 7.2	3.2.23.0	Yes[b]
RHEL 7.3	3.2.25.0	Yes[b]
RHEL 7.4, 7.5, 7.6, 7.7	3.2.25.1	Yes[b]
SLES 11	1.1.0.2	Yes[a]
SLES 11 SP1	2.1.2.1	Yes[a]
SLES 11 SP2	3.0.2.2	Yes[a]
SLES 11 SP3	3.1.2.1	Yes[b]
SLES 11 SP4	3.2.23.0	Yes[b]
SLES 12	3.2.23.0	Yes[b]
SLES 12 SP1	3.2.23.0	Yes[b]
SLES 12 SP2	3.2.25.1	Yes[b]
SLES 12 SP3	3.2.25.1	Yes[b]
SLES 12 SP4	3.2.25.1	Yes[b]
SLES 15	3.2.25.1	Yes[b]
SLES 15 SP1	3.2.25.1	Yes[b]

Table 17 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
Linux OL 6.x [x86_64] UEK R1 [2.6.23-100]	2.3.2.3	Yes[b]
Linux 6.x [x86_64] UEK R1 U2 [2.6.32-300]	3.0.2.2	Yes[b]
Linux OL 5.x [x86_64] UEK R2 [2.6.39-100] Linux OL 6.x [x86_64] UEK R2 [2.6.39-100]	3.0.2.2	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U1 [2.6.39-200] Linux OL 6.x [x86_64] UEK R2 U1 [2.6.39-200]	3.0.2.2	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U3 [2.6.39-400] Linux OL 6.x [x86_64] UEK R2 U3 [2.6.39-400]	3.0.2.2	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U4 [2.6.39-400.109] Linux OL 6.x [x86_64] UEK R2 U4 [2.6.39-400.109]	3.0.2.2	Yes[b]
Linux OL 5.x [x86_64] UEK R2 U5 [2.6.39-400.209] Linux OL 6.x [x86_64] UEK R2 U5 [2.6.39-400.209]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R3 [3.8.13-16] Oracle Linux 7.x UEK R3 [3.8.13-16]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R3 U1 [3.8.13-26] Oracle Linux 7.x UEK R3 U1[3.8.13-26]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R3 U2 [3.8.13-35] Oracle Linux 7.x UEK R3 U2 [3.8.13-35]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R3 U3 [3.8.13-44] Oracle Linux 7.x UEK R3 U3 [3.8.13-44]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R3 U4 [3.8.13-55] Oracle Linux 7.x UEK R3 U4 [3.8.13-55]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R3 U5 [3.8.13-68] Oracle Linux 7.x UEK R3 U5 [3.8.13-68]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R3 U6 [3.8.13-98] Oracle Linux 7.x UEK R3 U6 [3.8.13-98]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R3 U7 [3.8.13-118] Oracle Linux 7.x UEK R3 U7 [3.8.13-118]	3.0.2.2	Yes[b]
Oracle Linux 6.x UEK R4 [4.1.12]	3.2.25.1	Yes[b]

Table 17 Supported FCOE in-kernel drivers (continued)

Operating system	Driver Version	Supported Adapters
		CNA
Oracle Linux 7.x UEK R4 [4.1.12]		
Oracle Linux 6.x UEK R4 U1 [4.1.12-37] Oracle Linux 7.x UEK R4 U1 [4.1.12-37]	3.2.25.1	Yes[b]
Oracle Linux 6.x UEK R4 U2 [4.1.12-61] Oracle Linux 7.x UEK R4 U2 [4.1.12-61]	3.2.25.1	Yes[b]
Oracle Linux 6.x UEK R4 U4 [4.1.12-94] Oracle Linux 7.x UEK R4 U4 [4.1.12-94]	3.2.25.1	Yes[b]
Oracle Linux 6.x UEK R4 U5 [4.1.12-103] Oracle Linux 7.x UEK R4 U5 [4.1.12-103]	3.2.25.1	Yes[b]
Oracle Linux 6.x UEK R4 U6 [4.1.12-112] Oracle Linux 7.x UEK R4 U6 [4.1.12-112]	3.2.25.1	Yes[b]
Oracle Linux 6.x UEK R4 U7 [4.1.12-124] Oracle Linux 7.x UEK R4 U7 [4.1.12-124]	3.2.25.1	Yes[b]
Oracle Linux 7.x UEK R5 [4.1.35-1818]	3.2.25.1	Yes[b]
Oracle Linux 7.x UEK R5 U1 [4.1.35-1844]	3.2.25.1	Yes[b]
Oracle Linux 7.x UEK R5 U2 [4.1.35-1902]	3.2.25.1	Yes[b]
XenServer [x86_64] 6.0.2	2.3.2.7	Yes[b]
XenServer [x86_64] 6.1	3.1.0.0	Yes[b]
XenServer [x86_64] 6.2	3.2.1.1	Yes[b]
XenServer [x86_64] 6.5	3.2.1.1	Yes[b]
XenServer [x86_64] 7.0	3.2.1.1	Yes[b]
XenServer [x86_64] 7.1	3.2.23.0	Yes[b]
XenServer [x86_64] 7.2	3.2.25.1	Yes[b]
XenServer [x86_64] 7.3	3.2.25.1	Yes[b]
XenServer [x86_64] 7.4	3.2.25.1	Yes[b]

a. For models EM1010/1020 only.
b. For models EM1010/1020, EM-BR1860.

Table 18 Supported FCOE out-of-kernel drivers

Operating system	Driver Version	Supported Adapters
		CNA
RHEL 5.8 -5.9	3.2.3.0	Yes[a]
RHEL 6.3-6.4	3.2.3.0	Yes[a]
SLES10 SP3-SP4	3.2.3.0	Yes[a]
SLES 11 SP2-SP3	3.2.3.0	Yes[a]

a. For models EM1010/1020, EM-BR1860.

Utility

QLogic Host Connectivity Manager (GUI) is a management software application for configuring, monitoring, and troubleshooting QLogic Host Bus Adapter (HBAs), Converged Network Adapters (CNAs), and Fabric Adapters in a storage area network (SAN) environment.

The management software has two components:

- The agent, which runs on the host

- The management console, which is the graphical user interface client used to manage the adapter QLogic

You can manage the software on the host or remotely from another host. The communication between the management console and the agent is managed using JSON-RPC over HTTPS.

You can also use QLogic BCU CLI to manage your adapters locally.

The QLogic Host Connectivity Manager (GUI) and QLogic BCU CLI are available for download from the QLogic website.

Update firmware and boot code

The Brocade CNA firmware and bootcode are provided in a multi-boot firmware image, which is available from the QLogic website.

The image can be flashed using either the BCU CLI or Host Connectivity Manager on systems. See *BR-Series Adapters Administrator's Guide* available on QLogic website.

You can also download the multi-boot firmware liveCD to update firmware and boot code.

Cisco Unified Computing System

The Cisco Unified Computing System (UCS) is a next-generation data center platform that unites compute, network, storage access, and virtualization into a single system configuration. As shown in the following figure, configurations consist of a familiar chassis and blade server combination that works with Cisco's Fabric Interconnect switches to attach to NPIV-enabled fabrics. This allows for a centralized solution combining high-speed server blades that are designed for virtualization, FCoE connectivity, and centralized management.

Figure 8 Cisco Unified Computing System example
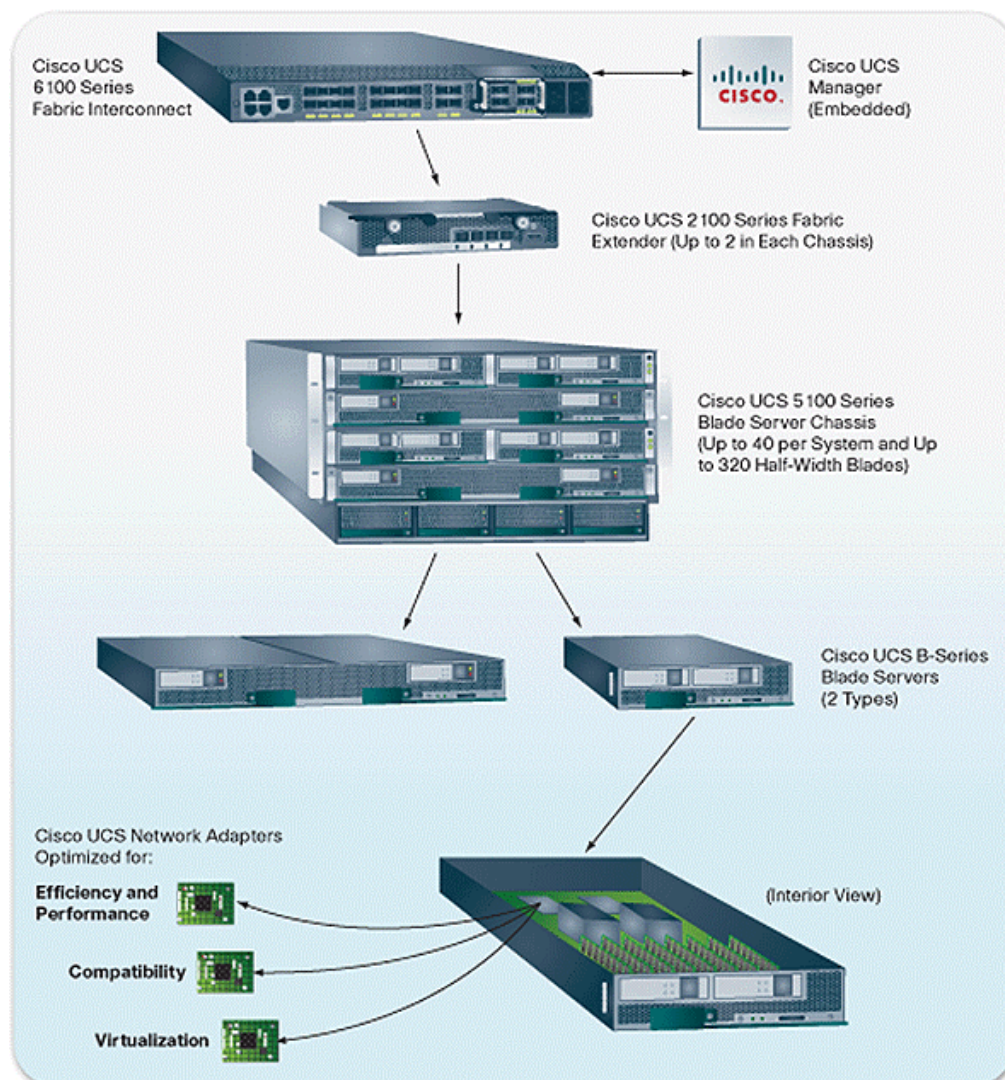
Fibre Channel ports on Fabric Interconnect switches must be configured as NP ports, which requires the connected Fabric switch to be NPIV-capable. See the latest Dell EMC Simple Support Matrix for currently supported switch configurations.

In each server blade, an Emulex or QLogic-based converged network adapter (CNA) mezzanine card is used to provide Ethernet and Fibre Channel connectivity for that blade to an attached network or SAN. These CNAs are based on currently supported PCI Express CNAs that Dell EMC supports in standard servers and use the same drivers, firmware, and BIOS to provide connectivity to both Dell EMC Fibre Channel and iSCSI storage array ports through the UCS Fabric Extenders and Fabric Interconnect switches that provide both 10 Gb Ethernet and/or Fibre Channel.

In-depth information about UCS and how it utilizes FCoE technology for its blade servers can be found in Cisco UCS documentation on the Cisco website.

The UCS Fabric Interconnect switches are supported with the same supported configurations as the Cisco NEX-5020. For information about supported features and topologies, see *Fibre Channel over Ethernet (FCoE) and Data Center Bridging (DCB) Case Studies TechBook* on Dell EMC E-Lab Interoperability Navigator at the **Resource Center** tab.

iSCSI Connectivity

Introduction

iSCSI (Internet Small Computer System Interface) is an IP-based storage networking standard for linking data storage facilities that are developed by the Internet Engineering Task Force. By transmitting SCSI commands over IP networks, iSCSI can facilitate block-level transfers over an IP network.

The iSCSI architecture is similar to a client/server architecture. In this case, the client is an initiator that issues an I/O request and the server is a target (such as a device in a storage system). This architecture can be used over IP networks to provide distance extension. This can be implemented between routers, host-to-switch, and storage array-to-storage array to provide asynchronous/synchronous data transfer.

iSCSI initiators come in two varieties: software and hardware. A software initiator is an operating system driver that is used along with an Ethernet card. The iSCSI driver handles all requests by determining if the packet is an Ethernet packet that is then passed to the network stack, or if it is an iSCSI packet that will then have the SCSI packet stripped out and pasted to the SCSI stack. Using a software initiator requires more CPU and memory resources on the host. A hardware initiator is an HBA which offloads some or all the iSCSI packet processing, which saves CPU and memory resources. These adapters will reduce the load of the iSCSI stack in the operating system.

iSCSI discovery

For an iSCSI initiator to establish an iSCSI session with an iSCSI target, the initiator needs the IP address, TCP port number, and iSCSI target name information. The goals of iSCSI discovery mechanisms are to provide low overhead support for small iSCSI setups and scalable discovery solutions for large enterprise setups.

Send target-An initiator may log in to an iSCSI target with session type of discovery and request a list of target WWUIs through a separate `SendTargets` command. All iSCSI targets are required to support the `SendTargets` command.

iSNS

The iSNS protocol is designed to facilitate the automated discovery, management, and configuration of iSCSI and Fibre Channel devices on a TCP/IP network. iSNS provides intelligent storage discovery and management services comparable to those found in Fibre Channel networks, allowing a commodity IP network to function in a similar capacity as a storage area network. iSNS also facilitates a seamless integration of IP and Fibre Channel networks, due to its ability to emulate Fibre Channel fabric services, and manage both iSCSI and Fibre Channel devices. iSNS thereby provides value in any storage network that is comprised of iSCSI devices, Fibre Channel devices, or any other combination.

SLP

iSCSI targets are registered with SLP as a set of service URLs, one for each address on which the target may be accessed. Initiators discover these targets using SLP service requests. Targets that do not directly support SLP, or are under the control of a management service, may be registered by a proxy service agent as part of the software providing this service.

Digests

Digests enable the checking of end-to-end, non-cryptographic data integrity beyond the integrity checks provided by the link layers and they cover the entire communication path including all elements that may change the network level PDUs such as routers, switches, and proxies.

Optional header and data digests protect the integrity of the header and data, respectively. The digests, if present, are located after the header and PDU-specific data and cover the header and the PDU data, each including the padding bytes, if any. The existence and type of digests are negotiated during the Login phase. The separation of the header and data digests is useful in iSCSI routing applications, where only the header changes when a message is forwarded. In this case, only the header digest should be recalculated.

(i) **Note:** Only header digests are currently supported in Linux iSCSI.

iSCSI error recovery

iSCSI supports three levels of error recovery, 0, 1, and 2:

- Error recovery level 0 implies session level recovery.

- Error recovery level 1 implies level 0 capabilities and digest failure recovery.

- Error recovery level 2 implies level 1 capabilities and connection recovery.

(i) **Note:** The Linux iSCSI implementation only supports Error level 0 recovery.

The most basic kind of recovery is called *session* recovery. In session recovery, whenever any kind of error is detected, the entire iSCSI session is terminated. All TCP connections connecting the initiator to the target are closed, and all pending SCSI commands are completed with an appropriate error status. A new iSCSI session is then established between the initiator and target, with new TCP connections.

Digest failure recovery starts if the iSCSI driver detects that data arrived with an invalid data digest and that data packet must be rejected. The command corresponding to the corrupted data can then be completed with an appropriate error indication.

Connection recovery can be used when a TCP connection is broken. Upon detection of a broken TCP connection, the iSCSI driver can either immediately complete the pending command with an appropriate error indication, or can attempt to transfer the SCSI command to another TCP connection. If necessary, the iSCSI initiator driver can establish another TCP connection to the target, and the iSCSI initiator driver can inform the target the change in allegiance of the SCSI command to another TCP connection.

iSCSI security

Historically, native storage systems have not had to consider security because their environments offered minimal security risks. These environments consisted of storage devices either directly attached to hosts or connected through a storage area network (SAN) distinctly separate from the communications network. The use of storage protocols, such as SCSI over IP-networks, requires that security concerns be addressed. iSCSI implementations must provide means of protection against active attacks(such as, pretending to be another identity, message insertion, deletion, modification, and replaying) and passive attacks (such as, eavesdropping, gaining advantage by analyzing the data sent over the line). Although technically possible, iSCSI should not be configured without security. iSCSI configured without security should be confined, in extreme cases, to closed environments without any security risk.

Security mechanisms

The entities that are involved in iSCSI security are the initiator, target, and IP communication end points. iSCSI scenarios in which multiple initiators or targets share a single communication end

points are expected. To accommodate such scenarios, iSCSI uses two separate security mechanisms:

- In-band authentication between the initiator and the target at the iSCSI connection level(carried out by exchange of iSCSI Login PDUs).

- Packet protection (integrity, authentication, and confidentiality) by IPsec at the IP level.

The two security mechanisms complement each other. The in-band authentication provides end-to-end trust (at login time) between the iSCSI initiator and the target while IPsec provides a secure channel between the IP communication end points.

Authentication method

The following authentication method is supported with Linux:

Challenge Handshake Authentication Protocol (CHAP) - CHAP is used to periodically verify the identity of the peer using a three-way handshake. This is done upon establishing initial link and may be repeated anytime after the link has been established.

CHAP provides protection against playback attack by the peer by using an incrementally changing identifier and a variable challenge value. The use of repeated challenges is intended to limit the time of exposure to any single attack. The authenticator is in control of the frequency and timing of the challenges. This authentication method depends upon a secret that is known only to the authenticator and that peer. The secret is not sent over the link.

iSCSI solutions

This section contains the following information:

- General best practices
- General supported configurations
- Dell EMC native iSCSI targets
- Native Linux iSCSI driver
- Software and hardware iSCSI initiator

General best practices

This section lists general best practices.

Network design

- The network should be dedicated solely to the IP technology being used, and no other traffic should be carried over it.

- The network must be a well-engineered network with no packet loss or packet duplication, if possible. This would lead to retransmission, which will affect overall system performance.

- While planning the network, ensure that the utilized throughput will never exceed the available bandwidth. Oversubscribing available bandwidth leads to network congestion, which might cause dropped packets and lead to a TCP slow start.

- Network congestion must be considered between switches as well as between the switch and the end device.

Header and data digest

Header digest is recommended when using a routed network (Layer 3) or when using Layer 2 network with VLAN tagging. Digests are not mandatory in a plain LAN (other than those mentioned above).

The Linux iSCSI software doesn't support Data Digest. Data Digest can also be responsible for a severe impact of overall system performance.

General supported configurations

This section lists general supported configurations.

Direct connect-Direct connect between an iSCSI-enabled host and storage system is the most basic configuration. This configuration allows proof of concept for pre-rollout testing and simple secure connectivity using iSCSI.

Figure 9 Direct connect example

iSCSI target   iSCSI Initiator

Network connect-Storage array iSCSI ports are connected to an IP network. The network can either be a local area network or a routed network. iSCSI servers are connected to the network as well and communicate with the storage array iSCSI ports. All standard network equipment is supported. However, it is recommended using enterprise class equipment because they provide more memory to address network congestion.

Figure 10 Network connect example

iSCSI target   TCP/IP network   iSCSI Initiator
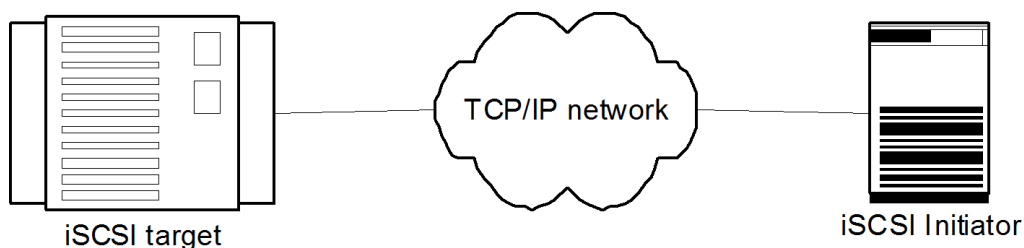
Dell EMC native iSCSI targets

This section discusses the following Dell EMC native iSCSI targets:

- PowerMAX and VMAX series
- VNX series iSCSI/FC system
- VNXe/Unity Series
- XtremIO
- VNX
- PowerStore

PowerMAX and VMAX series

The iSCSI channel director supports iSCSI channel connectivity to IP networks and to iSCSI-capable open systems server systems for block storage transfer between hosts and storage. The primary applications are storage consolidation and host extension for stranded servers and departmental work groups. The PowerMAX and VMAX series iSCSI director provides up to four 1 Gb/s Ethernet ports or up to two 10 Gb/s Ethernet ports, both with LC connectors. The iSCSI directors support the iSNS protocol. Authentication mechanism is Challenge Handshake Authentication Protocol (CHAP).

LUNs are configured in the same manner as for Fibre Channel directors and are assigned to the iSCSI ports. LUN masking is available.

References

For configuration of a PowerMAX and VMAX series iSCSI target, see the PowerMAX and VMAX configuration guides. For latest iSCSI host support, see Dell EMC E-Lab Navigator.

For configuration of an iSCSI server, see the applicable host connectivity guide.

VNX series iSCSI/FC system

VNX series systems can be configured to support iSCSI and Fibre Channel connectivity simultaneously. However, the Linux host can only be connected to the same VNX series system using either FC or iSCSI connectivity.

- iSCSI ports on the array can be 1 Gb/s or 10 Gb/s Ethernet mechanism is ports.
- iSNS protocol is supported. The authentication mechanism is CHAP.
- LUNs are configured in the same manner as for Fibre Channel array and are assigned to a storage group.

VNXe/Unity Series

This can be configured as an iSCSI array or as a Fibre Channel array. All iSCSI ports on the array are 10 Gb/s Ethernet ports. The iSNS protocol is supported. The authentication mechanism is CHAP.

LUNs are configured in the same manner as for Fibre Channel array and are assigned to a storage group.

References

For configuration of VNXe or Unity series targets, see the appropriate configuration guide.

For up-to-date iSCSI host support, see Dell EMC Simple Support Matrix, available on Dell EMC E-Lab Navigator.

XtremIO

XtremIO can be configured to support iSCSI and FC connectivity simultaneously.

LUNs are configured in the same manner as for FC array and are assigned to a storage group. See the Server/OS vendor support matrix for 10 Gb iSCSI compatibility.

References

For configuration of XtremIO targets, see the appropriate configuration guide.

For up-to-date iSCSI host support, see Dell EMC E-Lab Navigator. For configuration of the iSCSI server, see the appropriate host connectivity guide.

When iSCSI is used with XtremIO, the iscsi.conf file is used to overwrite iSCSI specific multipathing related settings.

Parameter	Value	Description
replacement_timeout	120	Specifies the number of seconds the iSCSI layer holds for a timed-out path/ session to reestablish before failing any commands on that path/session. The default value is 120.
FirtsBurstLength	<at least block size used>	Specifies the maximum amount of unsolicited data (in bytes) an iSCSI initiator can send to the target during the execution of a single SCSI command. Adjust this parameter when the block size used is larger than the

Parameter	Value	Description
		default setting for this parameter (256 KB).

Using these settings prevents commands from being split by the iSCSI initiator and enables instantaneous mapping from the host to the volume. To apply the adjusted iscsi.conf settings, run the following command on the Linux host: `service iscsi restart`.

VNX

This can be configured to support iSCSI and FC connectivity simultaneously; however, no Linux host is accepted to connect to the same VNX series using both FC and iSCSI connectivity. The host must be either iSCSI or FC.

iSCSI ports on the array can be 1 Gb/s or 10 Gb/s Ethernet ports. iSNS protocol is supported. The authentication mechanism is CHAP.

LUNs are configured in the same manner as for Fibre Channel array and are assigned to a storage group.

References

For configuration of VNX iSCSI targets, see the appropriate configuration guide. For up-to-date iSCSI host support, see Dell EMC E-Lab Navigator.

For configuration of an iSCSI server, see the appropriate host connectivity guide.

PowerStore

PowerStore can be configured to support iSCSI and FC connectivity simultaneously. PowerStore support both 25 Gb and 10 Gb speed target ports. However, the current PowerStore arrays only support one iSCSI subnet in whole system, which needs Linux host run specific configuration for iSCSI path redundancy. The iSNS protocol is supported. The authentication mechanism is CHAP. LUNs are configured in the same manner as for Fibre Channel array and are assigned to a storage group.

References

For configuration of PowerStore series targets, see the appropriate configuration guide. For up-to-date iSCSI host support, see *Dell EMC Simple Support Matrix*, available on Dell EMC E-Lab Navigator.

Native Linux iSCSI driver

This driver is found in the Linux kernel. The iSCSI driver provides a host with the ability to access storage through an IP network. The driver uses the iSCSI protocol to transport SCSI requests and responses over an IP network between the host and an iSCSI target device. The iSCSI protocol is an IETF-defined protocol for IP storage.

For more information about the IETF standards, see the Internet Engineering Task Forces (ITEF) website.

For more information about the iSCSI protocol, see RFC 3720.

Architecturally, the iSCSI driver combines with the host's TCP/IP stack, network drivers, and network interface card (NIC) to provide the same functions as a SCSI or a Fibre Channel(FC) adapter driver with a host bus adapter (HBA).The iSCSI driver provides a transport for SCSI requests and responses to storage devices over an IP network instead of using a directly attached SCSI bus channel or an FC connection. The storage router, in turn, transports these SCSI requests and responses received over the IP network between it and the storage devices that are attached to it.

For detailed information regarding the native Linux iSCSI driver, see Native Linux iSCSI Attach .

Software and hardware iSCSI initiator

A software initiator is a driver that handles all iSCSI traffic. The iSCSI driver pairs the network interface and the SCSI disk driver to transport the data. Any system with an Ethernet card can act as an iSCSI initiator if supported by the operating system and server vendor.

A hardware initiator is an iSCSI HBA. iSCSI HBAs are available from several vendors.

iSCSI HBAs provide PCI connectivity to SCSI devices using the iSCSI protocol. iSCSI enables the use of IP-based SANs, which are similar to Fibre Channel SANs. An iSCSI offload engine, or iSOE card, offers an alternative to a full iSCSI HBA. An iSOE offloads the iSCSI initiator operations for this particular network interface from the host processor, freeing up CPU cycles for the main host applications. iSCSI HBAs or iSOEs are used when the additional performance enhancement justifies the additional expense of using an HBA for iSCSI, rather than using a software-based iSCSI client (initiator).

Converged network adapters or CNAs, offer a class of network devices that provide support for iSCSI over Ethernet, allowing hardware to offload networking and storage connectivity across standard Ethernet protocol.

For HBA and CNA support matrix, see Dell EMC E-Lab Navigator.

Native Linux iSCSI Attach

This section provides information about Native Linux iSCSI Attach, including the following:

- Open-iscsi driver
- Installing the open-iscsi driver
- Setting initiator name in software iSCSI
- Selective targets login
- Starting and stopping the iSCSI driver
- Dynamic LUN discovery
- Mounting and unmounting iSCSI file systems automatically (RHEL, Asianux, and SLES)
- Excessive dropped session messages found in /var/log/messages

open-iSCSI driver

This section discusses the following open-iSCSI driver information:

- open-iscsi driver introduction
- Features
- README file and main pages
- Environment and system requirements
- Installing the open-iscsi driver

open-iSCSI driver introduction

This driver is found from RHEL 5 to RHEL 8, SLES 10 to SLES 15, and Asianux 3 and 4. The open-iSCSI driver is a high-performance, transport independent, multi-platform implementation of RFC3720 iSCSI.

Open-iscsi is partitioned into user and kernel parts.

The kernel portion of open-iscsi is from code licensed under GPL. The kernel part implements iSCSI data path (that is, iSCSI Read and iSCSI Write), and consists of three loadable modules:

- scsi_transport_iscsi.ko

- libiscsi.ko
- iscsi_tcp.ko

User space contains the entire control layer:

- Configuration manager
- iSCSI discovery
- Login and logout processing
- connection-level error processing
- Nop-In and Nop-Out handling

The user space open-iscsi consists of a daemon process called *iscsid* and a management utility called *iscsiadm*. See the main pages for the complete usage of iscsiadm().

The iSCSI driver creates a table of available target devices. After the table is completed, the host can use the iSCSI targets by using the same commands and utilities used by a storage device that is directly attached (for example, via a SCSI bus).

Features

The following major features are supported by the iSCSI driver:

- Highly optimized and very small-footprint data path
- Persistent configuration database
- SendTargets discovery
- CHAP
- PDU header digest
- Multiple sessions

A more detailed description of each of these features is described in the README file for iSCSI in your Red Hat (RHEL), Asianux, or SuSE (SLES) distribution and in the associated main pages.

For the most recent list of features, see the Open-iSCSI website.

README file and main pages

For RHEL 5, 6, 7 and 8, see the following:

- `/usr/share/doc/iscsi-initiator-utils-x.x.x/README`
- Main pages for:
  - *iscsid*
  - *iscsiadm*

For SLES 10, 11, 12 and 15, see the following:

- `/usr/share/doc/packages/open-iscsi/README`
- Main pages for:
  - *iscsid*
  - *iscsiadm*

For Asianux 3 and 4, see the following:

- `/usr/share/doc/iscsi-initiator-utils-x.x.x/README`
- Main pages for:
  - *iscsid*
  - *iscsiadm*

> (i) **Note:** Follow the configuration guidelines provided by Dell EMC outlines. Using improper
> setting scan cause erratic behavior. In particular, note the following:
>
> - The Linux iSCSI driver, which is part of the Linux operating system, does not distinguish
>   between NICs on the same subnet; therefore, to achieve load balancing and multipath
>   failover, storage systems connected to Linux servers must configure each NIC on a
>   different subnet.
>
> - The Linux server cannot be connected to the same storage system using both FC and
>   iSCSI connectivity. The host must be either iSCSI or FC.
>
> - iSCSI and NIC teaming/bonding is not supported simultaneously. There is no capability in
>   the drivers, at this time, to distinguish the paths for failover.

Figure 11 shows an example of Linux iSCSI in the preferred high availability configuration with
multiple NICs/HBAs to multiple subnets. This configuration illustrates a tested and supported
configuration that provides high availability for NIC/HBA failures, as well as single-path events
between the switches and the NICs/HBAs.

Figure 11 Linux iSCSI with multiple NICs/HBAs to multiple subnets example
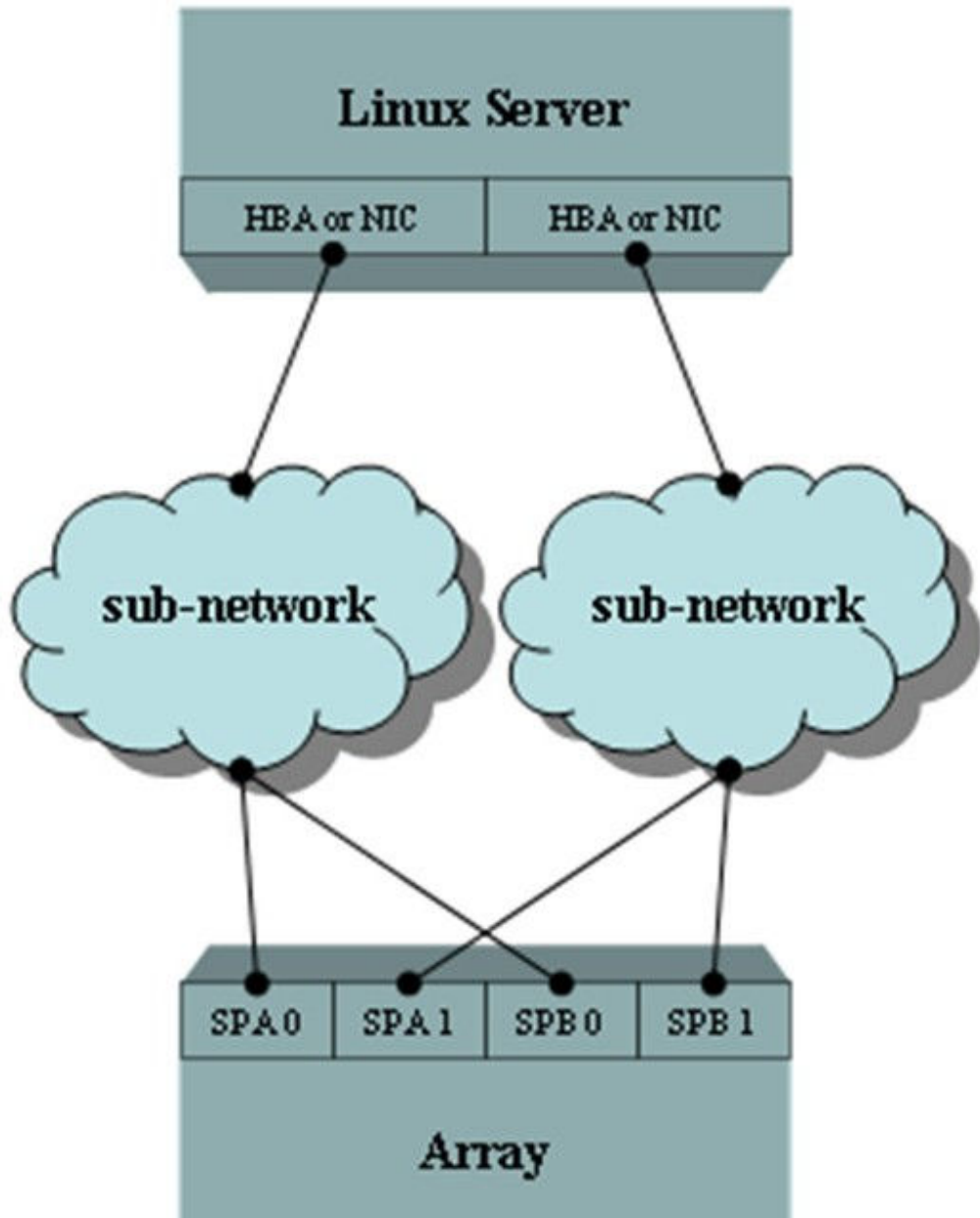
Figure 12 shows an example of Linux iSCSI in a supported configuration with a single NIC/HBA to a single subnet. This configuration does not provide high availability service, but does offer minimal high availability at the array end from events such as a service processor failure. However, for a trespass to occur on the array, a multipath drive must be installed and configured for the array. PowerPath and Linux native DM-MPIO are examples of these drivers.

Figure 12 Linux iSCSI with single NIC/HBA to single subnet example
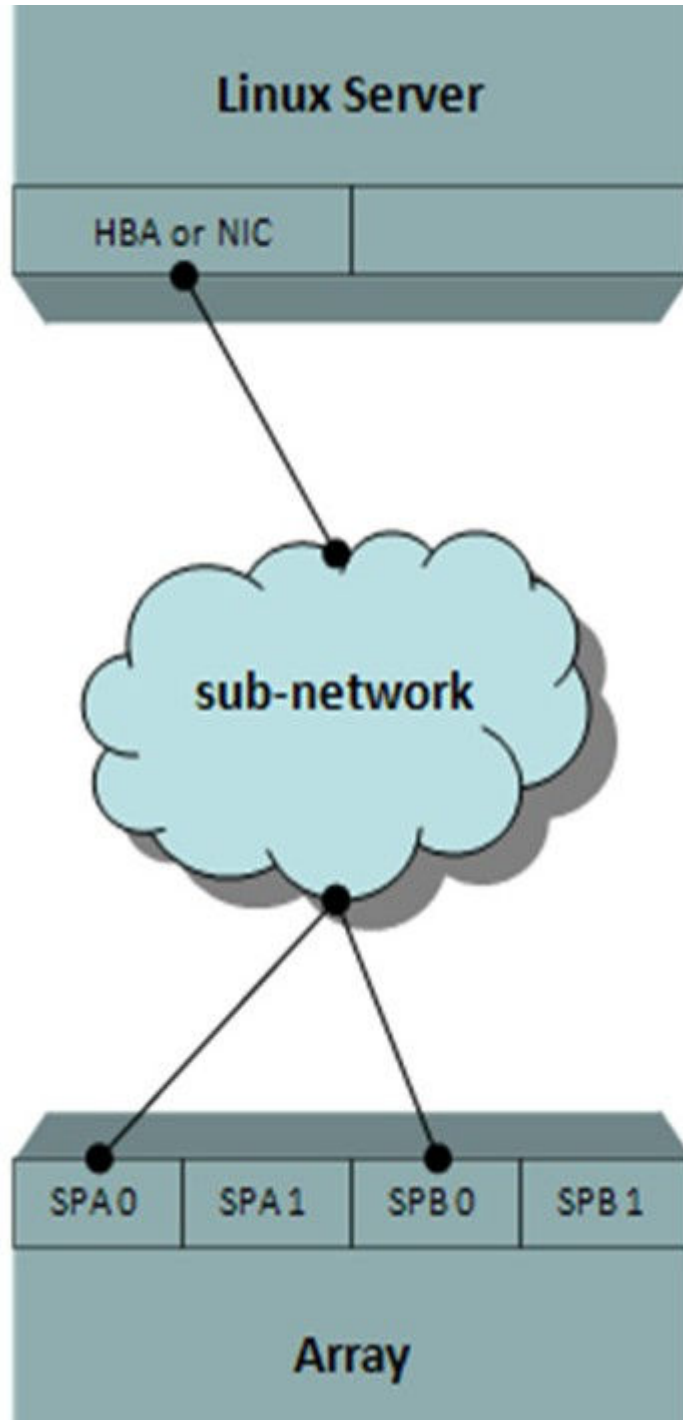
Environment and system requirements

See Dell EMC Simple Support Matrix, the applicable operating system distribution hardware compatibility list, and release notes for applicable multipath software.

Installing the open-iscsi driver

Before you begin

To complete the installation of the Linux iSCSI software initiator, consult the README files available within your Linux distribution and the Release Notes from the vendor.

(i) **Note:** Complete the following steps before continuing on to the SLES 10 installation subsections.

About this task

open-iscsi persistent configuration is implemented as a DBM database available on all Linux installations. The database contains the following two tables:

- Discovery table (discovery.db)

- Node table (node.db)

The iSCSI database files in SLES 10 GA are located in /var/lib/open-iscsi/. For SLES 10 SP1, they are located in `/etc/iscsi/`.

It is recommended to complete the following generic steps before proceeding to SLES 10 specific installation instruction. These recommendations are generic to all distributions unless otherwise noted.

Procedure

1. Edit the `/etc/iscsi/iscsid.conf` file.

   There are several variables within the file. The default file from the initial installation is configured to operate with the default settings. The syntax of the file uses a # symbol to comment out a line in the configuration file. You can enable a variable that is listed in the following table by deleting the # symbol preceding the variable in the `iscsid.conf` file. The entire set of variables is listed in each distribution's README file with both the default and optional settings; and, the configuration file itself.

   Table 19 RHEL 5/6/7/8, Asianux 4, SLES 10, SLES 11/12 (Linux 2.6 kernel and up)

Variable name	Default settings	Dell EMC recommended settings
node.startup	Asianux and SLES: Manual Red hat: Automatic	Auto
node.session.iscsi.InitialR2T	No	Yes
node.session.iscsi.ImmediateData	Yes	No
node.session.timeo.replacement_time o ut	120	60
node.conn[0].timoe.noop_out_interval	5	Higher in congested networks[a]
node.conn[0].timeo.noop_out_timeou t	5	Higher in congested networks[b]

   a. If you use of multipathing software, such as PowerPath or the native Linux DM-MPIO, you can decrease this time to 30 seconds for a faster failover. However, ensure that this timer is greater than the node.conn[0].timoe.noop_out_interval and node.conn[0].timeo.noop_out_timeout times combined.
   b. Ensure that this value does not exceed that of the value in node.session.timeo.replacement_timeout.

2. Set the following run levels for the iSCSI daemon to automatically start at boot and to shut down when the server is brought down:

- RHEL 5, RHEL 6 and Asianux 3.0:
  ```
  # chkconfig iscsid on
  ```

- SLES 10 and SLES 11:
  ```
  # chkconfig -s open-iscsion
  ```

- RHEL 7 and SLES 12:
  ```
  # systemctl enable iscsid # systemctl enable iscsi
  ```

3. See SLES 10, SLES 11, and SLES 12 to complete your installation on a SLES server or RHEL 5, 6 and 7 and Asianux 3, 4 to complete your installation on RHEL 5 and Asianux 3.0.

README file and main pages

It is recommended to use the YaST utility (another setup tool) on SLES to configure the iSCSI software initiator. It can be used to discover targets with the iSCSI SendTargets command, by adding targets to be connected to the server, and starting/stopping the iSCSI service.

Procedure

1. Open YaST and select **NetworkServices** > **iSCSI Initiator**. Open the **Discovered Targets** tab by entering the IP address of the target.

   With an XtremIO, Unity, or VNX series you only need to enter one of the target IP addresses and the array will return all its available targets for you to select. However, for PowerMAX and VMAX arrays you must enter each individual target you want to discover.

2. After discovering the targets, select the **Connected Targets** tab to log in to the targets you want to be connected to and select those you want to be log in to automatically at boot time.

RHEL 5, 6, 7 and 8 and Asianux 4

Before you begin

For RHEL 5, 6, 7 and 8, or Asianux 4, you must perform a series of iscsiadm(8) commands to configure the targets you want to connect to with open-iscsi. See the main pages for iscsiadm for a detailed explanation of the command and its syntax.

Procedure

1. Discover the targets that you want to connect your server to using iSCSI.

   For XtremIO, Unity, or VNX series you only need to perform a discovery on a single IP address and the array will return all its iSCSI configured targets. However, for the PowerMAX and VMAX arrays you must perform the discovery process on each individual target.

2. Before you perform the discovery of each port on the PowerMAX and VMAX series, configure the targeted PowerMAX and VMAX series iSCSI ports to accept the IQN of your initiator.

   See the appropriate PowerMAXand VMAX series installation and administration documentation available on Dell EMC Online Support.

   (i) Note: If you do not configure the initiator on a PowerMAX and VMAX series array first, SendTargets may encounter an iSCSI login failure.

Setting initiator name in software iSCSI

iSCSI Qualified Name (IQN) is a commonly used name format for an initiator and target in an iSCSI environment. Use the following format: iqn.yyyy-mm.{reversed domain name} e.g., iqn.1987-05.com.cisco:host.fileserver.xyz

(i) **Note:** The text after the colon, also known as an alias, will help to organize or label resources.

For more details, see iSCSI protocol specification (RFC 3720).

This is an example of the default initiator name that comes with the open-iscsi package: `iqn.1987-05.com.cisco:01.bca6fa632f97`.

Use the following command to change the default initiator name for the open-iscsi driver,-edit/etc/iscsi/initiatorname.conf.

`InitiatorName= iqn.2014-01.com.example:node1.target1`

For details and the location of the configuration files that are packaged with -open-iscsi, see the operating system vendor-specific documentation.

Selective targets login

- For open-iscsi, you can discover the targets using one of the array target IPs (for example, VNX array target IP: 11.10.10.20):

```
# iscsiadm -m discovery -t sendtargets -p 11.10.10.20
11.10.10.20:3260,1 iqn.1992-04.com.emc:cx.ck200073700372.a0
11.10.10.21:3260,2 iqn.1992-04.com.emc:cx.ck200073700372.a1
11.10.10.22:3260,3 iqn.1992-04.com.emc:cx.ck200073700372.a2
11.10.10.23:3260,4 iqn.1992-04.com.emc:cx.ck200073700372.a3
11.10.10.24:3260,5 iqn.1992-04.com.emc:cx.ck200073700372.b0
11.10.10.25:3260,6 iqn.1992-04.com.emc:cx.ck200073700372.b1
11.10.10.26:3260,7 iqn.1992-04.com.emc:cx.ck200073700372.b2
11.10.10.27:3260,8 iqn.1992-04.com.emc:cx.ck200073700372.b3
```

- Log in to selected targets a0/a1, b0/b1:

```
# iscsiadm -m node -l, or
# iscsiadm -m node --login
```

- Log in to selected targets a0/a1, b0/b1:

```
#iscsiadm -m node -T
iqn.1992-04.com.emc:cx.ck200073700372.a0 -p 11.10.10.20 -l
#iscsiadm -m node -T
iqn.1992-04.com.emc:cx.ck200073700372.a1 -p 12.10.10.21 -l
#iscsiadm -m node -T
iqn.1992-04.com.emc:cx.ck200073700372.b0 -p 11.10.10.24 -l
#iscsiadm -m node -T
iqn.1992-04.com.emc:cx.ck200073700372.b1 -p 12.10.10.25 -l
```

- Log out all targets:

```
# iscsiadm -m node -u, or

# iscsiadm -m node --logout
```

- **Log out selected targets a0/a1, b0/b1:**
```
#iscsiadm -m node -T

iqn.1992-04.com.emc:cx.ck200073700372.a0 -p 11.10.10.20 -u

#iscsiadm -m node -T

iqn.1992-04.com.emc:cx.ck200073700372.a1 -p 12.10.10.21 -u

#iscsiadm -m node -T

iqn.1992-04.com.emc:cx.ck200073700372.b0 -p 11.10.10.24 -u

#iscsiadm -m node -T

iqn.1992-04.com.emc:cx.ck200073700372.b1 -p 12.10.10.25 -u
```

Starting and stopping the iSCSI driver

- To manually start the iSCSI driver for RHEL 5, RHEL 6 and Asianux, type:
```
# service iscsid force-start # /etc/init.d/iscsid start
```
- To manually start the iSCSI driver for SLES 10 and SLES 11 type:
```
# /etc/init.d/open_iscsi start
```

If there are problems loading the iSCSI kernel module, diagnostic information is placed in /var/log/iscsi.log.

The open_iscsi driver is a sysfs class driver. You can access many of its attributes in the /sys/class/iscsi_<host, session, connection> directory.

See the main page for iscsiadm(8) for all administrative functions used to configure, gather statistics, target discovery, and so on.

- To manually stop the iSCSI driver for RHEL 5 and RHEL 6, type:
```
# /etc/init.d/iscsid stop
```
- To manually start the iSCSI driver for SLES 10 and SLES 11, type:
```
# /etc/init.d/open_iscsi stop
```

You must ensure that anything that has an iSCSI device open has closed the iSCSI device before shutting down iSCSI. This includes file systems, volume managers, and user applications.

If iSCSI devices are open when you attempt to stop the driver, the scripts display an error instead of removing those devices. This prevents you from corrupting the data on iSCSI devices. In this case, iscsid will no longer be running. If you want to continue using the iSCSI devices, it is recommended that you type /etc/init.d/iscsi start.

- To manually start the iSCSI driver on RHEL 7/8 and SLES 12/15, type:
```
# systemctl start iscsid # systemctl start iscsi
```
- To manually stop the iSCSI driver, type:
```
# systemctl stop iscsid
```

Dynamic LUN discovery

The iSCSI driver can dynamically discover target storage systems on the Ethernet; however, it cannot dynamically discover LUNs on the storage systems. The Linux operating system's SCSI mid-layer performs LUN discovery. Because of this, as with Fibre Channel, you must rescan the SCSI bus to discover additional LUNs. You can do this by either rebooting the server or reloading the iSCSI driver module.

Rescanning the SCSI bus must be performed with I/O quiecsed. To reload the iSCSI driver on RHEL, Asianux and SLES, use the following command as root:

- RHEL 5 and RHEL 6, and Asianux3 and Asianux 4:
  ```
  # /etc/init.d/iscsid restart
  ```
- SLES 10 and SLES 11:
  ```
  # /etc/init.d/open-iscsi restart
  ```
  Or
  ```
  # /sbin/iscsiadm -m session -R
  ```
- RHEL 7/8 and SLES 12/15:
  ```
  # systemctl restart iscsi
  # systemctl restartiscsid
  ```

(i) **Note:** This command will rescan all running sessions without performing a restart on the iSCSI daemon.

Mounting and unmounting iSCSI file systems automatically (RHEL, Asianux, and SLES)

For RHEL and Asianux, Dell EMC requires that you use the _netdev mount option on all file systems that are mounted on iSCSI devices. The file systems can be either physical or virtual devices (LVM, MD) that are composed of iSCSI devices. This way, the file systems are unmounted automatically by the `netfs initscript` (before iSCSI is stopped) during normal shutdown, and you can easily see which file systems are in network storage. To automatically mount the iSCSI file systems, make an entry in `/etc/fstab` for each file system that the operating systems init scripts will automatically mount and unmount.

- PowerPath device and a native iSCSI device examples:
  ```
  /dev/emcpowera1 /mnt/emcpowera1 ext2 _netdev 0 0
  /dev/sde1 /mnt/sde1 ext2 _netdev 0 0
  ```

For SLES, you can use YaST to choose which devices you want to mount upon system boot. For details, see the *SLES Storage Administration Guide*.

In SLES 10, use the hotplug option in the `/etc/fstab` file to mount iSCSI targets.

- Example entry of a PowerPath device:
  ```
  /dev/emcpowera1 /mnt/emcpowera1 ext4 rw,hotplug 0 0
  ```
- Example entry of a Linux native DM-MPIO device:
  ```
  /dev/disk/by-uuid/c346ca01-3deb-4b44-8b1f-afa28b86a182 /iscsidata
  reiserfs rw,hotplug 0 2
  ```

(i) **Note:** Do not use the Linux native SCSI /dev/sd device, because it does not persist between boots.

- You must also edit the udev rules for the hotplug event to trigger the mounting of the iSCSI device:
  ```
  #> vi /etc/udev/rules.d/85-mount-fstab.rules
  ```
- Example entry of a PowerPath device:
  ```
  # check fstab and possibly mount
  SUBSYSTEM=="block", ACTION=="add",
  KERNEL=="sd*[0-9]|hd*[0-9]|emcpower*[0-9]", RUN+="mount.sh"
  ```
- Example entry of a Linux native DM-MPIO device:
  ```
  # check fstab and possibly mount
  SUBSYSTEM=="block", ACTION=="add",
  KERNEL=="sd*[0-9]|hd*[0-9]|dm-*", RUN+="mount.sh"
  ```

- In SLES 11 and SLES 12, the hotplug option no longer works. Use the nofail option instead:
  ```
  /dev/emcpowera1 /mnt/emcpowera1 ext3 acl,user,nofail 0 0
  ```

For more information, see TID 7004427: */etc/fstab entry does not mount iSCSI device on boot up*, available on SUSE Support website.

Excessive dropped session messages found in /var/log/messages

Sometimes an Ethernet session may be dropped as a normal part of the performance of the network and the Linux network mid-layer's response to such events. Although this is not harmful to the operation of the system because the software initiator recovers the session, but it can prevent the system from achieving its best possible I/O performance.

- The Linux network mid-layer contains a parameter called `tcp_low_latency` to help reduce these occurrences. The `tcp_low_latency` parameter is set after system boot by invoking the following echo(1) command:
  ```
  echo 1 >> /proc/sys/net/ipv4/tcp_low_latency
  ```
- The equivalent sysctl(8) command as follows:
  ```
  # sysctl -w net.ipv4.tcp_low_latency=1
  ```
- This will persist until the next reboot. To persistently set non-default settings to the TCP stack add the following lines to `/etc/sysctl.conf`:
  ```
  net.ipv4.tcp_low_latency=1
  ```
- Use the `sysctl(8)` command to get the value:
  ```
  # sysctl net.ipv4.tcp_low_latency..
  ```

iSCSI write optimization in Unity, or VNX series

Flow control is a feature that gigabit Ethernet ports use to inhibit the transmission of incoming packets. If a buffer on a gigabit Ethernet port runs out of space, the port transmits a special packet that requests remote ports to delay sending packets for a period of time.

The VNX OE for Block version 31 and later, supports Flow Control to optimize 10 G iSCSI connectivity. Flow control can be enabled end-to-end on the switch and host side.

Enabling Flow Control on a Brocade switch

To configure flow control on a gigabit Ethernet port, perform the following task in privileged mode, as shown in the following table.

Table 20 Enabling flow control on a Brocade switch

Step	Task	Command
1	Set the flow control parameters on a gigabit Ethernet port.	• `no switchport` • `no cee` • `switchport` • `switchport mode access` • `no shutdown` • `qos flowcontrol tx on rx`
2	Verify the flow control configuration.	• `On sh run int te 0/7`

The following example shows how to turn transmit and receive flow control on, and how to verify the flow control configuration:

- Enable flow control (802.3x) on the interface by using the `qos flowcontrol tx on rx on CMSH` command:

```
ELARA-8K-21(config)#int te 0/7

ELARA-8K-21(conf-if-te-0/7)# no switchport ELARA-8K-21(conf-if-
te-0/7)# no cee

ELARA-8K-21(conf-if-te-0/7)# switchport

ELARA-8K-21(conf-if-te-0/7)# switchport mode access ELARA-8K-21(conf-
if-te-0/7)# no shutdown

ELARA-8K-21(conf-if-te-0/7)# qos flowcontrol tx on rx on
```

- Verify that it is enabled:
```
ELARA-8K-21# sh run int te 0/7

!interface TenGigabitEthernet 0/7

switchport

switchport mode access

no shutdown

qos flowcontrol tx on rx on
```

Enable flow control on a Cisco switch

To configure flow control on a gigabit Ethernet port, perform the following task in privileged mode, as shown in the following table:

Table 21 Enabling flow control in a Cisco switch

Step	Task	Command
1	Set the flow control parameters on a gigabit Ethernet port.	`set port flowcontrol {receive	send} mod/port {off	on	desired}`
2	Verify the flow control configuration.	`show port flowcontrol`

The following example shows how to turn transmit and receive flow control on and how to verify the flow control configuration:

```
Console> (enable) set port flowcontrol send 2/1 on
Port 2/1 flow control send administration status set to on (port will send
flowcontrol to far end)
Console> (enable) set port flowcontrol receive 2/1 on
Port 2/1 flow control receive administration status set to on (port will
require far end to send flowcontrol)
Console> (enable) show port flowcontrol 2/1

PortSend FlowControl                      RxPause
admin    oper       Receive FlowControl opcodes  TxPause unsupported adminoper
-----        -------     --------    --------     -------  ------- ---------
---------
2/1          on              on          on          on            0
0         0
```

Perform the task in the following table to enable flow control on a gigabit Ethernet NIC and disable window scaling on a Linux host.

Table 22 Enabling flow control in a Cisco switch

Step	Task	Command
1	Set the flow control parameters on a gigabit Ethernet NIC.	`ethtool -a eth<x> autoneg on|off rx on|off tx`
2	Verify the flow control configuration.	`$ ethtool -a eth<x>`
3	Disable window scaling.	`echo 0 > /proc/sys/net/ipv4/tcp_window_ scaling`
4	Verify window scaling is off.	`$ cat /proc/sys/net/ipv4/tcp_window_ scaling`

Linux `ethtool(8)` is used for enabling flow control on an NIC. For detailed information on `ethtool(8)`, see its main page. The following example shows how to configure flow control on eth1:

```
$ ethtool -A eth1 autoneg on rx on tx on
$ ethtool -a eth1 Autonegotiate: on RX: on
TX:on
$ echo 0 > /proc/sys/net/ipv4/tcp_window_scaling
$ cat /proc/sys/net/ipv4/tcp_window_scaling
```

IP routing

All Linux systems may have only one default IP route. When IP routing is required for multiple interfaces on different subnets, the administrator is required to assign the required route(s) to each of the server's interfaces. The administrator can use the `route(8)` command or other networking route menus or utilities to accommodate these changes.

Known problems and limitations

The following table lists the known problems and limitations:

Table 23 Known problems and limitations

Issue number	Description	Workaround/resolution
SuSE BZ: #4 74455 (SLES 10 SP3) #498369 (SLES 11)	There may be data corruption over iSCSI when working with large files.	SuSE acknowledged the problem. Errata will be posted to the relevant branch once it is available.
EMC Artifact # 55072	A start or restart of open-iscsi on SLES 10 SP3 fails. The file `/var/log/messages` displays a message similar to this one: `Jan 25 09:31:44 lin048126 kernel: scsi scan: 192 byte inquiry failed. Consider BLIST_INQUIRY_36 for this device.`	This is not seen in the GA kernel 2.6.16.60-0.54.5 but was a regression that appeared in a subsequent errata kernel. The issue is resolved with errata kernel 2.6.16.60-0.76.8 and greater.
EMC Artifact #55654 SUSE BZ #680464 (SLES10 SP3)	*Open-iscsi* on SLES 10 SP3 does not reestablish the connection after a logout request from a VNXe.	This is seen in the GA release with open-iscsi-2.0.868-0.6.11. The issue is resolved by updating to open-iscsi-2.0.868-0.11.1 and later.

Table 23 Known problems and limitations (continued)

Issue number	Description	Workaround/resolution
OPT # 454647 PowerPath 6.0	Kernel panic might occur when both FE ports are disabled in an iSCSI host.	No solution exists. Red Hat Bugzilla ID# 1142771 was opened.
OPT # 422109 PowerPath 5.7 SP2	I/O fails on SP reboot with Basic Failover for software iSCSI connections.	No solution exists.
OPT # 367442 PowerPath 5.6	Oracle Linux 5.6 (kernel-uek-2.6.32-100) does not show correct information during a target cable pull in a QLogic iSCSI infrastructure. The `powermt display dev=all` command shows paths as live while the target cable pull is done against the owner SP.	No solution exists.
OPT # 454647 PowerPath 6.0	Kernel panic might occur when both FE ports are disabled in the iSCSI host.	No solution exists. Red Hat Bugzilla ID# 1142771 was opened.
OPT# 422109 PowerPath 5.7 SP2	I/O fails on an SP reboot with basic failover for software iSCSI connections.	No solution exists.
OPT# 367442 PowerPath 5.6	Oracle Linux 5.6 (kernel-uek-2.6.32-100) does not show correct information during target cable pull in a QLogic iSCSI infrastructure. The `powermt display dev=all` command shows paths as alive while the target cable pull is done against the owner SP.	No solution exists.
anaconda component, BZ # 1027737	You cannot rescue a system by using an iSCSI disk; when starting anaconda in rescue mode on a system with an iSCSI disk, anaconda does not allow the user to connect to the disk.	Redhat 7.0
kernel component, BZ # 9 15855	The QLogic 1G iSCSI adapter present in the system can cause a call trace error when the qla4xx driver is sharing the interrupt line with the USB sub-system. This error has no impact on the system functionality. The error can be found in the kernel log messages that are located in the `/var/log/messages` file. To prevent the call trace from logging into the kernel log messages, add the nousb kernel parameter when the system is booting.	Redhat 7.0 To prevent the call trace from logging into the kernel log messages, add the nousb kernel parameter when the system is booting.
anaconda component, BZ # 984129 Redhat 6.5	For HP systems running in HP FlexFabric mode, the designated iSCSI function can only be used for iSCSI offload-related operations and will not be able to perform any other Layer 2 networking tasks such as DHCP. For an iSCSI boot from SAN, the	To work around this problem, if DHCP is desired, you must use the `asknetwork` installation parameter and provide a "dummy" static IP address to the corresponding network interface of the iSCSI function. This prevents Anaconda

Table 23 Known problems and limitations (continued)

Issue number	Description	Workaround/resolution
	same SAN MAC address is exposed to both the `correspondingifconfig` record and the iSCSI Boot Firmware Table (iBFT). Anaconda will skip the network selection prompt and attempt to acquire the IP address as specified by iBFT. For DHCP, Anaconda attempts to acquire DHCP using this iSCSI function, which will fail, and Anaconda will then try to acquire DHCP indefinitely.	from entering an infinite loop and allows it to instead request the iSCSI offload function to perform DHCP acquisition.
`iscsi-initiator-utils` component, BZ # 825185 Redhat 6.5	If the corresponding network interface has not been brought up by `dracut` or the tools from the `iscsi-initiator-utils` package, this prevents the correct MAC address from matching the offload interface, and host bus adapter (HBA) mode will not work without manual intervention to bring the corresponding network interface up.	Select the corresponding Layer 2 network interface when anaconda prompts you to select which network interface to install through. This will inherently bring up the offload interface for the installation.
BZ # 1001705 Redhat 6.5	When VDMS (Virtual Desktop Server Manager) attempted to add a new record to the iSCSI database, it failed with the following error: `iscsiadm: Error while adding record: no available memory.` The host is non-operational when connecting to storage.	An upstream patch was applied and the `/var/lib/iscsi` file is now successfully attached.
BZ # 983553 Redhat 6.5	Prior to this update, a single unreachable target could block rescans of others. The `iscsiadm` utility could halt in the D state and the rest of the targets could remain unscanned.	To fix this bug, iscsiadm was made terminable and all the targets were updated. Functioning sessions will now be rescanned properly without long delays.
BZ # 917600 Redhat 6.5	Support for managing flash nodes from the `open-iscsi` utility was added to this package.	If you use `iscsi-initiator-utils`, upgrade to the updated packages, which fix these bugs and add these enhancements.
BZ # 916994 Redhat 6.5	A kernel panic could occur during path failover on systems using multiple iSCSI, FC, or SRP paths to connect an iSCSI initiator and an iSCSI target. This happened because a race condition in the SCSI driver allowed removing a SCSI device from the system before processing its run queue. This led to a NULL pointer dereference.	The SCSI driver was modified and the race is now avoided by holding a reference to a SCSI device run queue while it is active.

Table 23 Known problems and limitations (continued)

Issue number	Description	Workaround/resolution
BZ # 865739 Redhat 6.5	Previously, the tgtd daemon did not report its exported targets properly if configured to report them to an Internet Storage Name Service (iSNS) server. Consequently, running the `iscsiadm -m discoverydb -t isns` command failed.	This bug was. fixed and `tgtd` now reports its exported targets correctly in the described scenario
OPT # 212991	Multiple Linux hosts have the same `iqn` identifiers.	The administrator should check the `/etc/iniatorname.iscsi` file and ensure that each host has a unique iqn name. To generate a new iqn identifier, use the `iSCSI iscsi-iname` utility. For RHEL, use `/sbin/iscsi-iname` and for SLES, use `/usr/sbin/iscsi-iname`.
OPT # 180063	Under a heavy load you may experience dropped Ethernet frames.	Though the frame may be dropped, the Ethernet protocol causes a retry of the data so no data is lost. However, there is a performance hit for this activity. See Excessive dropped session messages found in /var/log/messages.
OPT # 221738 OPT # 221745 DIMS # 126170	Sometimes when a path is lost from one NIC to one SP iSCSI port the paths may be lost from the same NIC to the same SP's other iSCSI port.	This issue is under investigation by Dell EMC Engineering. In a multipath environment under the control of PowerPath, I/O will continue through an alternate path via a trespass until the path is recovered.
SuSE BZ # 212760 SuSE BZ # 251675 SuSE TID # 7004390	`Open-iscsi()` does not find targets automatically when configured to log in at boot. This is first seen on SLES 10.	This is a network start script issue and is dependent on the network configuration. A work-around for this in SLES 10 and SLES 11 is to place a sleep 30 in the beginning of the start stanza in `/ etc/init.d/open-iscsi`. For RHEL 5 or Asianux, the same sleep is installed in `/etc/init.d/iscsid` after `modprobe -q iscsi_tcp`.
DIMS # 145916 SuSE BZ # 172447	A server may not automatically register with a VNX series via Unisphere/Navisphere or the Unisphere/Navisphere Server Utility when installed on SLES 10.	This was fixed in SLES 10 SP1.

CHAPTER 3

Managing Storage and Disk Paths in Linux Environment

This chapter includes the following topics:

Introduction

Dell EMC supports various mechanisms to address multiple paths to a device.

Having redundant access paths to a storage environment is an essential aspect in any storage topology. Online array code upgrades (NDU), online configurations changes, and any disturbances in the topology are best handled by a host when multiple paths are available, and when path management software is installed and configured on the host.

Some of the advantages of path management software include:

- Path failover and path monitoring: Periodically assessing the health of the Host storage connectivity and routing over a preconfigured alternate path if there is a path and/or component failure.

- Load balancing: Ability to improve performance by sharing I/O load across multiple channels.

- Device management: Ability to manage multiple native devices that are instances of a single device. In active-passive array environments the intelligence to route I/O to an active device.

- Device persistence: Upon reboot, when the *scsi* bus is rescanned, the native-device names may not remain persistent from the previous boot. Path Management software provides one mechanism to achieve persistence.

Dell EMC supports three path management options for load balancing, multipathing, and failover, each of which are discussed briefly:

- PowerPath
- Veritas Dynamic Multipathing
- Device-mapper multipath I/O (DM-MPIO)

PowerPath

PowerPath is the premier path-management software available for Fibre-Channel and iSCSI connectivity in a storage multipath environment.

PowerPath for Linux is tightly coupled with the operating system I/O stack. PowerPath provides dynamic load balancing through array-specific optimization algorithms. Load balancing is available on both active/active and active/passive systems including Dell EMC Fibre Channel and/or iSCSI connected storage arrays. Support is also available for non-Dell EMC systems. Consult the Dell EMC Simple Support Matrix for supported revisions of PowerPath on Linux.

Multiple data paths and load-balancing feature

PowerPath uses multiple data paths to share workloads and automate load balancing to ensure that data paths are used efficiently.

PowerPath allows for two or more data paths to be simultaneously used for read/write operations. Performance is enhanced by automatically and intelligently optimizing data access across all available paths. PowerPath is unique in the industry in providing intelligent multipath load-balancing capability.

The workload balancing feature of PowerPath ensures that no one path can become overloaded while others have underutilized bandwidth, thus reducing bottlenecks. When one or more paths become busier than others, PowerPath shifts the I/O traffic. It does so, while tracking of load characteristics from the busy paths to the less-utilized ones, further enhancing the throughput of the server.

Automatic path failover feature

The automatic path failover and dynamic recovery feature of PowerPath permits the dispersal of data access to an alternate data path, if a path failure occurs. It eliminates the possibility of disrupting an application due to the failure of an adapter, cable, or channel controller. If a path failure occurs, all outstanding and subsequent I/O requests are automatically directed to the alternative path.

The intelligent path management feature of PowerPath extends beyond failover. Instead of losing an active HBA is a path failure occurs, PowerPath fails over to an alternative path. Mission-critical applications continue without interruption, performance is optimized, and storage assets are maximized. The Auto Detect and Auto Restore features use periodic probing of inactive paths to check for path failures. Once a failure has been detected and repaired, the path is automatically restored. There is no user intervention required.

PowerPath documentation is available at Dell EMC Online Support and can be found using the words *Linux* and *PowerPath* in the title search.

Veritas Dynamic Multipathing

Veritas Storage Foundation bundles path management software as part of their product offering on Linux. While Veritas Dynamic Multipathing (DMP) provides both load balancing and failover on active/active arrays, DMP only provides failover support on active-passive arrays. Veritas DMP I/O load balancing is performed using a round-robin algorithm.

DMP is supported on hosts attaching to Dell EMC Fibre Channel and or iSCSI connected storage arrays. See the Dell EMC Simple Support Matrix for supported configurations.

Dell EMC supports Veritas Storage Foundation for Linux with Dell EMC Fibre Channel and or iSCSI connected storage arrays. This chapter provides general information to consider if you are planning to use Veritas Volume Manager (VxVM) and Veritas Filesystem (VxFS) in your environment. See the Veritas website for detailed technical and installation information.

Veritas DMP I/O load balancing is performed using a round-robin algorithm. DMP automatically detects the presence of PowerPath on a Linux server and defers all multipathing duties to PowerPath without user intervention.

Dell EMC does not support active coexistence between PowerPath and Veritas DMP. In other words, within the same Linux server, you cannot do the following simultaneously:

• Use PowerPath to control one set of storage array devices.

• Use DMP to control another set of storage array devices.

The same principle applies for Dell EMC Fibre Channel and or iSCSI-connected storage arrays. If PowerPath is installed, it must control all storage array devices.

Dell EMC supports stand-alone DMP failover functionality for the purposes of controlling non-VxVM volume managers. Support requires Storage Foundation 5.1 SP1 or later. See the Dell EMC Simple Support Matrix for supported Linux operating system versions.

(i) Note: On a PowerMAX or VMAX series, the C-bit must be enabled on the array port that is attached to the host for DMP usage.

(i) Note: PowerPath can also be used with Veritas Storage Foundation instead of DMP. PowerPath uses a Third Party Device Driver (TPD) framework in VxVM.

Device-mapper multipath I/O (DM-MPIO)

With kernel distributions 2.6 or later, the multipathing feature has been introduced as part of the operating system.

Native operating system Multipathing is also called as Device Mapper Multipath I/O (DM-MPIO). For supported operating systems along with DM-MPIO, see the Dell EMC Simple Support Matrix.

DM-MPIO provides failover support on Dell EMC storage arrays over Fibre Channel, FCoE, and iSCSI interfaces. The load-balancing algorithm with DM-MPIO is a basic implementation of round-robin scheme over the active paths to the arrays.

(i) **Note:** The coexistence of DM-MPIO and PowerPath on the same host is not supported on Linux.

For more details, see the Native Multipath Failover section.

Operating system distributions also provide information about configuring DM-MPIO on the host, this information is generally available in the `/usr/share/doc/` directory.

See the Dell EMC Simple Support Matrix and the documentation that is listed above, for the appropriate version that is required on the host.

Native Multipath Failover

This section contains the following topics:

- Storage arrays and code revisions
- Supported host bus adapters
- Supported operating systems
- Server platforms
- DM-MPIO on IBM zSeries
- Configuration requirements
- Useful utilities
- Known issues
- MPIO configuration for PowerMAX or VMAX series
- MPIO configuration for Unity storage, VNX Unified Storage, and CLARiiON
- MPIO configuration for Dell EMC Invista or VPLEX virtualized storage
- MPIO configuring for XtremIO storage
- Change the path selector algorithm
- Configure LVM2
- Disabling Linux Multipath
- PowerPath

Storage arrays and code revisions

The following table lists the Dell EMC storage systems that are supported along with native multipathing, the required code revisions, and the types of multipathing:

Table 24 Dell EMC storage systems supported with native multipathing

Storage array	Array code minimum requirements	Type of Multipath supported		
		Active/ Active	PNR	ALUA[a]
PowerMax 2000/8000	PowerMaxOS 5978	Yes	N/A	Yes
VMAX All Flash 250F/FX, 450F/FX, 850F/FX, 950F/FX	HYPERMAX OS 5977	Yes	N/A	Yes, HYPERMAX OS 5977.811.784 and later
VMAX3 400K/200k/100k	HYPERMAX OS 5977	Yes	N/A	Yes, HYPERMAX OS 5977.811.784 and later
VMAX 40K/20k	Enginuity 5876 microcode family	Yes	N/A	N/A
VMAX	Enginuity 5875/5876 microcode family	Yes	N/A	N/A
VMAX 10K (Systems with SN xxx987xxxx)	Enginuity 5876 microcode family	Yes	N/A	N/A
VMAX 10K (Systems with SN xxx959xxxx)	Enginuity 5876 microcode family	Yes	N/A	N/A
VMAXe	Enginuity 5875/5876 microcode family	Yes	N/A	N/A
Symmetrix DMX-4	Enginuity 5772 microcode family	Yes	N/A	N/A
Symmetrix DMX-3	Enginuity 5771 microcode family	Yes	N/A	N/A
Unity series	UnityOE V4.0	N/A	N/A	Yes
VNX series	VNX OE for Block version 31	N/A	Yes, per host	Yes, per host
VNXe series	VNXe OE for Block version 2.0	N/A	N/A	Yes
AX150/150i	FLARE 20	N/A	Yes, per host	N/A
VPLEX	GeoSynchrony 5.0	Yes	N/A	Yes. GeoSynchrony 5.5 and later through the Optimized Path

Table 24 Dell EMC storage systems supported with native multipathing (continued)

Storage array	Array code minimum requirements	Type of Multipath supported		
		Active/ Active	PNR	ALUA[a]
				Management (OPM) feature
XtremIO	XIOS 2.2.1	Yes	N/A	N/A
PowerStore	1.0.0.1.3.x	N/A	N/A	N/A

a. For ALUA deployment, see the following notes:

ⓘ **Note:** Check the following before deploying an array using Asymmetric Logical Unit Access (ALUA) feature on either Fibre Channel or iSCSI:

- Your Linux host operating system environment supports ALUA, see the ALUA supported Linux operating systems table. If this feature is enabled on a host that does not support ALUA, inconsistent device behaviors and or other host failures are expected.

- A failover software package is still required to make a logical unit accessible through the available host target ports. When the failover software detects a path failure, the failover software automatically reconfigures the target device to make it accessible using other available target ports.

- See the Dell EMC Simple Support Matrix for the latest supported configurations.

Supported host bus adapters

Linux DM-MPIO is supported on both Fibre Channel, iSCSI, and Fibre Channel over Ethernet environments.

Fibre Channel host bus adapters are supported with Dell EMC Fibre Channel and Fibre Channel over Ethernet storage arrays. Always see the Linux "Base Connectivity" section of the Dell EMC Simple Support Matrix for supported HBA models. See the appropriate install guide for configuring the host adapter and driver for the system.

Converged Network Adapters (CNAs), with the Linux *open-fcoe* driver stack, are supported with Fibre Channel and Fibre Channel over Ethernet storage arrays. Always see the Linux *Base Connectivity* section of the Dell EMC Simple Support Matrix for supported CNA models and see the appropriate install guide for configuring the host adapter and driver for the system.

Dell EMC-published HBA, and CNA driver configuration guides are available in the Dell EMC-approved sections of the applicable vendor.

Approved vendor hardware initiators and the generic Network Interface Card iSCSI software initiator are supported with Dell EMC iSCSI storage arrays. See the Linux *iSCSI Connectivity* section of the Dell EMC Simple Support Matrix for supported configurations and required driver revisions.

Supported operating systems

Native multipath failover is supported in production environments on Linux only in the following configurations:

- Red Hat Enterprise Linux 5 and newer updates

- Red Hat Enterprise Linux 6 and newer updates
- Red Hat Enterprise Linux 7 and newer updates
- Red Hat Enterprise Linux 8 and newer updates
- SuSE Linux Enterprise Server (SLES) 10 and newer service packs
- SuSE Linux Enterprise Server (SLES) 11 and newer service packs
- SuSE Linux Enterprise Server (SLES) 12 and newer service packs
- SuSE Linux Enterprise Server (SLES) 15 and newer service packs
- Oracle Linux (OL) Release 5 and newer updates
- Oracle Linux (OL) Release 6 and newer updates
- Oracle Linux (OL) Release 7 and newer updates
- Oracle Linux (OL) Release 8 and newer updates
- Asianux 4.0 and newer service packs

The listed operating systems are supported by letters of support that may be found under the **Extended Support** tab in the Dell EMC Simple Support Matrix. To configure the native multipathing, see the subsections containing Red Hat Enterprise Linux (Red Hat Enterprise Linux) equivalent release versions. Because Scientific Linux and CentOS are built from Red Hat Enterprise Linux, and Oracle VM server is built from Oracle Linux, functionality and guidance in this guide are considered equivalent unless otherwise noted.

- CentOS
- Scientific Linux
- Oracle VM server
- Dell EMC only supports kernels that are packaged with a distributed operating system.

Server platforms

The following system architectures are supported with native MPIO:

- Intel 32-bit
- Intel EM64T
- Intel IA64
- AMD64 Opteron
- PowerPC

See the Linux *Base Connectivity* section of the Dell EMC Simple Support Matrix for supported server models.

DM-MPIO on IBM zSeries

DM-MPIO is supported to run on an LPAR of an IBM zSeries system after the Red Hat Enterprise Linux 5 and SLES 10 releases.

DM-MPIO is also supported to run as a guest operating system with Red Hat Enterprise Linux 5 and SLES10 on z/VM 5.1 and 5.2 on the IBM zSeries. Only PowerMAX and VMAX volumes are supported while connected to the system over FCP. Both 31-bit and 64-bit architectures are supported.

ⓘ Note:

- For Dell EMC support, the operating system should be installed only on CKD devices. Support is not available when the operating system is installed on an FCP device. Disks that are not used for the operating system installation can either be FCP or CKD devices.

- DM-MPIO support is only available for FCP devices.

Configuration requirements

The following are configuration requirements when connected to Dell EMC storage:

- The maximum SCSI devices, paths, and LUNs follow the same guidelines that are specified in Operating system limits and guidelines.

- Boot from SAN of a multipathed LUN is introduced and supported by Dell EMC. It is supported with the release of Red Hat Enterprise Linux 5.1 and SLES 10 SP2. For more information, see Booting from SAN.

- Only round-robin I/O load sharing algorithm is supported with Native MPIO.

- Dell EMC does not support mixing Fibre Channel, Fibre Channel over Ethernet, and iSCSI to the same host system from the same storage array.

(i) Note: The `/etc/multipath.conf` configurations in the sections that follow are Dell EMC defaults used for DM-MPIO qualification with Dell EMC storage. The integrated defaults in DM-MPIO are based on Dell EMC defaults as well.

Dell EMC recognizes that not every SAN environment is similar to the one used in Dell EMC qualification. Also, each customer site may need to deviate from these defaults to tune DM-MPIO for their environment. If you should deviate from these settings, it is recommended that you return to the default settings as part of the root cause analysis.

Useful utilities

This section lists the system utilities that are useful when using DM-MPIO. See the man pages for detailed information and usage instructions.

Table 25 Useful utilities

Command name	Purpose (From the respective man pages)
dmsetup	dmsetup manages logical devices that use the device-mapper driver.
lvm	lvm provides the command-line tools for LVM2.
multipath	multipath is used to detect multiple paths to devices for fail-over or performance reasons and coalesces them.
udev	udev creates or deletes device node files that are located in the /dev directory. It provides a dynamic device directory containing only the files for present devices.
udevinfo	udevinfo queries the udev database for device information that is stored in the udev database. It can also query the properties of a device from its sysfs representation to help create udev rules that match this device.
udevmonitor	udevmonitor listens to the kernel uevents and events that are sent by a udev rule and prints the devpath of the event to the console. It can be used to analyze the event timing by comparing the timestamps of the kernel uevent with the udev event.

Table 25 Useful utilities (continued)

Command name	Purpose (From the respective man pages)
iostat	The iostat command is used for monitoring system input/output device loading. It does so observing the duration of time that the devices are active in relation to their average transfer rates.
hotplug	hotplug is a program that is used by the kernel to notify user mode software when some significant (hardware-related) events take place.
devmap_name	devmap_name queries the device-mapper for a name for the device. This is specified by major and minor number.
kpartx	This tool reads partition tables on specified device and creates device maps over partitions segments detected.
scsi_id	scsi_id queries a SCSI device using the SCSI INQUIRY vital product data (VPD) page 0x80 or 0x83. It uses the resulting data to generate a value that is unique across all SCSI devices that properly support page 0x80 or page 0x83.
lsblk	Lsblk lists all block devices (except RAM disks) in a tree-like format by default. The command reads the sysfs file system to gather information.

Known issues

This section lists known issues in native multipath failover, along with workarounds.

Table 26 Known issues

Problem description	Environments	Workaround/Fix
Device-mapper (dm) names and sd names may not be persistent across reboots.	All	Persistence can be achieved by: Use of LVM on top of the DM names. Use of scsi_id based names. Use of simple multipath names. Use of device aliases defined in the /etc/multipath.conf file.
During a host reboot having unfractured SnapView clones in a storage group of the host, the server may stop responding.	First seen in SLES 10	The workaround is to fracture or delete any unfractured cloned LUNs from the host storage group before that host is rebooted.
In a cluster configured with a VNX series, VNXe series or Unity series using ALUA, there is no follow-over capability. So, one host may fail the LUN back to the original SP after the host with the failing path has	All RedHat, Asianux, Oracle OL versions; also, all SLES 9 and all SLES10	See Red Hat BZ #437107 and SuSE BZ #370057 and #502626. As a workaround in the VNX series, VNXe series or Unity series, change the default owner of the LUN to a storage processor accessible by all hosts that are attached

Table 26 Known issues (continued)

Problem description	Environments	Workaround/Fix
trespassed the LUN to the secondary SP.		to the storage group. This is resolved in Red Hat Enterprise Linux 5.5, OL 5.5, Asianux 3.0 SP3, and in SLES 11 with kernel 2.6.27.37-0.1 and Multipath tools 0.4.8-40.6.1.
Path follow-over functionality is not available in MPIO. In certain scenarios, this may result in repeated trespasses when connected to VNX series, VNXe series or Unity series resulting in lower performance.	All RedHat, Asianux, Oracle OL versions; also SLES 9 and SLES10 SP1, SP2, and SP3	In the VNX series, VNXe series or Unity series, change the default owner of the LUN to a storage processor accessible by all hosts that are attached to the storage group under nonfailure conditions. This is resolved in Red Hat Enterprise Linux 5.5, OL 5.5 and in SLES 11 with kernel 2.6.27.37-0.1 and Multipath tools 0.4.8-40.6.1.
Multipath configuration with simple names results in multiple distinct devices being addressed with the same *mpath* name.	SLES 10	Edit the `/var/lib/multipath/bindings` file and delete all duplicate occurrences of the same simple name. Run the multipath command for reconfiguring all devices.

During an out of family code load (NDU) on a Symmetrix array, directors to the LUN may go offline simultaneously. This prevents access to the LUN and I/O will fail immediately when using the default DM-MPIO settings for Symmetrix.	See the Dell EMC Knowledge base solution emc212937.	To work around this issue, configure the *no_path_retry to* queue. For example:
		
		 ``` devices { ## Device attributes for EMC SYMMETRIX device { vendor "EMC" product "SYMMETRIX" no_path_retry queue } } ```
		 This causes I/O to pend until the path is restored. Another solution is to queue for some retries such as the following:
		
		 ``` devices { ## Device attributes for EMC SYMMETRIX device { vendor "EMC" product "SYMMETRIX" path_checker tur ```

Table 26 Known issues (continued)

Problem description	Environments	Workaround/Fix
		``
polling_interval 5
no_path_retry   6
}
}
```
These are tunable parameters and can be adjusted to accommodate the particular environment for the results you want to achieve. |
| In SLES 11, when a path fails, all references to the path and the devices are deleted from the kernel operating system. | SLES 11 | This is a change in the behavior of the Linux kernel. Expect that if a path fails, it is deleted from the output of the `multipath -l` command. |
| If you are restoring failed paths, the LUN will trespass back only after all paths are restored. | SLES 11 | This is newer behavior that appears in the later versions of DM-MPIO and is first seen in SLES 11. |
| Multiple cable failures at one time cause I/O to stop responding. | SLES 11- multipath-tools-0.4.8-40.4.1 | Bz 518007 - Reinsertion of the cables will not re-enable the paths. Cable pulls with greater than 10 seconds apart do not exhibit this behavior. Fixed in multipath-tools-0.4.8-40.5 |
| Dell EMC Knowledge base solution #emc227052 Dell EMC VPLEX requires its own specific configuration stanza in the `/etc/multipath.conf` file. | All supported releases | Add:
```
devices {
Device attributes
for EMC Invista
device {
vendor   "EMC"
product   "Invista"
path_checker   tur
no_path_retry   5
product_blacklist
"LUNZ"
}
}
```
This has been resolved as a default in:
• Red Hat Enterprise Linux 5.5 and higher
• SLES 10 SP3 and higher
• SLES 11 SP1 and higher |

**Table 26** Known issues (continued)

| Problem description | Environments | Workaround/Fix |
|---|---|---|
| | | So, you do not need to add this stanza to support VPLEX. |
| VPLEX requires its own specific configuration stanza in the `/etc/multipath.conf` file. | All supported releases | Add:<br><br>```<br>devices {<br>## Device attributes<br>for EMC Invista<br>device {<br>vendor      "EMC"<br>product     "Invista"<br>path_checker    tur<br>no_path_retry    5<br>product_blacklist<br>"LUNZ"<br>}<br>}<br>```<br><br>This has been resolved as a default in:<br><br>• Red Hat Enterprise Linux 5.5 and higher<br><br>• SLES 10 SP3 and higher<br><br>• SLES 11 SP1 and higher<br><br>So, you do not need to add this stanza to support VPLEX. |
| SuSE BZ number 254644 - root device that is not managed by DM-MPIO on VNX series. | SLES 10 | This can occur if you are attempting to configure many LUNs attached to the server and you run out of file descriptors (fd). The limit for the maximum number of open files is 1024 (ulimit -n). As multipath and multipathd require one fd for each path, you have to increase this limit to roughly 2 times the number of paths. Edit the script `/etc/init.d/multipathd` and change the variable MAX_OPEN_FDS to a number greater than the existing value. |
| Red Hat BZ #509095 - inconsistent multipath maps following storage addition. | Red Hat Enterprise Linux 5 | *multipath.conf* now includes the *bindings_file* default option. On installations where `/var/lib/multipath/bindings` is not on the same device as the root file system, this option should be set to a path on the |

Table 26 Known issues (continued)

| Problem description | Environments | Workaround/Fix |
|---|---|---|
| | | same device as the root file system. For example, `/etc/multipath_bindings`. **By setting this, multipath uses the same bindings during boot as it does during normal operation.** |
| My Celerra or VNXe storage device is detected as an Unknown product by DM-MPIO and requires configuration on my Linux server. | All supported operating systems listed in the Supported operating systems section. | Use this device stanza:<br><br>```device {<br>vendor "EMC" product<br>"Celerra"<br>path_grouping_policy<br>"multibus"<br>path_checker "tur"<br>no_path_retry "30"<br>}```<br><br>In the more recent release of the Linux operating system, the Celerra and VNXe arrays are detected by default and they are listed as Celerra. |
| Red Hat BZ number 467709 - a trespass storm may occur when DM-MPIO is configured in a cluster. | Always existed in all versions of Linux using DM-MPIO in a cluster configuration. | Fixed in Red Hat Enterprise Linux 6.3 Redhat Customer Portal 6.3. device-mapper-multipath technical notes. The multipathd daemon did not have a failover method to handle switching of path groups when multiple nodes were using the same storage. Also, if one node lost access to the preferred paths to a logical unit, while the preferred path of the other node was preserved, multipathd could end up switching back and forth between path groups. This update adds the follow over failback method to device-mapper-multipath. If the follow over failback method is set, multipathd does not fail back to the preferred path group, unless it returned to the online state recently. When multiple nodes are using the same storage, a path failing on one machine |

**Table 26** Known issues (continued)

| Problem description | Environments | Workaround/Fix |
|---|---|---|
| | | now no longer causes the path groups to continually switch back and forth. |
| SUSE BZ #802456 - Dell EMC active/active ALUA was not driving I/O down all paths. | First observed in SLES 11. Not observed in Red Hat Enterprise Linux. | Fixed in SLES 11 SP2 patch multipath-tools 8339 (multipath-tools-0.4.9-0.70.72.1 and kpartx-0.4.9-0.70.72.1). |

(i) **Note:** On VNX series, VNXe series or Unity series, running the multipath command at any time will result in an attempt to restore the device to its default owners. This does not impact availability of the device.

(i) **Note:** PowerMAX and VMAX series Layered Applications are supported along with native multipathing. Include any gatekeeper devices in the blacklist so that device mapper does not control these devices. Sample multipath.conf files are available in the following sections that detail the steps that are required.

(i) **Note:** The following host utilities are supported in Linux native multipathing environments:

- VNX series
- Admsnap

Support in Linux native multipathing environments is also available for MirrorView, SAN Copy, and SnapView clones.

Unfractured SnapViewclones are *not* supported in active storage groups in the Linux native multipathing environment. During a host reboot, having unfractured SnapView clones in the storage group of the host may cause the host to stop responding. The workaround is to fracture or delete any unfractured cloned LUNs from the host storage group before that host is rebooted.

# Change the path selector algorithm

From the Red Hat Enterprise Linux 6.0 release, two new path selector algorithms were added to DM-MPIO. This can be set as the selector algorithm to use for all your connected storage. You can set this algorithm in the *defaults stanza* or on an individual storage device in the device stanza. There are three selector algorithms:

**About this task**

- round-robin 0
  Loop through every path in the path group, sending the same amount of I/O to each.

- queue-length 0
  Send the next bunch of I/O down the path with the least amount of outstanding I/O.

- service-time 0
  Send the next bunch of I/O down the path with the shortest estimated service time. This is determined by dividing the total size of the outstanding I/O to each path by its relative throughput.

**Procedure**

1. To change to an alternate path selector, create a device stanza for the Dell EMC product you want to alter. See the following XtremIO in Red Hat Enterprise Linux 7:

```
devices {
device {
vendor XtremIO product XtremApp
path_selector "queue-length 0"
rr_min_io_rq 1 path_grouping_policy multibus path_checker tur
failback immediate fast_io_fail_tmo 15
}
}
```

ⓘ **Note:** If you change the path selector in the defaults stanza, it does not take effect.

Configure LVM2

When using LVM, it is recommended that Logical Volumes are created on DM-MPIO devices instead of SCSI devices for the following reasons:

- In a multipath environment, more than one SCSI sd device direct to the same physical device. Using LVM on sd devices results in the duplicate entries being reported during the LVM scan.

- Depending on the order of the scan, it is conceivable that the LVM volumes are tied to particular sd devices instead of the multipath infrastructure. This may result in multipath infrastructure not providing failover capabilities if a SAN fails or device becomes unavailability.

- The SCSI device (sd) names are not persistent across reboots or SAN changes.

By default, LVM2 does not scan for multipath devices. Also, LVM scans for all available block devices. For LVM2 operation in a multipath environment with DM-MPIO, the sd devices need to be filtered out. You must also be include the device mapper devices as part of the volume scan operation. The sections that follow outline the procedures for Red Hat Enterprise Linux 4, Red Hat Enterprise Linux 5, SLES 9, and SLES 10.

ⓘ **Note:** The sections that follow provide sample filters that may be used to configure LVM, but your environment may differ and require a different filter.

Configuring LVM2 for DM-MPIO on Red Hat Enterprise Linux

Perform the procedure described in this section to configure LVM2 for DM-MPIO on Red Hat Enterprise Linux.

Procedure

1. Add the following line to `/etc/lvm/lvm.conf` file, to enable scanning of device-mapper block devices.

   `types = [ "device-mapper", 1]`

2. Filter all sd devices from the system and choose to scan for multipath devices by adding the following line:

   `filter = [ "a/dev/mpath/.*/", "r/.*/" ]`

   ⓘ **Note:** This filters the boot device if it is under LVM control. If your configuration is booted under LVM, perform the next step.

3. If Multipath does not control devices with Logical Volumes, then enable selective scanning of those devices. For instance, in the below example, the partition sda2 contains a Logical Volume. However, the sd device is not under multipath control. To enable scanning for this device, set the following filter as follows.

```
filter = [ "a/dev/sda [1-9]$/","a/dev/mpath/.*/", "r/.*/" ]
```

4. Save the edits to the `/etc/lvm/lvm.conf` file.

5. Run the command `lvmdiskscan` and ensure that the required SCSI devices are scanned and that the LVM volume groups and partitions are available.

Configuring LVM2 for DM-MPIO on SLES

To configure LVM2 for DM-MPIO on SLES:

Procedure

1. Add the following line to `/etc/lvm/lvm.conf` file, to enable scanning of device-mapper block devices.

   ```
   types = [ "device-mapper", 1]
   ```

2. To filter sd devices from the system and scan for multipath devices in the `/dev/disk/by-name` persistent directory that udev generates, replace the default filter as follows:

   - For SLES 10:
     ```
     filter = [ "a/dev/mapper/mpath.*/", "r/.*/" ]
     ```

   - For SLES 11 and 12:
     ```
     filter = [ "a|/dev/disk/by-id/dm-uuid-.*-mpath-.*|", "r|.*|"
     ```

3. If multipath does no control devices with Logical Volumes, then enable selective scanning of those devices. For instance, in this example, the partition sda2 contains a Logical Volume. However, the sd device is not under multipath control. To enable scanning for this device, set the following filter as follows.

   - For SLES 10:
     ```
     filter = [ "a/dev/mapper/mpath.*/", "a/dev/sda2$/", "r/.*/" ]
     ```

   - For SLES 11 and 12:
     ```
     filter = [ "a|/dev/disk/by-id/dm-uuid-.*-mpath-.*|", "a|/dev/
     sda2$/" "r|.*|" ]
     ```

   Save the edits to the `/etc/lvm/lvm.conf` file. Run the command `lvmdiskscan` and ensure that the required SCSI devices are scanned and that the LVM volume groups and partitions are available.

Disable Linux Multipath

If you decide to upgrade to PowerPath for path management, it is necessary to disable Linux Multipath before installing PowerPath. It not, both multipathing applications hold locks on the same devices, making the server unstable for use.

Before you begin

Before you install PowerPath for Linux, perform the following steps:

Procedure

1. As root, edit the `/etc/multipath.conf` file by commenting out all its present entries. To do so, insert a **#** symbol at the beginning of each line entry in the configuration file and then add the following:

   ```
   blacklist {
   devnode "*"
   }
   ```

2. From root, stop the multipath daemon, disable the multipath daemon, clear the device mapper table, and remove the Linux Multipath utility package:

```
#> /etc/init.d/multipathd stop
#> /sbin/chkconfig multipathd off
#> /sbin/dmsetup remove_all
#> /sbin/rpm -e `rpm -qa | grep multipath`
```

3. Reboot the server, and install PowerPath per the PowerPath installation documentation, available at Dell EMC Online Support.

PowerPath

The following Path Management software was examined. The findings suggest that PowerPath did not appear to be affected in this scenario.

Linux Native MPIO

Linux Native MPIO did not appear to be affected in this scenario in earlier releases of Linux. However, changes have been made to the Linux kernel in Red Hat Enterprise Linux 6, SLES 10 SP4, and SLES 11, that show an altered behavior. The path is faulted when an under-provisioned pool is exhausted. The following figures show examples of SLES.

Figure 13   Filesystem mpathq-part1 made from thin LUN

```
SUSE10SP4:~ # mount
/dev/sda2 on / type reiserfs (rw,acl,user_xattr)
proc on /proc type proc (rw)
sysfs on /sys type sysfs (rw)
debugfs on /sys/kernel/debug type debugfs (rw)
udev on /dev type tmpfs (rw)
devpts on /dev/pts type devpts (rw,mode=0620,gid=5)
securityfs on /sys/kernel/security type securityfs (rw)
/dev/mapper/mpathq-part1 on /zoner/mpathq-part1 type ext3 (rw)
```

Figure 14   Provisioned size: 200G; actual allocated size:17G

```
SUSE10SP4:~ # df -lh
Filesystem                Size  Used Avail Use% Mounted on
/dev/sda2                  67G   12G   55G  18% /
udev                      5.9G  336K  5.9G   1% /dev
/dev/mapper/mpathq-part1
                          197G   17G  171G   9% /zoner/mpathq-part1
SUSE10SP4:~ #
SUSE10SP4:~ #
```

Figure 15 I/O error being seen in /var/log/messages when pool being exhausted

```
May  5 22:05:43 SUSE10SP4 kernel: sdk: Current: sense key: Data Protect
May  5 22:05:43 SUSE10SP4 kernel:     ASC=0x27 ASCQ=0x7
May  5 22:05:43 SUSE10SP4 kernel: end_request: I/O error, dev sdk, sector 148868727
May  5 22:05:43 SUSE10SP4 kernel: sd 7:0:0:6: SCSI error: return code = 0x08000002
May  5 22:05:43 SUSE10SP4 kernel: sdk: Current: sense key: Data Protect
May  5 22:05:43 SUSE10SP4 kernel:     ASC=0x27 ASCQ=0x7
May  5 22:05:43 SUSE10SP4 kernel: end_request: I/O error, dev sdk, sector 145600151
May  5 22:05:43 SUSE10SP4 kernel: sd 7:0:0:6: SCSI error: return code = 0x08000002
May  5 22:05:43 SUSE10SP4 kernel: sdk: Current: sense key: Data Protect
May  5 22:05:43 SUSE10SP4 kernel:     ASC=0x27 ASCQ=0x7
May  5 22:05:43 SUSE10SP4 kernel: end_request: I/O error, dev sdk, sector 148368735
May  5 22:05:43 SUSE10SP4 kernel: sd 7:0:0:6: SCSI error: return code = 0x08000002
May  5 22:05:43 SUSE10SP4 kernel: sdk: Current: sense key: Data Protect
May  5 22:05:43 SUSE10SP4 kernel:     ASC=0x27 ASCQ=0x7
May  5 22:05:43 SUSE10SP4 kernel: end_request: I/O error, dev sdk, sector 148751511
May  5 22:05:43 SUSE10SP4 kernel: sd 7:0:0:6: SCSI error: return code = 0x08000002
May  5 22:05:43 SUSE10SP4 kernel: sdk: Current: sense key: Data Protect
May  5 22:05:43 SUSE10SP4 kernel:     ASC=0x27 ASCQ=0x7
May  5 22:05:43 SUSE10SP4 kernel: end_request: I/O error, dev sdk, sector 146037351
May  5 22:05:43 SUSE10SP4 kernel: sd 7:0:0:6: SCSI error: return code = 0x08000002
May  5 22:05:43 SUSE10SP4 kernel: sdk: Current: sense key: Data Protect
May  5 22:05:43 SUSE10SP4 kernel:     ASC=0x27 ASCQ=0x7
May  5 22:05:43 SUSE10SP4 kernel: end_request: I/O error, dev sdk, sector 147046479
```

Figure 16 Thin LUN paths both marked as failed

```
SUSE10SP4:~ # multipath -l
mpathq (3600601608a903200882fdebf05b6e211) dm-6 DGC,VRAID
[size=200G][features=1 queue_if_no_path][hwhandler=1 emc]
\_ round-robin 0 [prio=0][enabled]
 \_ 7:0:0:6 sdk 8:160 [failed][undef]
\_ round-robin 0 [prio=0][enabled]
 \_ 6:0:0:6 sdo 8:224 [failed][undef]
mpathp (3600601608a9032001f6a87a7d0afe211) dm-3 DGC,VRAID
[size=6.0G][features=1 queue_if_no_path][hwhandler=1 emc]
\_ round-robin 0 [prio=0][active]
 \_ 6:0:0:0 sdb 8:16  [active][undef]
\_ round-robin 0 [prio=0][enabled]
 \_ 7:0:0:0 sdc 8:32  [active][undef]
mpatho (3600601608a9032000bba20f6d0afe211) dm-5 DGC,RAID 1
[size=5.0G][features=1 queue_if_no_path][hwhandler=1 emc]
\_ round-robin 0 [prio=0][active]
 \_ 6:0:0:1 sdf 8:80  [active][undef]
\_ round-robin 0 [prio=0][enabled]
 \_ 7:0:0:1 sdd 8:48  [active][undef]
mpathn (3600601608a903200f182ae6cd1afe211) dm-4 DGC,RAID 10
[size=5.0G][features=1 queue_if_no_path][hwhandler=1 emc]
\_ round-robin 0 [prio=0][active]
 \_ 6:0:0:2 sdi 8:128 [active][undef]
\_ round-robin 0 [prio=0][enabled]
 \_ 7:0:0:2 sde 8:64  [active][undef]
mpathm (3600601608a9032002a6c340eceafe211) dm-0 DGC,RAID 5
[size=2.1T][features=1 queue_if_no_path][hwhandler=1 emc]
\_ round-robin 0 [prio=0][active]
 \_ 7:0:0:3 sdg 8:96  [active][undef]
\_ round-robin 0 [prio=0][enabled]
 \_ 6:0:0:3 sdl 8:176 [active][undef]
mpathl (3600601608a903200165525ccd0afe211) dm-2 DGC,RAID 5
[size=5.0G][features=1 queue_if_no_path][hwhandler=1 emc]
\_ round-robin 0 [prio=0][active]
```

CHAPTER 4

Host Connectivity with Dell EMC Products

This chapter includes the following topics:

PowerMax/VMAX All Flash

This section provides support information about the PowerMax, VMAX All Flash, and VMAX3 series.

Dell EMC PowerMax

Dell EMC PowerMax provides a platform that is ready for current and next generation data center data storage requirements. The PowerMax Family uses NVMe technology for customer application data.

PowerMax is built using a 100% NVMe storage back-end, allowing it to reach optimal IOPS densities by eliminating the flash media issues that are found in traditional SAS and SATA interfaces. NVMe over Fibre Channel 32 Gb front-end connections are supported starting with the PowerMaxOS 5978.444 release running on PowerMax storage systems allows for end-to-end NVMe accessibility, the PowerMaxOS 5978.444 and later releases running on PowerMax storage systems. It also introduces capability for traditional Fibre Channel FCP connections with an operation link speed of 32 Gb. NVMe over Fibre Channel and 32 Gb FCP both require the installation of the 32 Gb Fibre Channel I/O module. The PowerMaxOS 5978.479 and later releases leverages the 32 Gb Fibre Channel I/O module for 32 Git SRDF link connection support. All PowerMaxOS releases maintain downward compatibility with the legacy 16 Gb Fibre Channel I/O module.

There are two PowerMax models as follows:

PowerMax 2000-PowerMax 2000 is designed to provide high efficiency and flexibility to the data center, providing 1.7 million IOPS (8 K RRH) and up to 1 PB of effective capacity in 20U total space. It is the entry NVMe scale out array sold with the Essentials and Pro software packages.

PowerMax 8000-PowerMax 8000 designed to provide optimal scalability, performance, and IOPS density. It can consolidate disparate workloads on a mass scale as 8 Bricks can support 10 million IOPS (8 K RRH) and provide up to 4 PB of effective capacity in just two floor tiles of space. It is the flagship NVMe scale out array sold with the Essentials and Pro software packages.

Dell EMC PowerMax Operating System environment

PowerMaxOS 5978 is the only supported software operating environment for PowerMax storage arrays. It manages the storage and controls communications with the host systems and introduces support of dedupe and other new feature enhancements.

For detailed product specifications including the 32G Frontend I/O and 32G SRDF protocol support information, see Dell EMC PowerMax Family documentation available on https://shop.dellemc.com and *Dell EMC PowerMax Family Product Guide* on Dell EMC Online Support.

Dell EMC VMAX All Flash

Dell EMC VMAX All Flash range of storage arrays use only high-density flash drives. The range contains four models that combine high scale, low latency, and rich data services:

- VMAX 250F with a maximum capacity of 1.16 PBe (Petabytes effective)
- VMAX 450F with a maximum capacity of 2.3 PBe
- VMAX 850F with a maximum capacity of 4.4 PBe
- VMAX 950F with a maximum capacity of 4.42 PBe

Dell EMC VMAX All Flash Operating Environment

Dell EMC VMAX All Flash storage systems are supported with both HYPERMAX OS 5977 and PowerMaxOS 5978, however not all new feature enhancements that are introduced with

PowerMax storage systems running PowerMaxOS 5978 would be available with Dell EMC VMAX All Flash storage systems also running PowerMaxOS 5978.

HYPERMAX OS 5977 provides emulations that perform specific data service and control functions in the HYPERMAX environment, and it introduces an open application platform for running data services and provide file system storage with eNAS and embedded management services for Unisphere. The entire feature set available with HYPERMAX OS 5977 running on Dell EMC VMAX All Flash storage systems would be available with PowerMaxOS 5978 running on Dell EMC VMAX All Flash storage systems except FCoE front-end connections.

For detailed product specifications including the Frontend I/O protocol support information, see *VMAX All Flash Product Guide* and *VMAX All Flash: Family Overview* available on https://www.dellemc.com/.

PowerMax/VMAX All Flash/VMAX3 - Device types

There are two classes of devices available on PowerMax/VMAX All Flash/VMAX3 storage arrays:

- Compatibility device - The Compatibility device is the default VMAX device type, advertises T10 SPC-3 compliancy and behaves almost like the legacy VMAX devices.

- Mobility device - The Mobility device is an optionally configurable device type that is supported with HYPERMAX OS 5977.811.784 and later, advertises T10 SPC-4 compliancy, and has numerous INQUIRY and VPD page changes implemented intended to allow for support of future feature enhancements. Implicit Asymmetric Logical Unit Access (ALUA) is advertised by the Mobility device, however the Mobility device is configured into only a single active target port group per PowerMax or VMAX storage system and a second target port group is configured on a remote PowerMax or VMAX system only when the Mobility devices are configured in an SRDF Metro configuration which is restricted only to Windows and Solaris operating system environments for now. For all supported SRDF/Metro Mobility ID/ALUA, see the Support Matrix available on Dell EMC E-Lab Navigator.
Mobility devices are also not currently supported with NDM migration.

Local replication services

TimeFinder SnapVX

Dell EMC TimeFinder SnapVX creates and manages point-in-time snapshots of critical data that can be used for backups, decision support, and to refresh data warehouse, test, and development environments. SnapVX snapshots do not require target volumes. SnapVX snapshots share back-end allocations with the source volume and other snapshots on the source volume.

TimeFinder SnapVX is supported on VMAX All Flash arrays running HYPERMAX OS 5977 and later, and snapshots are always consistent. Consistency across multiple arrays is achieved when source devices are in a composite group.

SnapVX provides very low impact snapshots and clones for VMAX LUNs. SnapVX supports up to 256 snapshots per source volume, which are tracked as versions with less overhead and simple relationship tracking. Users can assign names to identify their snapshots, and can set automatic expiration dates on each snapshot. SnapVX provides the ability to manage consistent point-in-time copies for storage groups with a single operation. Up to 1024 target volumes can be linked per source volume, providing read/write access as pointers or full-copy clones. TimeFinder in HYPERMAX OS also provides compatibility modes for users who rely on their TimeFinder Mirror, Clone, or VP Snap command scripts. This allows users to leverage their existing scripts while learning how to take advantage of the new features of SnapVX.

For details, see the *TimeFinder SNAPVX Local Replication* document available on PowerMax and VMAX All Flash Technical Documentation page on https://www.dellemc.com/ or https://www.dell.com/support.

Remote Replication Services

SRDF

The Symmetrix Remote Data Facility (SRDF) maintains real-time or near real-time copies of data on a production storage array at one or more remote storage arrays and only one SRDF device mirror can be read/write enabled at any point in time.

SRDF/Metro

HYPERMAX OS 5977.691.684 and Solutions Enabler/Unisphere for VMAX 8.1 first introduced the support for SRDF/Metro for VMAX3 and VMAX All Flash families of storage arrays and all versions of PowerMaxOS on Dell EMC PowerMax arrays.

With SRDF/Metro, the SRDF secondary device is read/write accessible to the host and takes on the external identity of the primary device (geometry, device WWN, and so on). By providing this external identity on the secondary device, both the primary and secondary devices may then appear as a single virtual device across the two SRDF paired arrays for presentation to a single host or host cluster.

With both the local and remote devices being read/write accessible concurrently, the host or hosts (in a cluster) can read and write to both primary and secondary devices with SRDF/Metro ensuring that each copy remains current, consistent, and addressing any write conflicts which may occur between the paired SRDF devices. A single PowerMax/VMAX3/VMAX All Flash array may simultaneously support multiple SRDF groups that are configured for SRDF/Metro operations and multiple SRDF groups that are configured for non-SRDF/Metro operations. PowerMaxOS Q3 2019 release with SRDF/Metro supports online device expansion though which devices could be expanded within an active SRDF/Metro group. These operations can be accomplished using the latest version of Unisphere for PowerMax or Solutions enabler. Both SRDF and SRDF/Metro support FC or GiGE links for connectivity between arrays.

For more detailed information about SRDF/metro configuration, see *Dell EMC SRDF Introduction* and *SRDF/Metro Overview and Best Practices Technical Note* available on https://www.dell.com/support.

Non-Disruptive Migration services

Dell EMC's Non-Disruptive Migration (NDM) allows user to perform online data migrations that are simple and completely non-disruptive to the host and application. NDM is designed to help automate the process of migrating hosts and applications to a new PowerMax array with no downtime. NDM leverages SRDF replication technologies to move the application data to the new array. It also uses auto-provisioning, with PowerPath or a supported host multipathing solution, to manage host access to the data during the migration process.

NDM provides PowerMax user the following benefits:

- Allows non-disruptive migration with hosts and applications online. Potentially supported source storage systems are legacy VMAX v2, VMAX3 series, VMAX All Flash systems, and PowerMax. Potentially supported target systems are VMAX3 series, VMAX All Flash systems, and PowerMax.

- Ease of use with control operations that automate the setup and configuration of the migration environment.

- Managed by familiar and simple user interfaces using Solutions Enabler and Unisphere.

- Migrations can be easily canceled and failed back to the source array for any reason prior to commit.

- Completely built in and does not require any additional software or licensing costs.

For more detailed features, see the *Dell EMC PowerMax and VMAX: Non-Disruptive Migration Best Practices and Operational Guide* white paper available on https://www.dellemc.com/.

See Dell EMC Simple Support Matrix available on Dell EMC E-Lab Navigator for host interoperability with various operating system platforms and multipathing software supported for Non-Disruptive migrations.

Use advanced query option on Dell EMC E-lab Navigator for specific configuration search.

PowerMax/VMAX All Flash-Storage provisioning

PowerMax and VMAX All flash series initial configuration is performed by a Dell EMC Customer Engineer (CE) through the PowerMax or VMAX Management Module Control Station (MMCS). The CE will configure the storage arrays settings for each Fibre Channel port. The procedures in this document assume that any switches and storage systems to be used in this configuration have been installed, and that the front-end adapter ports have been connected to the appropriate switch ports and switch zoning is completed.

(i) **Note:** It is recommended to use Acess Logic (ACLX) to mask volumes.

Storage provisioning operations can be accomplished by Solutions enabler or Unisphere for PowerMax software.

Dell EMC Solutions Enabler

Dell EMC Solutions Enabler installation provides your host with SYMAPI, CLARAPI, and STORAPI shared libraries for use by Solutions Enabler applications, and the Symmetrix Command Line Interface (SYMCLI) for use by storage administrators and systems engineers. SYMCLI is a specialized library of UNIX-formatted commands that can be invoked one at a time. It supports single command-line entries and scripts to map and perform control operations on devices and data objects toward the management of your storage complex. It also monitors device configuration and status of devices that make up the storage environment. The target storage environments are typically PowerMax and VMAX arrays.

For detailed procedure, see *Dell EMC Solutions Enabler Array Controls and Management Guide* available on https://www.dell.com/support.

Dell EMC Unisphere for PowerMax

Dell EMC Unisphere for PowerMax enables management and monitoring of PowerMax arrays along with legacy VMAX All Flash and VMAX3 arrays. Unisphere is installed on a dedicated Windows or Linux server, or deployed as a Virtual Appliance (vAPP).

For detailed procedure, see *Dell EMC Unisphere for PowerMax Installation Guide* available on Dell EMC online Support under PowerMax and VMAX technical documents and videos section.

Dell EMC PowerMax/VMAX All Flash/VMAX3 - Director bit settings

Dell EMC PowerMax/VMAX All Flash/VMAX3 supports various director bit settings against different front-end protocol such as Fibre Channel FCP, NVMe over Fibre Channel (NVMeOF), iSCSI, and FCoE. They vary between switch-based connectivity and direct attach connectivity with various operating system types.

For detailed director bit settings, see the Simple Support Matrix available on Dell EMC E-Lab Navigator.

Dell EMC PowerMax Fibre Channel FCP connectivity

Dell EMC PowerMax arrays running with PowerMaxOS Q3 2019 introduces 32 Gb/s Fibre Channel modules support. Earlier versions of PowerMaxOS on PowerMax arrays and HYPERMAX OS on VMAX All Flash arrays supports 16 Gb/s front-end module for host connectivity.

For details, see the *Understanding storage provisioning* and *Understanding Host Management* sections in *Dell EMC Unisphere for PowerMax Product Guide* available on https://www.dellemc.com/ or Dell EMC Oline Support.

Dell EMC PowerMax NVMe-OF connectivity

Dell EMC PowerMax NVMe-OF connectivity is not supported with Windows environment.

Dell EMC PowerMax iSCSI connectivity

Dell EMC PowerMax array supports high-density quad port 10 Gb/s interface module. For detailed iSCSI host connectivity information, see *Dell EMC PowerMax: iSCSI Implementation for Dell EMC Storage Arrays Running PowerMaxOS* available on https://www.dellemc.com/ or Dell EMC Online Support under PowerMax and VMAX All Flash Technical Documentation page.

PowerMAX and VMAX multipathing behavior

The PowerMAX and VMAX series arrays present active accessible LUNs on all paths that are configured to see the LUN.

Multipath handles active accessible LUNs by using the policy *multibus*, which is a round-robin policy that distributes the I/O operations over the available channels. VMAX3 and VMAX All Flash and PowerMAX support ALUA access on Mobility ID devices from HYPERMAX OS 5977.811.784 and later.

MPIO configuration for PowerMAX or VMAX series

This section discusses methods for configuring the PowerMAX or VMAX series arrays for the following:

- RedHat Enterprise Linux (RHEL)
- Oracle Linux and VM server
- SuSE Linux Enterprise server

Red Hat Enterprise Linux

This section details the procedure for configuring native multipath failover for Symmetrix family of arrays in a Red Hat Enterprise Linux host.

About this task

The Red Hat Enterprise Linux native MPIO already contains default configuration parameters for PowerMAX or VMAX series arrays to provide optimal performance in most environments. There is no need to create a device stanza for these arrays unless you want to modify the default behavior. The /etc/multipath.conf file is installed by default when you install the device-mapper-multipath package.

The sample output from *multipath -ll* is as follows:

```
mpathax (3600009700001976002385330303244433) dm-72 EMC     ,SYMMETRIX
size=50G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
```

```
|- 11:0:4:31 sdew 129:128 active ready running
`- 12:0:9:31 sdis 135:192 active ready running
```

Procedure

1. If you want multipath devices to be created as */dev/mapper/mpathn*, set the value of the `user_friendly_name` in *multipath.conf* to **Yes**.

2. Remove # to enable the devnode blacklist. You may want to add the WWID for the Symmetrix VMware Configuration Manager database, as in this example. The VMware Configuration Manager database is a read-only device that an array uses. By blacklisting it, you eliminate any error messages that may occur due to its presence.

```
blacklist {
wwid 35005076718d4224
devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*" devnode "^hd[a-z]"
devnode "^cciss!c[0-9]d[0-9]*"
}
```

3. If you want to check active MPIO configuration, run `multipathd show config`. The device stanza begins with the following:

```
vendor "EMC"
product"Symmetrix"
```

Oracle Linux and VM server

All the Oracle Linux versions using the stock Red Hat kernel, Oracle enhanced Red Hat kernel, or UEK kernel use the same configurations as Red Hat Enterprise Linux.

See RedHat Enterprise Linux (RHEL) for more information.

SuSE Linux Enterprise server

Use this sample procedure to configure native multipath failover (MPIO) for the PowerMAX and VMAX series of arrays on a SuSE Linux Enterprise Server (SLES) host.

About this task

The SuSE native MPIO already contains default configuration parameters for family of Symmetrix arrays to provide optimal performance in most environments. There is no need to create a device stanza for these arrays unless you want to modify the default behavior. The /etc/multipath.conf file is installed by default when you install the device-mapper-multipathpackage.

The sample output from *multipath -ll* is as follows:

```
# multipath -ll
mpathax (36000097000019760023853303033937) dm-35 EMC,SYMMETRIX
size=18G features='2 queue_if_no_path retain_attached_hw_handler'
hwhandler='0' wp=rw
`-+- policy='service-time 0' prio=1 status=active
|- 2:0:1:11 sdai 66:32 active ready running
`- 1:0:1:11 sdbd 67:112 active ready running
```

Procedure

1. If you want multipath devices to be created as */dev/mapper/mpathn*, set the value of the `user_friendly_name` in *multipath.conf* to **Yes**.

2. Remove # to enable the devnode blacklist. You may want to add the WWID for the Symmetrix VMware Configuration Manager database, as in this example. The VMware Configuration Manager database is a read-only device that an array uses. By blacklisting it, you eliminate any error messages that may occur due to its presence.

```
blacklist {
wwid 35005076718d4224
devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*" devnode "^hd[a-z]"
devnode "^cciss!c[0-9]d[0-9]*"
}
```

3. If you want to check active MPIO configuration, run `multipathd show config`. The device stanza begins with the following:

```
vendor "EMC"
product"Symmetrix"
```

PowerMax/VMAX All Flash

This section provides support information about the PowerMax, VMAX All Flash, and VMAX3 series.

Dell EMC PowerMax

Dell EMC PowerMax provides a platform that is ready for current and next generation data center data storage requirements. The PowerMax Family uses NVMe technology for customer application data.

PowerMax is built using a 100% NVMe storage back-end, allowing it to reach optimal IOPS densities by eliminating the flash media issues that are found in traditional SAS and SATA interfaces. NVMe over Fibre Channel 32 Gb front-end connections are supported starting with the PowerMaxOS 5978.444 release running on PowerMax storage systems allows for end-to-end NVMe accessibility, the PowerMaxOS 5978.444 and later releases running on PowerMax storage systems. It also introduces capability for traditional Fibre Channel FCP connections with an operation link speed of 32 Gb. NVMe over Fibre Channel and 32 Gb FCP both require the installation of the 32 Gb Fibre Channel I/O module. The PowerMaxOS 5978.479 and later releases leverages the 32 Gb Fibre Channel I/O module for 32 Git SRDF link connection support. All PowerMaxOS releases maintain downward compatibility with the legacy 16 Gb Fibre Channel I/O module.

There are two PowerMax models as follows:

PowerMax 2000-PowerMax 2000 is designed to provide high efficiency and flexibility to the data center, providing 1.7 million IOPS (8 K RRH) and up to 1 PB of effective capacity in 20U total space. It is the entry NVMe scale out array sold with the Essentials and Pro software packages.

PowerMax 8000-PowerMax 8000 designed to provide optimal scalability, performance, and IOPS density. It can consolidate disparate workloads on a mass scale as 8 Bricks can support 10 million IOPS (8 K RRH) and provide up to 4 PB of effective capacity in just two floor tiles of space. It is the flagship NVMe scale out array sold with the Essentials and Pro software packages.

Dell EMC PowerMax Operating System environment

PowerMaxOS 5978 is the only supported software operating environment for PowerMax storage arrays. It manages the storage and controls communications with the host systems and introduces support of dedupe and other new feature enhancements.

For detailed product specifications including the 32G Frontend I/O and 32G SRDF protocol support information, see Dell EMC PowerMax Family documentation available on https://shop.dellemc.com and *Dell EMC PowerMax Family Product Guide* on Dell EMC Online Support.

Dell EMC VMAX All Flash

Dell EMC VMAX All Flash range of storage arrays use only high-density flash drives. The range contains four models that combine high scale, low latency, and rich data services:

- VMAX 250F with a maximum capacity of 1.16 PBe (Petabytes effective)
- VMAX 450F with a maximum capacity of 2.3 PBe
- VMAX 850F with a maximum capacity of 4.4 PBe
- VMAX 950F with a maximum capacity of 4.42 PBe

Dell EMC VMAX All Flash Operating Environment

Dell EMC VMAX All Flash storage systems are supported with both HYPERMAX OS 5977 and PowerMaxOS 5978, however not all new feature enhancements that are introduced with PowerMax storage systems running PowerMaxOS 5978 would be available with Dell EMC VMAX All Flash storage systems also running PowerMaxOS 5978.

HYPERMAX OS 5977 provides emulations that perform specific data service and control functions in the HYPERMAX environment, and it introduces an open application platform for running data services and provide file system storage with eNAS and embedded management services for Unisphere. The entire feature set available with HYPERMAX OS 5977 running on Dell EMC VMAX All Flash storage systems would be available with PowerMaxOS 5978 running on Dell EMC VMAX All Flash storage systems except FCoE front-end connections.

For detailed product specifications including the Frontend I/O protocol support information, see *VMAX All Flash Product Guide* and *VMAX All Flash: Family Overview* available on https://www.dellemc.com/.

PowerMax/VMAX All Flash/VMAX3 - Device types

There are two classes of devices available on PowerMax/VMAX All Flash/VMAX3 storage arrays:

- Compatibility device - The Compatibility device is the default VMAX device type, advertises T10 SPC-3 compliancy and behaves almost like the legacy VMAX devices.
- Mobility device - The Mobility device is an optionally configurable device type that is supported with HYPERMAX OS 5977.811.784 and later, advertises T10 SPC-4 compliancy, and has numerous INQUIRY and VPD page changes implemented intended to allow for support of future feature enhancements. Implicit Asymmetric Logical Unit Access (ALUA) is advertised by the Mobility device, however the Mobility device is configured into only a single active target port group per PowerMax or VMAX storage system and a second target port group is configured on a remote PowerMax or VMAX system only when the Mobility devices are configured in an SRDF Metro configuration which is restricted only to Windows and Solaris operating system environments for now. For all supported SRDF/Metro Mobility ID/ALUA, see the Support Matrix available on Dell EMC E-Lab Navigator.
  Mobility devices are also not currently supported with NDM migration.

Local replication services

TimeFinder SnapVX

Dell EMC TimeFinder SnapVX creates and manages point-in-time snapshots of critical data that can be used for backups, decision support, and to refresh data warehouse, test, and development environments. SnapVX snapshots do not require target volumes. SnapVX snapshots share back-end allocations with the source volume and other snapshots on the source volume.

TimeFinder SnapVX is supported on VMAX All Flash arrays running HYPERMAX OS 5977 and later, and snapshots are always consistent. Consistency across multiple arrays is achieved when source devices are in a composite group.

SnapVX provides very low impact snapshots and clones for VMAX LUNs. SnapVX supports up to 256 snapshots per source volume, which are tracked as versions with less overhead and simple relationship tracking. Users can assign names to identify their snapshots, and can set automatic expiration dates on each snapshot. SnapVX provides the ability to manage consistent point-in-time copies for storage groups with a single operation. Up to 1024 target volumes can be linked per source volume, providing read/write access as pointers or full-copy clones. TimeFinder in HYPERMAX OS also provides compatibility modes for users who rely on their TimeFinder Mirror, Clone, or VP Snap command scripts. This allows users to leverage their existing scripts while learning how to take advantage of the new features of SnapVX.

For details, see the *TimeFinder SNAPVX Local Replication* document available on PowerMax and VMAX All Flash Technical Documentation page on https://www.dellemc.com/ or https://www.dell.com/support.

Remote Replication Services

SRDF

The Symmetrix Remote Data Facility (SRDF) maintains real-time or near real-time copies of data on a production storage array at one or more remote storage arrays and only one SRDF device mirror can be read/write enabled at any point in time.

SRDF/Metro

HYPERMAX OS 5977.691.684 and Solutions Enabler/Unisphere for VMAX 8.1 first introduced the support for SRDF/Metro for VMAX3 and VMAX All Flash families of storage arrays and all versions of PowerMaxOS on Dell EMC PowerMax arrays.

With SRDF/Metro, the SRDF secondary device is read/write accessible to the host and takes on the external identity of the primary device (geometry, device WWN, and so on). By providing this external identity on the secondary device, both the primary and secondary devices may then appear as a single virtual device across the two SRDF paired arrays for presentation to a single host or host cluster.

With both the local and remote devices being read/write accessible concurrently, the host or hosts (in a cluster) can read and write to both primary and secondary devices with SRDF/Metro ensuring that each copy remains current, consistent, and addressing any write conflicts which may occur between the paired SRDF devices. A single PowerMax/VMAX3/VMAX All Flash array may simultaneously support multiple SRDF groups that are configured for SRDF/Metro operations and multiple SRDF groups that are configured for non-SRDF/Metro operations. PowerMaxOS Q3 2019 release with SRDF/Metro supports online device expansion though which devices could be expanded within an active SRDF/Metro group. These operations can be accomplished using the latest version of Unisphere for PowerMax or Solutions enabler. Both SRDF and SRDF/Metro support FC or GiGE links for connectivity between arrays.

For more detailed information about SRDF/metro configuration, see *Dell EMC SRDF Introduction* and *SRDF/Metro Overview and Best Practices Technical Note* available on https://www.dell.com/support.

Non-Disruptive Migration services

Dell EMC's Non-Disruptive Migration (NDM) allows user to perform online data migrations that are simple and completely non-disruptive to the host and application. NDM is designed to help automate the process of migrating hosts and applications to a new PowerMax array with no downtime. NDM leverages SRDF replication technologies to move the application data to the new array. It also uses auto-provisioning, with PowerPath or a supported host multipathing solution, to manage host access to the data during the migration process.

NDM provides PowerMax user the following benefits:

- Allows non-disruptive migration with hosts and applications online. Potentially supported source storage systems are legacy VMAX v2, VMAX3 series, VMAX All Flash systems, and PowerMax. Potentially supported target systems are VMAX3 series, VMAX All Flash systems, and PowerMax.
- Ease of use with control operations that automate the setup and configuration of the migration environment.
- Managed by familiar and simple user interfaces using Solutions Enabler and Unisphere.
- Migrations can be easily canceled and failed back to the source array for any reason prior to commit.
- Completely built in and does not require any additional software or licensing costs.

For more detailed features, see the *Dell EMC PowerMax and VMAX: Non-Disruptive Migration Best Practices and Operational Guide* white paper available on https://www.dellemc.com/.

See Dell EMC Simple Support Matrix available on Dell EMC E-Lab Navigator for host interoperability with various operating system platforms and multipathing software supported for Non-Disruptive migrations.

Use advanced query option on Dell EMC E-lab Navigator for specific configuration search.

PowerMax/VMAX All Flash-Storage provisioning

PowerMax and VMAX All flash series initial configuration is performed by a Dell EMC Customer Engineer (CE) through the PowerMax or VMAX Management Module Control Station (MMCS). The CE will configure the storage arrays settings for each Fibre Channel port. The procedures in this document assume that any switches and storage systems to be used in this configuration have been installed, and that the front-end adapter ports have been connected to the appropriate switch ports and switch zoning is completed.

(i) **Note:** It is recommended to use Acess Logic (ACLX) to mask volumes.

Storage provisioning operations can be accomplished by Solutions enabler or Unisphere for PowerMax software.

Dell EMC Solutions Enabler

Dell EMC Solutions Enabler installation provides your host with SYMAPI, CLARAPI, and STORAPI shared libraries for use by Solutions Enabler applications, and the Symmetrix Command Line Interface (SYMCLI) for use by storage administrators and systems engineers. SYMCLI is a specialized library of UNIX-formatted commands that can be invoked one at a time. It supports single command-line entries and scripts to map and perform control operations on devices and data objects toward the management of your storage complex. It also monitors device configuration and status of devices that make up the storage environment. The target storage environments are typically PowerMax and VMAX arrays.

For detailed procedure, see *Dell EMC Solutions Enabler Array Controls and Management Guide* available on https://www.dell.com/support.

Dell EMC Unisphere for PowerMax

Dell EMC Unisphere for PowerMax enables management and monitoring of PowerMax arrays along with legacy VMAX All Flash and VMAX3 arrays. Unisphere is installed on a dedicated Windows or Linux server, or deployed as a Virtual Appliance (vAPP).

For detailed procedure, see *Dell EMC Unisphere for PowerMax Installation Guide* available on Dell EMC online Support under PowerMax and VMAX technical documents and videos section.

Dell EMC PowerMax/VMAX All Flash/VMAX3 - Director bit settings

Dell EMC PowerMax/VMAX All Flash/VMAX3 supports various director bit settings against different front-end protocol such as Fibre Channel FCP, NVMe over Fibre Channel (NVMeOF), iSCSI, and FCoE. They vary between switch-based connectivity and direct attach connectivity with various operating system types.

For detailed director bit settings, see the Simple Support Matrix available on Dell EMC E-Lab Navigator.

Dell EMC PowerMax Fibre Channel FCP connectivity

Dell EMC PowerMax arrays running with PowerMaxOS Q3 2019 introduces 32 Gb/s Fibre Channel modules support. Earlier versions of PowerMaxOS on PowerMax arrays and HYPERMAX OS on VMAX All Flash arrays supports 16 Gb/s front-end module for host connectivity.

For details, see the *Understanding storage provisioning* and *Understanding Host Management* sections in *Dell EMC Unisphere for PowerMax Product Guide* available on https://www.dellemc.com/ or Dell EMC Oline Support.

Dell EMC PowerMax NVMe-OF connectivity

Dell EMC PowerMax NVMe-OF connectivity is not supported with Windows environment.

Dell EMC PowerMax iSCSI connectivity

Dell EMC PowerMax array supports high-density quad port 10 Gb/s interface module. For detailed iSCSI host connectivity information, see *Dell EMC PowerMax: iSCSI Implementation for Dell EMC Storage Arrays Running PowerMaxOS* available on https://www.dellemc.com/ or Dell EMC Online Support under PowerMax and VMAX All Flash Technical Documentation page.

Midrange Storage

This section describes host connectivity of the Dell EMC Midrange storage arrays.

Storage configuration

General recommendations for storage pools

Dell EMC Unity supports two types of storage pools: traditional pools and dynamic pools. The following recommendations are applicable to both types of pool.

Dell EMC recommends using fewer storage pools within Dell EMC Unity to reduce complexity and increase flexibility. However, it may be appropriate to configure multiple storage pools to:

- Separate workloads with different I/O profiles
- Separate pools where FAST Cache is active and not active
- Dedicate resources to meet specific performance goals
- Separate resources for multi-tenancy
- Create smaller failure domains

Storage pool capacity

Storage pool capacity is used for multiple purposes:

- To store all data written into storage objects (LUNs, file systems, data stores, and VVols) in that pool.
- To store data that is needed for Snapshots of storage objects in that pool.
- To track changes to replicated storage objects in that pool
- To perform efficient data relocations for FAST VP

Storage pools must maintain free capacity to operate properly. By default, Dell EMC Unity will raise an alert if a storage pool has less than 30% free capacity, and will begin to automatically invalidate Snapshots and Replication sessions if the storage pool has less than 5% free capacity. Dell EMC recommends that a storage pool always have at least 10% free capacity.

Raid protection

Dell EMC Unity applies RAID protection to the storage pool to protect user data against drive failures. Choose the RAID type that best suits your needs for performance, protection, and cost.

- RAID-1/0 provides the highest level of performance from a given set of drive resources, with the lowest CPU requirements; however, only 50% of the total drive capacity is usable.
- RAID-5 provides the best usable capacity from a set of drive resources, but at lower overall performance and availability than RAID-1/0.
- RAID-6 provides better availability than RAID-5 and better usable capacity than RAID-1/0, but has the lowest performance potential of the three RAID types.

Traditional pools

Traditional Storage Pools apply RAID protection to individual groups of drives within the storage pool. Traditional pools are the only type of pool available on Dell EMC Unity hybrid systems, and are also available on all-Flash systems.

Raid protection

For traditional pools, Dell EMC recommends RAID-5 for drives in Extreme Performance and Performance tiers, and RAID-6 for drives in the Capacity tier.

Assuming that roughly the same number of drives will be configured in a traditional pool, Dell EMC recommends smaller RAID widths as providing the best performance and availability, at the cost of slightly less usable capacity.

Example: When configuring a traditional pool tier with RAID-6, use 4+2 or 6+2 as opposed to 10+2 or 14+2.

When choosing RAID-1/0, 1+1 can provide better performance with the same availability and usable capacity as larger RAID widths (assuming that the same total number of drives are used), and also provides more flexibility.

All-flash pool

All-flash pools provide the highest level of performance in Dell EMC Unity. Use an all-flash pool when the application requires the highest storage performance at the lowest response time.

Snapshots and Replication operate most efficiently in all-flash pools. Data Reduction is only supported in an all-flash pool.

FAST Cache and FAST VP are not applicable to all-flash pools.

Dell EMC recommends using only a single drive size and a single RAID width within an all-flash pool.

Dynamic pools

Dynamic Storage Pools apply RAID protection to groups of drive extents from drives within the pool, and allow for greater flexibility in managing and expanding the pool. Dynamic pools are only available on Dell EMC Unity all-Flash systems, and therefore must be all-Flash pools; dynamic pools cannot be built with HDDs.

RAID protection

At the time of creation, dynamic pools use the largest RAID width possible with the number of drives that are specified, up to the following maximum widths:

- RAID-1/0: 4+4
- RAID-5: 12+1
- RAID-6: 14+2

With dynamic pools, there is no performance or availability advantage to smaller RAID widths. To maximize usable capacity with parity RAID, Dell EMC recommends to initially create the pool with enough drives to guarantee the largest possible RAID width.

- For RAID-5, initially create the pool with at least 14 drives.
- For RAID-6, initially create the pool with at least 17 drives.

Spare capacity

Hot spares are not needed with dynamic pools. A dynamic pool automatically reserves the capacity of one drive, as spare space in the pool, for every 32 drives. If a drive fails, the data that was on the failed drive is rebuilt into the spare capacity on the other drives in the pool. Also, unbound drives of the appropriate type can be used to replenish the spare capacity of a pool, after the pool rebuild has occurred.

Example: For an All-Flash pool, use only 1.6 TB SAS Flash 3 drives, and configure them all with RAID-5 8+1.

Hybrid pool

Hybrid pools can contain HDDs (SAS and NL-SAS drives) and flash drive, and can contain more than one type of drive technology in different tiers. Hybrid pools typically provide greater capacity at a lower cost than all-flash pools, but also typically have lower overall performance and higher response times. Use hybrid pools for applications that do not require consistently low response times, or that have large amounts of mostly inactive data.

Performance of a hybrid pool can be improved by increasing the amount of capacity in the flash drive tier, so that more of the active dataset resides on and is serviced by the flash drives. See the FAST VP section.

Hybrid pools can have up to three tiers (Extreme Performance, Performance, and Capacity). Dell EMC recommends using only a single drive speed, size, and RAID width within each tier of a hybrid pool.

Example:

- For the Extreme Performance tier, use only 800 GB SAS flash 2 drives, and configure them all with RAID-5 8+1.

- For the Performance tier, use only 1.2 TB SAS 10K RPM drives, and configure them with RAID-5 4+1.

- For the Capacity tier, use only 6 TB NL-SAS drives, and configure them all with RAID-6 6+2.

Storage object types

By default, Dell EMC Unity creates thin storage objects. Thin storage objects are virtually provisioned and space efficient. In general, Dell EMC recommends using thin storage objects, as they provide the best capacity utilization, and are required for most features. Thin storage objects are recommended when any of the following features will be used:

- Data Reduction

- Snapshots

- Thin Clones

- Asynchronous Replication

Thick storage objects will reserve capacity from the storage pool, and dedicate it to that particular storage object. Thick storage objects guarantee that all advertised capacity is available for that object. Thick storage objects are not space efficient, and therefore do not support the use of space-efficient features. If it is required to enable a space-efficient feature on a thick storage object, it is recommended to first migrate the thick storage object to a thin storage object, and enable the feature during the migration (for Data Reduction) or after migration has completed (for Snapshots, Thin Clones, and Asynchronous Replication).

In addition to capacity for storing data, storage objects also require pool capacity for metadata overhead. The overhead percentage is greater on smaller storage objects. For better capacity utilization, Dell EMC recommends configuring storage objects that are at least 100GB in size, and preferably at least 1TB in size.

Features

FAST VP

Fully Automated Storage Tiering (FAST) for Virtual Pools (VP) accelerates performance of a specific storage pool by automatically moving data within that pool to the appropriate drive technology, based on data access patterns. FAST VP is applicable to hybrid pools only within a Dell EMC Unity hybrid system.

The default and recommended FAST VP policy for all storage objects is **Start High then Auto-tier**. This policy places initial allocations for the storage object in the highest tier available, and monitors activity to this storage object to determine the correct placement of data as it ages.

FAST VP is most effective if data relocations occur during or immediately after normal daily processing. Dell EMC recommends scheduling FAST VP relocations to occur before backups or nightly batch processing. For applications which are continuously active, consider configuring FAST VP relocations to run constantly.

Dell EMC recommends maintaining at least 10% free capacity in storage pools, so that FAST VP relocations can occur efficiently. FAST VP relocations cannot occur if the storage pool has no free space.

FAST Cache

FAST Cache is a single global resource that can improve performance of one or more hybrid pools within a Dell EMC Unity hybrid system. FAST Cache can only be created with SAS Flash 2 drives, and is only applicable to hybrid pools. Dell EMC recommends to place a Flash tier in the hybrid pool before configuring FAST Cache on the pool. FAST Cache can improve access to data that is resident in the HDD tiers of the pool.

Enable FAST Cache on the hybrid pool if the workload in that pool is highly transactional, and has a high degree of locality that changes rapidly.

For applications that use larger I/O sizes, have lower skew, or do not change locality as quickly, it may be more beneficial to increase the size of the Flash tier rather than enable FAST Cache.

FAST Cache can increase the IOPS achievable from the Dell EMC Unity system, and this will most likely result in higher CPU utilization (to service the additional I/O). Before enabling FAST Cache on additional pools or expanding the size of an existing FAST Cache, monitor the average system CPU utilization to determine if the system can accommodate the additional load. See Table 3 for recommendations.

Data Reduction

Dell EMC Unity Data Reduction by compression is available for Block LUNs and VMFS datastores in an all-flash pool starting with Dell EMC Unity OE 4.1. Data reduction via compression is available for file systems and NFS datastores in an all-flash pool starting with Dell EMC Unity OE 4.2. Beginning with Dell EMC Unity OE 4.3, data reduction includes both compression and deduplication.

Be aware that data reduction increases the overall CPU load on the system when storage objects service reads or writes of reduceable data, and may increase latency. Before enabling data reduction on a storage object, Dell EMC recommends to monitor the system and ensure that the system has available resources to support data reduction (See Table 3 to the Hardware Capability Guidelines). Enable data reduction on a few storage objects at a time, and then monitor the system to be sure it is still within recommended operating ranges, before enabling data reduction on more storage objects.

For new storage objects, or storage objects that are populated by migrating data from another source, Dell EMC recommends to create the storage object with data reduction enabled, before writing any data. This provides maximum space savings with minimal system impact.

Advanced Deduplication

Dell EMC Unity Advanced Deduplication is an optional extension to Data Reduction, that you can enable to increase the capacity efficiency of data reduction enabled storage objects. Beginning with Dell EMC Unity OE 4.5, advanced deduplication is available for storage objects in dynamic pools on Dell EMC Unity 450F, 550F, and 650F All-Flash systems.

As with data reduction, advanced deduplication is only applied to data when it is written to the storage object. LUN Move can be utilized to deduplicate existing data on Block storage objects.

For new storage objects, or storage objects that will be populated by migrating data from another source, it is recommended to create the storage object with advanced deduplication enabled, before writing any data. This provides maximum space savings with minimal system impact.

Snapshots

Dell EMC recommends including a Flash tier in a hybrid pool where snapshots will be active.

Snapshots increase the overall CPU load on the system, and increase the overall drive IOPS in the storage pool. Snapshots also use pool capacity to store the older data being tracked by the snapshot, which increases the amount of capacity used in the pool, until the snapshot is deleted. Consider the overhead of snapshots when planning both performance and capacity requirements for the storage pool.

Before enabling snapshots on a storage object, it is recommended to monitor the system and ensure that existing resources can meet the additional workload requirements (See Table 2 for Hardware Capability Guidelines). Enable snapshots on a few storage objects at a time, and then monitor the system to be sure it is still within recommended operating ranges, before enabling more snapshots.

Dell EMC recommends to stagger snapshot operations (creation, deletion, and so on). This can be accomplished by using different snapshot schedules for different sets of storage objects. It is also recommended to schedule snapshot operations after any FAST VP relocations have completed.

Snapshots are deleted by the system asynchronously; when a snapshot is in the process of being deleted, it will be marked as *Destroying*. If the system is accumulating Destroying snapshots over time, it may be an indication that existing snapshot schedules are too aggressive; taking snapshots less frequently may provide more predictable levels of performance. Dell EMC Unity will throttle snapshot delete operations to reduce the impact to host I/O. Snapshot deletes will occur more quickly during periods of low system utilization.

Thin Clones

Dell EMC recommends including a flash tier in a hybrid pool where thin clones will be active.

Thin clones use snapshot technology to provide space-efficient clones of block objects. Consider the overhead of snapshots when planning performance and capacity requirements for a storage pool which will have thin clones.

Asynchronous replication

Dell EMC recommends including a Flash tier in a hybrid pool where asynchronous replication is active. This is applicable to both the source and the destination pools.

Dell EMC recommends configuring multiple replication interfaces per SP, and distributing replication sessions across them. Link Aggregation Control Protocol (LACP) can also be used to aggregate bandwidth for a replication interface. Configure Jumbo frames (MTU 9000) when possible.

Asynchronous replication takes snapshots on the replicated storage objects to create the point-in-time copy, determine the changed data to transfer, and maintain consistency during the transfer. Consider the overhead of snapshots when planning performance and capacity requirements for a storage pool that has replicated objects.

When possible, fill the source storage object with data before creating the replication session. The data will then be transmitted to the destination storage object during initial synchronization. This is typically the fastest way to populate the destination storage object with asynchronous replication.

Setting smaller RPO values on replication sessions will not make them transfer data more quickly; but smaller RPOs result in more frequent snapshot operations. Choosing larger RPOs, or manually synchronizing during nonproduction hours, may provide more predictable levels of performance.

Synchronous replication/Metrosync for file

Dell EMC recommends including a Flash tier in a hybrid pool where synchronous replication will be active. This is applicable to both the source and the destination pools.

Synchronous replication transfers data to the remote system over the first Fibre Channel port on each SP. When planning to use synchronous replication, it may be appropriate to reduce the number of host connections on this port. When the CNA ports are configured as FC, CNA port 4 is defined as the replication port. If the CNA ports are configured as Ethernet, then port 0 of the lowest numbered FC I/O Module is the replication port.

When possible, create the synchronous replication session before filling the source storage object with data, as this alleviates the need to perform initial synchronization of the replication session. This is typically the fastest way to populate the destination storage object with synchronous replication.

When sizing a disaster recovery solution using synchronous replication, consider provisioning a destination system that has similar performance capabilities as the source system. This can help maintain the same level of application performance after a failover event.

SAN Copy

SAN Copy provides one-time migration of Block resources from a third-party array, using either iSCSI or FC connections. When using FC, note that SAN Copy must use different ports than the FC ports which are designated for Synchronous Replication. This is true even if Synchronous Replication is not actively being used.

To lessen the impact of SAN Copy migrations on other host activity, consider reducing the number of host connections on the FC ports used for SAN Copy.

NDMP

Dell EMC Unity supports 2-way NDMP for file data, which enables the system to send file data directly to a backup device using FC connections. Make sure that NDMP uses different ports than the FC ports which are designated for Synchronous Replication. This is true even if Synchronous Replication is not actively being used.

To lessen the impact of 2-way NDMP backups on other host activity, consider reducing the number of host connections on the FC ports that are used for NDMP.

Data at Rest Encryption

Data at Rest Encryption (D@RE) is Controller Based Encryption that does not impact performance; therefore Dell EMC recommends ordering Dell EMC Unity systems as encryption-enabled, if appropriate for your environment.

(i) Note: Encryption can only be enabled at the time of system installation with the appropriate license.

If encryption is enabled, Dell EMC recommends making external backups of the encryption keys after system installation, and immediately following any change in the system's drives (such as, creating or expanding a storage pool, adding new drives, replacing a faulted drive, and so on).

Host I/O limits

Dell EMC recommends setting Host I/O Limits on workloads which might monopolize pool resources and starve other applications of their required performance. Consider some of the following opportunities to utilize Host I/O Limits:

- Limit the bandwidth available to large-block applications, such as backup jobs, which may be increasing the latency on other small-block workloads.
- Limit the IOPS capability of Thin Clones which are used in Test/Dev environments, so that they do not impact the capability of the associated source objects.
- Limit the IOPS / bandwidth that is available to non-essential applications that are hosted on the same Dell EMC Unity system as your critical applications.

Application considerations

Host alignment for block LUNs

Alignment only needs to be done for host operating systems which still use a 63-block disk header. If alignment is required, perform the alignment using a host-based method, and align with a 1MB offset.

See the *Host Connectivity Guide* on Dell EMC Online Support to determine if alignment is required for your operating system, and how to perform the alignment.

Hosts

Under the **ACCESS** category in the main navigation menu, users can configure hosts (Windows or Linux/UNIX) for storage access. VMware hosts can be configured on the VMware (Hosts) page. Before a network host can access block storage or NFS file systems, the user must define a configuration for the host and associate it with a storage resource. SMB file systems can automatically be accessed by authorized users once provisioned. Users can use the Hosts page, as shown in the following figure to configure host configurations. This can be done on an individual host-by-host basis or through subnet and netgroup configurations that allow access to multiple hosts or network segments. For block resources, before the user starts to configure a host, the user should ensure that initiator interfaces are configured and initiator registration completed. Once a host configuration is completed, users can go to the properties of a storage resource and specify the hosts, subnets, or netgroups from which they want the resource to be accessed.

Figure 17 Hosts

VMware (ACCESS)

The VMware host access page is specifically for VMware ESXi hosts and their associated vCenter servers. Unisphere provides VMware discovery capabilities through the VMware page, as shown in

the following figure. These discovery capabilities collect virtual machine and datastore storage details from vSphere and display them in the context of the storage system. Imported VMware hosts automatically register their initiators, allowing for ease of management. The vCenters tab allows users to add a vCenter and associated ESXi hosts in a single workflow, while the ESXi hosts tab allows users to add standalone ESXi hosts as needed. The Virtual Machines tab and Virtual Drives tab display imported information about virtual machines and their VMDKs from any added ESXi host.

For more information about VMware access and integration capabilities, see the *Dell EMC Unity: Virtualization Integration white paper* on Dell EMC Online Support.

Figure 18 VMware (Access)

Initiators

To ensure that hosts can access block storage resources, the user must register initiators between the storage system and configured hosts. On the Initiators page, as shown in the following figure, users can manually register one or more Fibre Channel or iSCSI initiators. Initiators are endpoints from which Fibre Channel and iSCSI sessions originate, where each initiator is uniquely identified by its World Wide Name (WWN) or iSCSI Qualified Name (IQN). The link between a host initiator and a target port on the storage system is called the initiator path. Each initiator can be associated with multiple initiator paths. The Initiator Paths tab shows all data paths that are currently available to the initiators connected to the system either by FC or iSCSI. For iSCSI paths to show up, iSCSI interfaces must be configured on the Block Page. These initiators can then be discovered and registered by hosts using the iSCSI initiator tool (that is, the Microsoft iSCSI Initiator). For Fibre Channel paths, FC zoning on the appropriate switch is needed for the initiator paths to be seen as available by the system. Once the paths are available, users can configure their connected hosts on the Hosts Page.

Figure 19 Initiators

With the release of Dell EMC Unity OE version 4.3, Initiators can now have advanced settings customized through Unisphere. In order to access these settings, select an Initiator and then click the pencil icon to bring up the Edit Initiator window. Clicking Advanced at the bottom to reveal the Initiator Source Type, Fail-over Mode, Unit Serial Number, and LunZ Enabled settings, as shown in the following figure. For more information about configuring Host Initiator Parameters, please reference the Online Help through Unisphere.

Figure 20 Edit Initiator - Advanced

Snapshot schedule

Dell EMC Unity enables you to take point-in-time snapshots for all storage resources (block or file) to meet protection and recovery requirements in the event of corruption or accidental deletion.

The Snapshot Schedule page, as shown in the figure , enables users to set the system to periodically take snapshots of storage resources automatically. Automating these administrative tasks takes away some of the management aspects of data protection. After enabling a snapshot schedule for a resource, each snapshot that is taken is time-stamped with the date and time of when it was created, and contains a point-in-time image of the data in the storage resource. The default snapshot schedules available on the system are:

*   Default protection - A snapshot is taken at 08:00 (UTC) every day, and the snapshot is retained for 2 days.

*   Protection with shorter retention - A snapshot is taken at 08:00 (UTC) every day, and the snapshot is retained for 1 day.

*   Protection with longer retention - A snapshot is taken at 08:00 (UTC) every day, and the snapshot is retained for 7 days.

> (i) **Note:** Times are displayed in a user's local time in a 12-hour format and that default snapshot schedules cannot be modified; but custom snapshot schedules can be configured by selecting the intervals, times, and days for the system to take snapshots regularly.

With the Dell EMC Unity OE version 4.4 or later, user-defined Snapshot Schedules can be replicated using the Synchronous Replication connection that is established between two physical systems. Reference the new **Sync Replicated** column in the Snapshot Schedule page, as shown in the following figure. Applying a replicated Snapshot Schedule is only enabled in synchronously replicated file resources.

Figure 21 Snapshot schedule

For more information about the snapshot technology available on Dell EMC Unity systems, see the *Dell EMC Unity: Snapshots and Thin Clones* and *Dell EMC Unity: MetroSync for File* white papers on Dell EMC Online Support.

Front-end connectivity

Dell EMC Unity provides multiple options for front-end connectivity, using on-board ports directly on the DPE, and using optional I/O Modules. This section discusses recommendations for the different types of connectivity.

In general, front-end ports need to be connected and configured symmetrically across the two storage processors (SPs), to facilitate high availability and continued connectivity if there is SP failure.

Example - A NAS Server is configured so that NAS clients connect using port 0 of the first I/O Module on SPA; therefore port 0 of the first I/O Module on SPB must be cabled so that it is accessible to the same networks.

For best performance, Dell EMC recommends using all front-end ports that are installed in the system, so that workload is spread across as many resources as possible.

Example - configuring the 4-port Fibre Channel I/O Module, zone different hosts to different ports so that all eight ports across the 2 SPs are used; do not zone all hosts to the first port of each I/O Module.

Fibre Channel

When configured for Fibre Channel (FC), Dell EMC Unity CNA ports and I/O Module ports can be configured with 8 GB or 16 GB SFPs. All FC ports can negotiate to lower speeds. 16 GB FC is recommended for the best performance.

Dell EMC recommends single-initiator zoning when creating zone sets. For high availability purposes, a single host initiator should be zoned to at least one port from SPA and one port from SPB. For load balancing on a single SP, the host initiator can be zoned to two ports from SPA and two ports from SPB. When zoning additional host initiators, zone them to different SP ports when possible, to spread the load across all available SP ports.

Utilize multipathing software on hosts that are connected using FC, such as Dell EMC PowerPath, which coordinates with the Dell EMC Unity system to provide path redundancy and load balancing.

iSCSI

Dell EMC Unity supports iSCSI connections on multiple 1 Gb/s and 10 GB/s port options. 10GBase-T ports can autonegotiate to 1 GB/s speeds. 10 GB/s is recommended for the best performance. If possible, configure Jumbo frames (MTU 9000) on all ports in the end-to-end network, to provide the best performance.

To achieve optimal iSCSI performance, use separate networks and VLANs to segregate iSCSI traffic from normal network traffic. Configure standard 802.3x Flow Control (Pause or Link Pause) on all iSCSI Initiator and Target ports that are connected to the dedicated iSCSI VLAN.

Dell EMC Unity supports 10 GbE and 1GBase-T ports that provide iSCSI offload. Specifically, the CNA ports (when configured as 10GbE or 1GBase-T) and the 2-port 10GbE I/O Module ports provide iSCSI offload. Using these modules with iSCSI can reduce the protocol load on SP CPUs by 10-20%, so that those cycles can be used for other services.

Utilize multipathing software on hosts that are connected using iSCSI, such as Dell EMC PowerPath, which coordinates with the Dell EMC Unity system to provide path redundancy and load balancing.

Network-attached storage (NAS)

Dell EMC Unity supports NAS (NFS, FTP, and/or SMB) connections on multiple 1 GB/s and 10 GB/s port options. 10GBase-T ports can auto-negotiate to 1 GB/s speed. 10 GB/s is recommended for the best performance. If possible, configure Jumbo frames (MTU 9000) on all ports in the end-to-end network, to provide the best performance.

Dell EMC recommends configuring standard 802.3x Flow Control (Pause or Link Pause) on all storage ports, switch ports, and client ports that are used for NAS connectivity.

Dell EMC Unity provides network redundancy for NAS using Link Aggregation Control Protocol (LACP) and Fail-Safe Networking (FSN). Combine FSN and LACP with redundant switches to provide the highest network availability. In addition to redundancy, LACP can also improve performance with multiple 1GBase-T connections, by aggregating bandwidth. LACP can be configured across any Ethernet ports that have the same speed, duplex, and MTU.

(i) Note: LACP cannot be enabled on ports that are also used for iSCSI connections.

While LACP creates a link aggregation with multiple active links, FSN provides redundancy by configuring a primary link and a standby link. The standby link is inactive unless the entire primary link fails. If FSN is configured with links of different performance capability (such as a link aggregation of 10 GB/s ports, and a stand-alone 1 GB/s port), Dell EMC recommends that you configure the highest performing link as the primary.

NAS Servers are assigned to a single SP. All file systems that are serviced by that NAS Server have I/O processed by the SP on which the NAS Server is resident. For load-balancing, Dell EMC recommends that you create at least two NAS Servers per Dell EMC Unity system: one on SPA, and one on SPB. Assign file systems to each NAS Server such that front-end workload is approximately the same for each SP.

Connectivity Options

The following tables provide maximum expected IOPS and bandwidth from the different ports that are available in the Dell EMC Unity system. (The capability of a port does not guarantee that the system can reach that level, nor does it guarantee that performance will scale with additional ports. System capabilities are highly dependent on other configuration parameters.)

SAS ports are used by the SPs to move data to and from the back-end drives; all other ports can be used to provide access to hosts.

Table 3 provides maximum expected IOPS and bandwidth from a 12Gb SAS port. The base Dell EMC Unity configuration contains four ports.

The following tables provide maximum expected IOPS from the different ports that are available in the Dell EMC Unity system. (The capability of a port does not guarantee that the system can reach that level, nor does it guarantee that performance will scale with additional ports. System capabilities are highly dependent on other configuration parameters.)

Table 27 Ports and expected IOPS

Port	Maximum IOPS per Port
16 GB FC CNA or 4-port I/O Module	45,000
8 GB FC CNA	45,000
10 GbE iSCSI CNA or 2-port I/O Module	25,000
10 GbE iSCSI 4-port I/O Module	30,000
10GBase-T iSCSI On-board or 4-port I/O Module	30,000
1GBase-T iSCSI CNA, On-board or 4-port I/OModule	3,000

The following table provides maximum expected IOPS from the front-end ports which provide File protocols (NFS and SMB).

Table 28 Front-end ports and expected IOPS

Port	Maximum IOPS per Port
10 GbE NAS CNA or 2-port I/O Module	60,000
10 GbE NAS 4-port I/O Module	60,000
10GBase-T NAS On-board or 4-port I/O Module	60,000
1GBase-T NAS CNA, On-board or 4-port I/OModule	6,000

Unity series and VNX series behavior

Unity series and VNX series systems support active/passive and ALUA access. The arrays present configured LUNs as:

- *Active* on the paths that are connected to the service processor that is the current owner of the LUN.

- *Passive* on the paths that are connected to the other service processor.

The default owner may or may not be the current owner of the LUN. In such an event, DM-MPIO attempts to trespass the LUN to the default owner. If the default owner of the LUN is not accessible from the host, the LUN is trespassed to the other storage processor.

VNX and Unity arrays support two types of asymmetric logical unit path management types. (See ALUA supported Linux operating systems, for your supported type.) The two types of Asymmetric Logical Unit Access (ALUA) are *explicit* and *implicit*, explained briefly as follows:

- Explicit Asymmetric Logical Unit Access
  SCSI target devices can set the target port group asymmetric access state of each target port group using the `SCSI Set Target Port Groups` command. SCSI target devices must have explicit asymmetric logical unit access management capabilities to do so.

- Implicit Asymmetric Logical Unit Access
  SCSI target devices can set the target port group asymmetric access state of each target port group using mechanisms other than the `SCSI Set Target Port Groups` command. SCSI target devices must have explicit asymmetric logical unit access management capabilities to do so.

The ALUA feature allows one port to differ from other ports that are connected to the same target device. The access implementation allows target devices with multiple target ports to be implemented in separate physical groupings. They each have designated but differing levels of access to the target ports.

The explicit ALUA method of implementation allows the existing MPIO driver to send non-ALUA commands and task management functions to a logical unit as if it were not an ALUA device. As a result the ALUA-based logical unit is treated like any non-ALUA logical unit.

The improvement of using ALUA in the *implicit* case is that no host-based trespass is required to access a logical unit should a preferred path fail.

The Service Processor (SP) ownership requirement for a target device has not changed, however. Commands and task management functions can be routed to a logical unit through any available target port. The drawback of using a nonpreferred port to access a target device is that the performance may not be optimal. Also, when a logical unit is being accessed through a nonpreferred port for an extended time, the array attempts to trespass the logical unit so that the port is on the nonpreferred SP.

This Table lists ALUA support in the Linux operating systems:

Table 29 ALUA supported Linux operating systems

Linux distribution	Implicit ALUA support [a]	Explicit ALUA support [b]
Red Hat Enterprise Linux	Red Hat Enterprise Linux 5.1 or later Red Hat Enterprise Linux 6.0 or later RHEL7.0 and later RHEL8.0 and later	Red Hat Enterprise Linux 6.0 Red Hat Enterprise Linux 7.0 or later Red Hat Enterprise Linux 8.0 and later
SuSE Linux Enterprise Server	SLES 10 SP1 or later SLES 11 or later SLES 12 or later	SLES 10 SP 2 or later SLES 11 or later SLES 12 or later
Oracle Linux and equivalent VM server	OL 5.1 and later OL6.0 and later OL7.0 and later OL8.0 and later	OL6.1 and later OL7.0 and later OL8.0 and later

Table 29 ALUA supported Linux operating systems (continued)

a. Does not require any alteration to the /etc/multipath.conf file.
b. Requires alteration to the /etc/multipath.conf file. For the required changes, see the appropriate MPIO configuration section for the Linux distribution release that you are configuring.

In a target port group, all target ports that support asymmetric access to logical units, will be in one of the following target ports asymmetric access states. This is with regard to their ability to access a particular logical unit:

- Active/Active
  While in an active/active state the target port group could access the logical unit. All commands operate exactly as specified in the appropriate command set standards.

- Active/Enabled
  While in the active/enabled state the device server shall support all commands that the logical unit supports. These commands shall operate exactly as specified in the appropriate command set standards. The execution of certain commands, may operate with lower performance than they would if the target port group were in the active/non-optimized state. It is true for commands involving data transfer or caching.

- Active/Standby
  While in the standby state all target ports in a target port group can perform a limited set of commands. The standby state is intended to create a state from which it should be possible to provide a higher level of accessibility, to a logical unit by transitioning to either the active/active or active/enabled states.

Commands that operate in the standby state are necessary for:

- Diagnosing and testing the logical unit and its paths.

- Identifying the path.

- Identifying the logical unit.

- Determining the operational state of the logical unit.

- Determining the active/inactive state of the unit.

- Manage or delete logical unit or element reservations.

- Testing service delivery subsystem.

The *group_by_prio* policy is used to handle these and other ALUA-specific scenarios. It is used along with the priority that is determined by the /sbin/mpath_prio_emc utility and ancillary function calls within the multipath framework. As part of the *group-by-prio* policy, the I/O is distributed across all active paths to the LUN in a round-robin manner.

Use of the recommended Dell EMC parameters in /etc/multipath.conf automatically handles implicit/explicit ALUA and other policy decisions. Unless specific site conditions require alterations, the default values should be used.

MPIO configuration for Unity storage and VNX Unified Storage

This section contains the following information for configuring the Unity series and VNX (Unified Storage) series:

- Blacklisting the Unity series and VNX series LUNZ

- Failover mode

- RedHat Enterprise Linux (RHEL)

- Red Hat Linux 5.0

- Red Hat Linux 6.0
- Red Hat Linux 7.0
- Red Hat Enterprise Linux 7.2 and later
- Oracle Linux and VM server
- SuSE Linux Enterprise server

Blacklisting the Unity series or VNX series LUNZ

When zoned to a Unity series or VNX series, a host may not boot if a LUNZ is exposed to the system. This is fixed by allowing DM-MPIO to blacklist and skip over these devices. To do this, it is necessary to add the following to the device node blacklist stanza:

```
device {
vendor "DGC" product "LUNZ"
}
```

An example of this stanza follows:

```
blacklist {
## Replace the wwid with the output of the command
## 'scsi_id -g -u -s /block/[internal scsi disk name]' ## Enumerate the wwid
for all internal scsi disks.
## Optionally, the wwid of VCM database may also be listed here. wwid
20010b9fd080b7321
devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*" devnode "^hd[a-z][[0-9]*]"
devnode "^cciss!c[0-9]d[0-9]*[p[0-9]*]"

device {
vendor    "DGC"
product   "LUNZ"
}
}
```

The blacklist parameter name has changed in both Red Hat Enterprise Linux 4.7 and later and Red Hat Enterprise Linux 5.3 and later, to **bl_product**. Substitute **bl_product** for product in the above example.

Failover mode

This section introduces two main failover modes that are supported on a Dell EMC midrange array, PNR, and ALUA. For other failover modes, see Dell EMC product documentation.

PNR

Passive Not Ready (PNR) is configured on the array by selecting failover mode 1. This provides active paths on the service processor (SP) that is the default owner of the LUN. The other SP provides passive Not Ready paths for the same LUN until it is trespassed to that SP.

ALUA

The Unity series and VNX series support Asymmetrical Logical Unit Access (ALUA) when configured for failover mode 4. See Unity series and VNX series behavior for more information about this functionality.

(i) Note: Dell EMC does not support having PNR and ALUA devices that are attached to the same host for ALUA devices in VNX series storage systems. Unity series does not support PNR mode.

Red Hat Enterprise Linux

The section discusses methods for configuring the Unity series and VNX (Unified Storage) series storage arrays. Red Hat Enterprise Linux multipath configuration natively contains stanzas to support these storage arrays. If you want to use ALUA failover mode, changes are required to the *multipath.conf* file.

Red Hat Linux 5.0 (and point releases)

The device stanza in discussed in this section is the default PNR stanza that is included in Red Hat Enterprise Linux and does not need to be added. The default stanza is as follows:

```
## Use user friendly names, instead of using WWIDs as names. defaults {
user_friendly_names yes
}
# Device attributes for EMC VNX series PNR device {
vendor "DGC"
product "*"
path_grouping_policy group_by_prio
getuid_callout "/sbin/scsi_id -g -u -s /block/%n" prio_callout "/sbin/
mpath_prio_emc /dev/%n" path_checker emc_clariion
path_selector "round-robin 0"
features "1 queue_if_no_path"
no_path_retry 300
hardware_handler "1 emc" failback immediate
}
```

The sample output from *multipath -ll* is as follows:

```
mpath3 (36006016005d01800b207271bb8ecda11) DGC,RAID 5
[size=3.1G][features=1 queue_if_no_path][hwhandler=1 emc]
\_ round-robin 0 [prio=2][undef]
\_ 5:0:0:21 sdar 66:176 [undef][ready]
\_ 1:0:0:21 sdj 8:144 [undef][ready]
\_ round-robin 0 [prio=0][undef]
\_ 4:0:0:21 sdad 65:208 [undef][ready]
\_ 6:0:0:21 sdbf 67:144 [undef][ready]
```

If you have ALUA, make the following updates to the *multipath.conf* file:

```
## Use user friendly names, instead of using WWIDs as names. defaults {
user_friendly_names yes
}
devices {
# Device attributed for EMC VNX series ALUA device {
vendor     "DGC"
product    "*"
prio_callout    "/sbin/mpath_prio_alua /dev/%n" path_grouping_policy
group_by_prio
features    "1 queue_if_no_path"
failback    immediate
hardware_handler    "1 alua"
}
```

The sample output from *multipath -ll* is as follows:

```
multipath -ll
mpath3 (3600601600f40190096e0d49c0f27df11) DGC,RAID 5 [size=20G][features=1
queue_if_no_path][hwhandler=1 alua][rw]
```

```
\_ round-robin 0 [prio=100][active]
\_ 5:0:3:0 sdd 8:48 [active][ready]
\_ 6:0:3:0 sdh 8:112 [active][ready]
\_ round-robin 0 [prio=20][enabled]
\_ 5:0:4:0 sde 8:64 [active][ready]
\_ 6:0:4:0 sdi 8:128 [active][ready]
```

Red Hat Enterprise Linux 6.0 (and point releases)

Only Red Hat Enterprise Linux 6.5 and later versions support the active/active ALUA mode. If you have PNR, make the following updates to the *multipath.conf* file:

```
## Use user friendly names, instead of using WWIDs as names. defaults {
user_friendly_names yes
}
blacklist {

devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*" devnode "^hd[a-z]"
devnode "^cciss!c[0-9]d[0-9]*"
}
# Device attributes for EMC VNX series PNR devices {
device {
vendor "DGC"
product ".*" product_blacklist "LUNZ"
path_grouping_policy group_by_prio
getuid_callout "/lib/udev/scsi_id --whitelisted --device=/dev/%n"
path_checker emc_clariion
features "1 queue_if_no_path"
hardware_handler "1 emc" prio emc
failback immediate rr_weight uniform no_path_retry 60
}
```

The sample output from *multipath -ll* is as follows:

```
mpathg (36006016093203700a4994a72d0e5e311) undef DGC,VRAID
size=1.0G features='1 queue_if_no_path' hwhandler='1 emc' wp=undef
|-+- policy='round-robin 0' prio=1 status=undef
| |- 0:0:1:2 sdh 8:112 undef ready running
| `- 1:0:1:2 sdp 8:240 undef ready running
`-+- policy='round-robin 0' prio=0 status=undef
|- 0:0:0:2 sdc 8:32    undef ready running
`- 1:0:0:2 sdl 8:176 undef ready running
```

If you have ALUA, make the following updates to the *multipath.conf* file:

```
## Use user friendly names, instead of using WWIDs as names. defaults {
user_friendly_names yes
}
devices {

device { vendor "DGC" product ".*"
product_blacklist "LUNZ" path_grouping_policy group_by_prio path_checker
emc_clariion
features "1 queue_if_no_path"
hardware_handler "1 alua" prio alua
failback immediate rr_weight uniform no_path_retry 60
}
}
```

Red Hat Enterprise Linux 6.3

For RHEL6.3, make the following updates to the *multipath.conf* file:

```
## Use user friendly names, instead of using WWIDs as names. defaults {
user_friendly_names     yes
}
blacklist {
devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*" devnode "^hd[a-z][[0-9]*]"
devnode "^cciss!c[0-9]d[0-9]*"
}
devices { device { vendor "DGC" product ".*"
product_blacklist "LUNZ" path_grouping_policy group_by_prio
getuid_callout "/lib/udev/scsi_id --whitelisted --device=/dev/%n"
path_checker emc_clariion
features "1 queue_if_no_path"
hardware_handler "1 alua" prio tpg_pref
failback immediate no_path_retry 300
}}
```

The sample output from *multipath -ll* is as follows:

```
# multipath -ll
mpatha (36006016014003500f6b47e53ff61f6ed) dm-2 DGC     ,VRAID
size=2.0G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='round-robin 0' prio=50 status=active
| |- 7:0:3:0    sdf 8:80    active ready running
| `- 10:0:3:0 sdn 8:208 active ready running
`-+- policy='round-robin 0' prio=10 status=enabled
|- 7:0:1:0    sdc 8:32    active ready running
`- 10:0:2:0 sdl 8:176 active ready running
```

Red Hat Enterprise Linux 7.0 (and point releases)

For Red Hat Enterprise Linux 7.0 and later releases, if you have PNR, make the following updates to the *multipath.conf* file:

(i) **Note:** The default configuration parameters for the Unity series and VNX series arrays are included as part of the multipath package. They do not need to be separately defined.

(i) **Note:** See the Red Hat documentation to ensure that the correct syntax is followed for your release. MPIO continues to evolve with each release of Red Hat. The following is only an example:

```
## Use user friendly names, instead of using WWIDs as names. defaults {
user_friendly_names     yes
}

devices { device { vendor "DGC" product ".*"
product_blacklist "LUNZ" path_grouping_policy "group_by_prio" path_checker
"emc_clariion"
features "1 queue_if_no_path"
hardware_handler "1 emc" prio "emc"
failback immediate rr_weight "uniform" no_path_retry 60
}

}
```

The sample output from *multipath -ll* is as follows:

```
mpathb (360060160782918000be69f3182b2da11) DGC,VARID
[size=3G][features=1 queue_if_no_path][hwhandler=1 emc]
\_  round-robin 0 [prio=2][undef]
\_  10:0:3:0 sdaq 66:160 [undef][ready]
\_  11:0:3:0 sdcw 70:64 [undef][ready]
\_  round-robin 0 [prio=0][undef]
\_  11:0:2:0 sdce 69:32 [undef][ready]
\_  10:0:2:0 sdy 65:128 [undef][ready]
```

If you have ALUA, make the following updates to the *multipath.conf* file:

```
## Use user friendly names, instead of using WWIDs as names. defaults {
user_friendly_names    yes
}
devices { device { vendor "DGC" product ".*"
product_blacklist "LUNZ" path_grouping_policy group_by_prio path_checker
emc_clariion
features "1 queue_if_no_path"
hardware_handler "1 alua" prio alua
failback immediate rr_weight uniform no_path_retry 60
}
}
```

The sample output from *multipath -ll* is as follows:

```
mpathbd (36006016029d03d0029bb3057cf338c55) dm-57 DGC    ,VRAID
size=3.0G features='2 queue_if_no_path retain_attached_hw_handler'
hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- 8:0:3:4    sdcu 70:32    active ready running
`-+- policy='service-time 0' prio=10 status=enabled
`- 10:0:2:4    sdcl 69:144    active ready running
```

Red Hat Enterprise Linux 7.2 and later

Since the release of Red Hat Enterprise Linux 7.2, the default stanza parameter
`retain_attached_hw_handler yes` can recognize PNR and ALUA automatically. It does not
need changes to the *multipath.conf* file. However, the manual configuration of PNR and ALUA also
works.

(i) Note: See the Red Hat documentation to ensure that the correct syntax is followed for your
release. MPIO continues to evolve with each release of Red Hat. The following is only an
example:

```
defaults { user_friendly_names    yes
}
devices { device { vendor "DGC" product ".*"
product_blacklist "LUNZ" path_grouping_policy "group_by_prio" path_checker
"emc_clariion"
features "1 queue_if_no_path"
hardware_handler "1 emc" prio "emc"
failback immediate rr_weight "uniform" no_path_retry 60
retain_attached_hw_handler yes detect_prio yes
}
}
```

Oracle Linux and VM Server

All the Oracle Linux versions using the stock Red Hat kernel, Oracle enhanced Red Hat kernel, or UEK kernel use the same configurations as Red Hat Enterprise Linux.

See information about Red Hat Enterprise Linux configuration for Unity storage, VNX Unified Storage, and CLARiiON in the preceding sections in this document.

SuSE Linux Enterprise Server (SLES)

The section discusses methods for configuring the Unity series and VNX (Unified Storage) series storage arrays.

SLES 10 (and point releases)

If you have the PNR, make the following updates to the *multipath.conf* file.

On a SLES 10 system, the default configuration parameters for the Unity series and VNX series arrays are included as part of the multipath package. This is done to eliminate the need to define them separately.

(i) **Note:** See the SuSE documentation to ensure that the correct syntax is followed for your release. MPIO continues to evolve with each release of SLES. The following is only an example:

```
device {
vendor     "DGC"
product    "*" path_grouping_policy group_by_prio
getuid_callout "/sbin/scsi_id -g -u -s/block/%n" prio     emc
hardware_handler "1 emc" features "    1 queue_if_no_path" no_path_retry    60
path_checker    emc_clariion failback    immediate
}
```

The sample output from *multipath -ll* is as follows:

```
mpathb (360060160782918000be69f3182b2da11) DGC,RAID 3
[size=3G][features=1 queue_if_no_path][hwhandler=1 emc]
\_   round-robin 0 [prio=2][undef]
\_   10:0:3:0 sdaq 66:160 [undef][ready]
\_   11:0:3:0 sdcw 70:64 [undef][ready]
\_   round-robin 0 [prio=0][undef]
\_   11:0:2:0 sdce 69:32 [undef][ready]
\_   10:0:2:0 sdy 65:128 [undef][ready]
```

If you have ALUA, make the following updates to the *multipath.conf* file:

- SLES10, SLE10SP1:
  There is no explicit ALUA mode supporting SLES 10 and SLES 10 SP1. Only implicit ALUA mode is supported. See the ALUA supported Linux operating system table for details.

- SLES10SP2 and later:
  Explicit ALUA is only supported with SLES 10 SP2 and greater. You only need to perform the following modifications for explicit ALUA support:

```
defaults { user_friendly_names     yes
}
devices {
# Device attributesd for EMC VNX series ALUA device {
vendor     "DGC"
```

```
product        "*"
path_grouping_policy      group_by_prio
getuid_callout     "/sbin/scsi_id -g -u -s /block/%n" prio      alua
hardware_handler      "1 alua"
features      "1 queue_if_no_path"
no_path_retry      60
path_checker      emc_clariion
failback      immediate
}
}
```

The sample output from *multipath -ll* is as follows:

```
mpathf (3600601601bf024006e16cd89575fde11) dm-1 DGC,RAID 5 [size=5.0G]
[features=1 queue_if_no_path][hwhandler=1 alua]
\_ round-robin 0 [prio=100][enabled]
\_ 1:0:0:0 sde 8:64 [active][ready]
\_ 0:0:0:0 sda 8:0 [active][ready]
\_ round-robin 0 [prio=20][enabled]
\_ 1:0:1:0 sdg 8:96 [active][ready]
\_ 0:0:1:0 sdc 8:32 [active][ready]
```

SLES 11 (and point releases)

If you have the PNR, make the following updates to the *multipath.conf* file:

On a SLES 11 system, the default configuration parameters for the Unity series and VNX series arrays are included as part of the multipath package. This is done to eliminate the need to define them separately.

ⓘ Note: See the SuSE documentation to ensure that the correct syntax is followed for your release. MPIO continues to evolve with each release of SLES. The following is only an example:

```
device {
vendor     "DGC"
product     ".*"
product_blacklist     "LUNZ"
getuid_callout     "/lib/udev/scsi_id --whitelisted --device=/dev/%n"
prio_callout     "/sbin/mpath_prio_emc /dev/%n"
features     "1 queue_if_no_path"
hardware_handler     "1 emc" path_grouping_policy     group_by_prio failback
immediate

rr_weight     uniform
no_path_retry     60
path_checker     emc_clariion
prio     emc
}
```

The sample output from *multipath -ll* is as follows:

```
mpathb (360060160782918000be69f3182b2da11) DGC,RAID 3
[size=3G][features=1 queue_if_no_path][hwhandler=1 emc]
\_ round-robin 0 [prio=2][undef]
\_ 10:0:3:0 sdaq 66:160 [undef][ready]
\_ 11:0:3:0 sdcw 70:64 [undef][ready]
\_ round-robin 0 [prio=0][undef]
\_ 11:0:2:0 sdce 69:32 [undef][ready]
\_ 10:0:2:0 sdy 65:128 [undef][ready]
```

If you have ALUA, make the following updates to the multipath.conf file:

> (i) **Note:** Support is for ALUA from kernel 3.0.101-0.21 and later since it includes some crucial fixes for ALUA support.

```
devices {
# Device attributed for EMC VNX series ALUA device {
vendor    "DGC"
product   "*"
hardware_handler    "1 alua" prio    ""alua""
}
}
```

The sample output from *multipath -ll* is as follows:

```
mpathc (3600601601bf02400e354642b9765de11) dm-1 DGC,RAID 3 [size=3.0G]
[features=1 queue_if_no_path][hwhandler=1 alua][rw]
\_  round-robin 0 [prio=8][active]
\_  1:0:0:1 sdc 8:32 [active][ready]
\_  2:0:1:1 sdr 65:16 [active][ready]
\_  round-robin 0 [prio=2][enabled]
\_  1:0:1:1 sdh 8:112 [active][ready]
\_  2:0:0:1 sdm 8:192 [active][ready]
```

SLES 12 (and point releases)

If you have the PNR, make the following updates to the *multipath.conf* file:

> (i) **Note:** On a SLES 12 system, the default configuration parameters for the Unity series and VNX series arrays are included as part of the multipath package. This is done to eliminate the need to define them separately.

> (i) **Note:** See the SuSE documentation to ensure that the correct syntax is followed for your release. MPIO continues to evolve with each release of SLES. The following is only an example:

```
device {
vendor "DGC" product ".*"
product_blacklist "LUNZ" path_grouping_policy "group_by_prio" path_checker
"emc_clariion"
features "1 queue_if_no_path"
hardware_handler "1 emc" prio "emc"
failback "immediate" rr_weight "uniform" no_path_retry 60
}
```

The sample output from *multipath -ll* is as follows:

```
Mpathd (36006016021003500e04ad9cc785fe411) undef DGC,VRAID size=20G
features='1 queue_if_no_path' hwhandler='1 emc' wp=undef
|-+- policy='service-time 0' prio=4 status=undef
| |- 0:0:3:0 sdr 65:16 undef ready running
| `- 1:0:2:0 sdo 8:224 undef ready running
`-+- policy='service-time 0' prio=1 status=undef
|- 0:0:2:0 sdq 65:0    undef ready running
`- 1:0:3:0 sdp 8:240 undef ready running
```

If you have ALUA, make the following updates to the multipath.conf file:

```
device {
vendor "DGC" product ".*"
product_blacklist "LUNZ" path_grouping_policy "group_by_prio" path_checker
"emc_clariion"
features "1 queue_if_no_path"
hardware_handler "1 alua" prio "alua"
failback "immediate" rr_weight "uniform" no_path_retry 60
}
```

The sample output from *multipath -ll* is as follows:

```
mpathgd    (36006016021003500e04ad9cc785fe411)    dm-4 DGC,VRAID
size=20G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='round-robin 0' prio=50 status=active
| |- 0:0:3:0 sdr 65:16 active ready running
| `- 1:0:2:0 sdo 8:224 active ready running
`-+- policy='round-robin 0' prio=10 status=enabled
|- 0:0:2:0 sdq 65:0    active ready running
`- 1:0:3:0 sdp 8:240 active ready running
```

VPLEX

For detailed information about VPLEX, see VPLEX documentation available at Dell EMC Online Support.

VPLEX documentation

See the following documents for configuration and administration operations:

- *Dell EMC VPLEX with GeoSynchrony 5.0 Product Guide*
- *Dell EMC VPLEX with GeoSynchrony 5.0 CLI Guide*
- *Dell EMC VPLEX with GeoSynchrony 5.0 Configuration Guide*
- *Dell EMC VPLEX Hardware Installation Guide*
- *Dell EMC VPLEX Release Notes*
- *Implementation and Planning Best Practices for Dell EMC VPLEX Technical Notes*
- *VPLEX online help*, available on the Management Console GUI
- *VPLEX Procedure Generator*, available at Dell EMC Online Support
- *Dell EMC Simple Support Matrix, Dell EMC VPLEX, and GeoSynchrony*, available at Dell EMC E-Lab Navigator

For the most up-to-date support information, see Dell EMC Simple Support Matrix.

Prerequisites

Before configuring VPLEX in the Linux environment, complete the following on each host:

- Confirm that all necessary remediation has been completed.
  This ensures that OS-specific patches and software on all hosts in the VPLEX environment are at supported levels according to the Dell EMC Simple Support Matrix.
- Confirm that each host is running VPLEX-supported failover software and has at least one available path to each VPLEX fabric.

ⓘ **Note:** See the Dell EMC Simple Support Matrix for the most up-to-date support information and prerequisites.

- If a host is running PowerPath, confirm that the load-balancing and failover policy is set to *Adaptive*.

ⓘ **Note:** For optimal performance in an application or database environment, ensure alignment of your host's operating system partitions to a 32 KB block boundary.

Veritas DMP settings with VPLEX

If a host attached to VPLEX is running Veritas DMP multipathing, change the following values of the DMP tunable parameters on the host to improve the way DMP handles transient errors at the VPLEX array in certain failure scenarios:

- `dmp_lun_retry_timeout` for the VPLEX array to 60 s using the following command:
  `vxdmpadm setattr enclosure emc-vplex0 dmp_lun_retry_timeout=60`

- `recoveryoption` to throttle and `iotimeout` to 30 using the following command:
  `vxdmpadm setattr enclosure emc-vplex0 recoveryoption=throttle iotimeout=30`

- If the Vertitas DMP version is 6.0.1 or later, and to support VPLEX in clustered environment, update the `VRTSaslapm` package to 6.0.100.100 or later.

Host configuration for Linux: Fibre Channel HBA configuration

This section details Fibre Channel (FC) HBA-related configuration that must be addressed when using FC with VPLEX.

The values that are provided are required and optimal for most cases. However, sometimes, the values need to be tuned if the performance of the VPLEX shows high front-end latency in the absence of high back-end latency and this has visible impact on host applications. This can be caused by too many outstanding IOs at a specified time per port.

ⓘ **Note:** For further information about how to monitor VPLEX performance, see the Performance and Monitoring section of the *VPLEX Administration Guide*. If host applications show a performance issue with the required settings, contact Dell EMC Support for further recommendations.

Setting queue depth for QLogic

About this task

ⓘ **Note:** Changing the HBA queue depth is designed for advanced users. Increasing the queue depth might cause hosts to overstress the arrays that are connected to the Linux host, resulting in performance degradation while communicating with them.

- Execution throttle setting - This setting controls the number of outstanding I/O requests per HBA port. Set the HBA exe- cution throttle to the QLogic 65535 default value. You can do this on the HBA firmware level by using the HBA BIOS or QConverageConsole CLI utility that is provided by the HBA vendor.

- Queue depth - This setting controls the amount of outstanding I/O requests per a single path. For Linux, the HBA queue depth can be adjusted using host CLI and text editor.

ⓘ **Note:** When the execution throttle in the HBA level is set to a value lower than the queue depth, it can limit the queue depth to a lower value than set.

Use the following procedures to adjust the queue depth setting for QLogic HBAs:

- Set the QLogic HBA adapter queue depth in Linux to 32 decimal.

- Set the QLogic HBA adapter Execution Throttle to its default value of 65535.

Follow the appropriate procedure according to the HBA type. For any additional information, see the HBA vendor documentation.

To set the queue depth on the Qlogic FC HBA:

Procedure

1. On the Linux host, verify the existing queue depth. Type the following on the host prompt:

   `# cat /sys/module/qla2xxx/parameters/ql2xmaxqdepth 32`
   ⓘ Note: The queue depth value is a decimal.

2. Select one of the following options according to the version and set the queue depth value to 20 hexadecimal (32 decimal):

   - For SuSE and Red Hat version 5.x
     Edit the `/etc/modprobe.conf` file and add or modify the following parameter:

     `qla2xxx ql2xmaxqdepth=<new_queue_depth>`

   - For later SuSE and Red Hat releases:
     Create or edit the `/etc/modprobe.d/qla2xxx.conf` file and add or modify the following parameter:

     `qla2xxx ql2xmaxqdepth=<new_queue_depth>`

   ⓘ Note: The queue depth value means 32 outstanding IOs per ITL. If a host has 4 paths, then there are 32 outstanding IOs per path, resulting in a total 128.

3. Save the file.

4. Type the following command to rebuild the RAMdisk:

   ```
   # mkinitrd -v --with=qla2xxx -f /boot/initramfs-w-qla2xxx-'uname -
   r'.img 'uname -r'
   I: Executing /sbin/dracut -v -f --add-drivers qla2xxx

   /boot/initramfs-w-qla2xxx-2.6.32-504.el6.x86_64.img
   2.6.32-504.el6.x86_64
   ```

5. Type the following command to find the name of the new RAMdisk:

   ```
   # ls /boot/ | grep qla initramfs-w-
   qla2xxx-2.6.32-504.el6.x86_64.img
   ```

6. Type the following command to synchronize data on a disk with memory:

   `#sync`

7. Add an entry to GRUB config file (`/boot/grub/grub.conf`) with the RAMdisk that was rebuilt, as follows:

   ```
   title <OS Name>
   root (hd0,0)
   kernel....
   initrd....
   initrd /<the name of the new RAMdisk>
   ```

8. Type the following command to reboot the host:

   `# /sbin/shutdown -r now`

   Or

 # reboot

9. After the host is booted, type the following command to verify that the
   `<new_queue_depth>` is set.

   `# cat /sys/module/qla2xxx/parameters/ql2xmaxqdepth 20`

   (i) **Note:** The value is in hexadecimal: **20 hexadecimal (32 decimal)**. If the execution throttle
   is less than the queue depth, then the execution throttle overrides the queue depth
   setting.

Setting the execution throttle on the QLogic FC HBA

To set the execution throttle on the Qlogic FC HBA:

Procedure

1. Install QConvergeConsole CLI on the host.

2. Run `qaucli` from the installed directory and select the **Adapter Configuration** from the
   menu:

   `# /opt/QLogic_Corporation/QConvergeConsoleCLI/qaucli`
   Using config file: `/opt/QLogic_Corporation/QConvergeConsoleCLI/qaucli.cfg`

   Installation directory: `/opt/QLogic_Corporation/QConvergeConsoleCLI`

   Working dir: `/opt/QLogic_Corporation/QConvergeConsoleCLI`

   Using config file: `/opt/QLogic_Corporation/QConvergeConsoleCLI/iscli.cfg`

   `Loading iSCSI Data...`

   QConvergeConsole

   CLI - Version 1.1.3 (Build 29) Main Menu

   ```
   1: Adapter Information

   2: Adapter Configuration

   3: Adapter Updates

   4: Adapter Diagnostics

   5: Adapter Statistics

   6: FabricCache CLI

   7: Refresh

   8: Help

   9: Exit
   ```

   Please Enter Selection:

   1. Select Adapter Configuration Please Enter Selection: 2

   QConvergeConsole

   CLI - Version 1.1.3 (Build 29)

   Fibre Channel Adapter Configuration

   ```
   1: Adapter Alias

   2: Adapter Port Alias

   3: HBA Parameters

   4: Persistent Names (udev)

   5: Boot Devices Configuration
   ```

```
6: Virtual Ports (NPIV)
```

```
7: Target Link Speed (iiDMA)
```

```
8: Export (Save) Configuration
```

```
9: Generate Reports
```

1. Select the HBA Parameters

Please Enter Selection: 3

QConvergeConsole

CLI - Version 1.1.3 (Build 29)

Fibre Channel Adapter Configuration

HBA Model QLE2562 SN: LFD1047J02485

```
1: Port 1: WWPN: 21-00-00-24-FF-31-62-16 Online
```

```
2: Port 2: WWPN: 21-00-00-24-FF-31-62-17 Online
```

(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)

Please Enter Selection:

3. Select HBA Port to configure

Please Enter Selection: 1

QConvergeConsole

Version 1.1.3 (Build 29)

HBA Parameters Menu

```
================================================
HBA     : 0 Port: 1
SN      : LFD1047J02485
HBA Model    : QLE2562
HBA Desc.    : QLE2562 PCI Express to 8Gb FC Dual Channel
FW Version    : 4.03.01
WWPN    : 21-00-00-24-FF-31-62-16
WWNN    : 20-00-00-24-FF-31-62-16
Link    : Online
================================================
```

```
1: Display HBA Parameters
```

```
2: Configure HBA Parameters
```

```
3: Restore Defaults
```

(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)

Please Enter Selection:

4. Select Configure HBA Parameters

Please Enter Selection: 2

QConvergeConsole

Version 1.1.3 (Build 29)

Configure Parameters Menu

```
==================================================
HBA    : 0 Port: 1
SN     : LFD1047J02485
HBA Model    : QLE2562
HBA Desc.    : QLE2562 PCI Express to 8Gb FC Dual Channel
FW Version   : 4.03.01
WWPN   : 21-00-00-24-FF-31-62-16
WWNN   : 20-00-00-24-FF-31-62-16
Link   : Online
==================================================
```

```
1: Connection Options

2: Data Rate

3: Frame Size

4: Enable HBA Hard Loop ID

5: Hard Loop ID

6: Loop Reset Delay (seconds)

7: Enable BIOS

8: Enable Fibre Channel Tape Support

9: Operation Mode

10: Interrupt Delay Timer (100 microseconds)

11: Execution Throttle

12: Login Retry Count

13: Port Down Retry Count

14: Enable LIP Full Login

15: Link Down Timeout (seconds)

16: Enable Target Reset

17: LUNs per Target

18: Enable Receive Out Of Order Frame

20: Commit Changes

21: Abort Changes
```

(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)

Please Enter Selection:

5. Select Execution Throttle

Please Enter Selection: 11

Enter Execution Throttle [1-65535] [32]:

6. Set the value to 65535.

Note: The current value is in the second set of square brackets. The first is the allowable range.

Enter Execution Throttle [1-65535] [32]: 65535

QConvergeConsole

Version 1.1.3 (Build 29)

Configure Parameters Menu

```
=======================================================
HBA    : 0 Port: 1
SN     : LFD1047J02485
HBA Model    : QLE2562
HBA Desc.    : QLE2562 PCI Express to 8Gb FC Dual Channel
FW Version   : 4.03.01
WWPN   : 21-00-00-24-FF-31-62-16
WWNN   : 20-00-00-24-FF-31-62-16
Link   : Online
=======================================================
```

1: Connection Options

2: Data Rate

3: Frame Size

4: Enable HBA Hard Loop ID

5: Hard Loop ID

6: Loop Reset Delay (seconds)

7: Enable BIOS

8: Enable Fibre Channel Tape Support

9: Operation Mode

10: Interrupt Delay Timer (100 microseconds)

11: Execution Throttle

12: Login Retry Count

13: Port Down Retry Count

14: Enable LIP Full Login

15: Link Down Timeout (seconds)

16: Enable Target Reset

17: LUNs per Target

18: Enable Receive Out Of Order Frame

19: Enable LR

20: Commit Changes

21: Abort Changes

(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)

Please Enter Selection:

7. Commit changes

(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)

Please Enter Selection: 20

HBA Parameters Update Complete. Changes have been saved to HBA instance 0.

Press <Enter> to continue:

8. Press Any Key to Continue

QConvergeConsole

Version 1.1.3 (Build 29)

HBA Parameters Menu

```
================================================
HBA    : 0 Port: 1
SN     : LFD1047J02485
HBA Model    : QLE2562
HBA Desc.    : QLE2562 PCI Express to 8Gb FC Dual Channel
FW Version   : 4.03.01
WWPN   : 21-00-00-24-FF-31-62-16
WWNN   : 20-00-00-24-FF-31-62-16
Link   : Online
================================================
```

1: Display HBA Parameters

2: Configure HBA Parameters

3: Restore Defaults

(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)

Please Enter Selection:

9. Validate that the Execution Throttle is set to the expected value of 65535

```
-------------------------------------------------------
HBA Instance 0: QLE2562 Port 1 WWPN 21-00-00-24-FF-31-62-16
PortID 74-3B-00
Link: Online
Connection Options    : 2
-------------------------------------------------------------
Loop Preferred, Otherwise Point-to-Point
```

Data Rate	Auto
Frame Size	2048
Hard Loop ID	0
Loop Reset Delay (seconds)	5
Enable Host HBA BIOS	Disabled
Enable Hard Loop ID	Disabled
Enable FC Tape Support	Enabled
Operation Mode	0 - Interrupt for every I/O completion
Interrupt Delay Timer (100us)	0
Execution Throttle	65535
Login Retry Count	8
Port Down Retry Count	45
Enable LIP Full Login	Enabled
Link Down Timeout(seconds)	45
Enable Target Reset	Enabled
LUNs Per Target	256
Out Of Order Frame Assembly	Disabled

Enable LR	Disabled
Enable Fabric AssignedWWN	N/A

10. Press Enter to continue.

QConvergeConsole

Version 1.1.3 (Build 29)

HBA Parameters Menu

```
=====================================================
HBA     : 0 Port: 1
SN      : LFD1047J02485
HBA Model      : QLE2562
HBA Desc.      : QLE2562 PCI Express to 8Gb FC Dual Channel
FW Version     : 4.03.01
WWPN    : 21-00-00-24-FF-31-62-16
WWNN    : 20-00-00-24-FF-31-62-16
Link    : Online
=====================================================
```

1: Display HBA Parameters

2: Configure HBA Parameters

3: Restore Defaults

(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)

Please Enter Selection:

11. Press 0 to go back to previous Menu

Please Enter Selection: 0

QConvergeConsole

CLI - Version 1.1.3 (Build 29)

Fibre Channel Adapter Configuration

HBA Model QLE2562 SN: LFD1047J02485

1: Port 1: WWPN: 21-00-00-24-FF-31-62-16 Online

2: Port 2: WWPN: 21-00-00-24-FF-31-62-17 Online

(p or 0: Previous Menu; m or 98: Main Menu; ex or 99: Quit)

Please Enter Selection:

12. Repeat steps 3-12 for each port on each adaptor connecting to VPLEX.

13. Press 99 to Quit from QConvergeConsoleCLI.

Setting queue depth and queue target for Emulex

About this task

(i) Note: Changing the HBA queue depth is designed for advanced users. Increasing the queue depth may cause hosts to overstress other arrays that are connected to the Linux host, resulting in performance degradation while communicating with them.

- HBA Queue depth - This setting controls the number of outstanding I/O requests per HBA port. On Linux, the Emulex HBA queue depth can be adjusted using Host CLI.

- LUN Queue depth - This setting controls I/O depth limiting on a per target LUN basis. On Linux, the Emulex HBA queue depth can be adjusted by using CLI.

The following procedures detail adjusting the HBA queue depth and LUN queue depth settings for Emulex HBAs as follows:

- Set the Emulex HBA adapter queue depth in Linux to *8192*
- Set the Emulex HBA LUN queue depth in Linux to *32*

(i) Note: This means 32 outstanding IOs per LUN, so if a host has 4 paths then 8 outstanding IOs per path.

Follow the appropriate procedure according to the HBA type. For any additional information, see the HBA vendor documentation.

To set the queue depth on the Emulex FC HBA:

Procedure

1. On Linux host, verify the existing queue depth by typing the following command on the host prompt:

   ```
   # cat /etc/modprobe.d/lpfc.conf
   options lpfc lpfc_hba_queue_depth=32
   options lpfc lpfc_lun_queue_depth=30
   options lpfc lpfc_sg_seg_cnt=256
   ```

2. Select one of the following options according to the version:

   - For SuSE and Red Hat 5.x and earlier:
     Edit the file `/etc/modprobe.conf` and add or modify the following parameter:

     ```
     options lpfc lpfc_lun_queue_depth= <new_queue_depth>
     ```

   - For later SuSE and Red Hat releases:
     Create or edit the `/etc/modprobe.d/lpfc.conf` file and add or modify the following parameter:

     ```
     options lpfc lpfc_lun_queue_depth=<new_queue_depth>
     ```

3. Save the file.

4. Make a backup copy of RAMdisk by typing the following command:

   ```
   # cp /boot/initramfs-2.6.32-358.el6.x86_64.img
   initramfs-2.6.32-358.el6.x86_64.img.bak
   ```

5. Type the following command to rebuild the RAMdisk:

   ```
   # mkinitrd /boot/initramfs-2.6.32-358.el6.x86_64.img
   2.6.32-358.el6.x86_64 --force
   ```

6. Type the following command to synchronize data on disk with memory:

   ```
   #sync
   ```

7. Type the following command to reboot the host:

   ```
   # /sbin/shutdown -r now
   ```

   Or

   ```
   # reboot
   ```

8. Type following command to verify that the new queue depths are set:

   ```
   cat /sys/class/scsi_host/host*/lpfc_hba_queue_depth
   ```

```
cat /sys/class/scsi_host/host*/lpfc_lun_queue_depth
```

Provisioning and exporting storage

This section provides information about the following:

- VPLEX with GeoSynchrony v4.x
- VPLEX with GeoSynchrony v5.x
- VPLEX with GeoSynchrony v6.x

VPLEX with GeoSynchrony v4.x

To begin using VPLEX, you must provision and export storage so that hosts and applications can use the storage. Storage provisioning and exporting refers to the following tasks required to take a storage volume from a storage array and make it visible to a host:

Procedure

1. Discover available storage.
2. Claim and name storage volumes.
3. Create extents from the storage volumes.
4. Create devices from the extents.
5. Create virtual volumes on the devices.
6. Create storage views to allow hosts to view specific virtual volumes.
7. Register initiators with VPLEX.
8. Add initiators (hosts), virtual volumes, and VPLEX ports to the storage view.

   You can provision storage using the GUI or the CLI. See *Dell EMC VPLEX Management Console Help* or the *Dell EMC VPLEX CLI Guide*, at Dell EMC Online Support, for more information.

   The following figure illustrates the provisioning and exporting process.

Figure 22 VPLEX provisioning and exporting storage process

VPLEX with GeoSynchrony v5.x

VPLEX allows easy storage provisioning among heterogeneous storage arrays. After a storage array LUN volume is encapsulated within VPLEX, all its block-level storage is available in a global directory and coherent cache. Any front-end device that is zoned properly can access the storage blocks.

Two methods available for provisioning: EZ provisioning and Advanced provisioning. For more information, see *Dell EMC VPLEX with GeoSynchrony 5.0 Product Guide*, available on Dell EMC Online Support.

VPLEX with GeoSynchrony v6.x

VPLEX provides easy storage provisioning among heterogeneous storage arrays. Use the web-based GUI to simplify everyday provisioning or create complex devices.

Use the following ways to provision storage in VPLEX:

- Integrated storage provisioning (VIAS-VPLEX Integrated Array Services based provisioning)
- EZ provisioning
- Advanced provisioning

All provisioning features are available in the Unisphere for VPLEX GUI.

For more information, see the *Dell EMC VPLEX with GeoSynchrony 6.0 Product Guide*, available on Dell EMC Online Support.

Storage volumes

A storage volume is a LUN exported from an array. When an array is discovered, the storage volumes view shows all exported LUNs on that array. You must claim, and optionally name, these

storage volumes before you can use them in a VPLEX cluster. Once claimed, you can divide a storage volume into multiple extents (up to 128), or you can create a single full-size extent using the entire capacity of the storage volume.

(i) **Note:** To claim storage volumes, the GUI supports only the Claim Storage wizard, which assigns a meaningful name to the storage volume. Meaningful names help you associate a storage volume with a specific storage array and LUN on that array, and help during troubleshooting and performance analysis.

This section provides information about the following:

- Claiming and naming storage volumes
- Extents
- Devices
- Distributed devices
- Rule sets
- Virtual volumes

Claiming and naming storage volumes

You must claim storage volumes before you can use them in the cluster (except the metadata volume, which is created from an unclaimed storage volume). Only after claiming a storage volume you can use it to create extents, devices, and then virtual volumes.

Extents

An extent is a slice (range of blocks) of a storage volume. You can create a full size extent using the entire capacity of the storage volume, or you can carve the storage volume up into several contiguous slices. Extents are used to create devices, and then virtual volumes.

Devices

Devices combine extents or other devices into one large device with specific RAID techniques, such as mirroring or striping. Devices can only be created from extents or other devices. The storage capacity of a device is not available until you create a virtual volume on the device and export that virtual volume to a host.

You can create only one virtual volume per device. There are two types of devices:

- Simple device - A simple device is configured using one component, which is an extent.
- Complex device - A complex device has more than one component, combined using a specific RAID type. The components can be extents or other devices (both simple and complex).

Distributed devices

Distributed devices are configured using storage from both clusters and are used only in multicluster plexes. A distributed device's components must be other devices and those devices must be created from storage in different clusters in the plex.

Rule sets

Rule sets are predefined rules that determine how a cluster behaves when it loses communication with the other cluster, for example, during an inter-cluster link failure or cluster failure. In these situations, until communication is restored, most I/O workloads require specific sets of virtual volumes to resume on one cluster and remain suspended on the other cluster.

VPLEX provides a Management Console on the management server in each cluster. You can create distributed devices using the GUI or CLI on either management server. The default rule set used by

the GUI makes the cluster that is used to create the distributed device detach during communication problems, allowing I/O to resume at the cluster. For more information, on creating and applying rule sets, see the *Dell EMC VPLEX CLI Guide*, available at Dell EMC Online Support.

There are cases in which all I/O must be suspended resulting in a data unavailability. VPLEX with GeoSynchrony 5.0 is introduces the new functionality of VPLEX Witness. When a VPLEX Metro or a VPLEX Geo configuration is augmented by VPLEX Witness, the resulting configuration provides the following features:

- High availability for applications in a VPLEX Metro configuration (no single points of storage failure)

- Fully automatic failure handling in a VPLEX Metro configuration

- Significantly improved failure handling in a VPLEX Geo configuration

- Better resource utilization

For information about VPLEX Witness, see the *Dell EMC VPLEX with GeoSynchrony 5.0 Product Guide*, available at Dell EMC Online Support.

Virtual volumes

Virtual volumes are created on devices or distributed devices and presented to a host by a storage view. Virtual volumes can be created only on top-level devices and always use the full capacity of the device.

System volumes

VPLEX stores configuration and metadata on system volumes that are created from storage devices. There are two types of system volumes. Each is briefly discussed in this section:

- Metadata volumes
- Logging volumes

Metadata volumes

VPLEX maintains its configuration state, referred to as metadata, on storage volumes provided by storage arrays. Each VPLEX cluster maintains its own metadata, which describes the local configuration information for this cluster and any distributed configuration information shared between clusters.

For more information about metadata volumes for VPLEX with GeoSynchrony v4.x, see the *Dell EMC VPLEX CLI Reference Guide* available on Dell EMC Online Support.

For more information about metadata volumes for VPLEX with GeoSynchrony v5.x, see the *Dell EMC VPLEX with GeoSynchrony 5.0 and Point Releases Product Guide* available on Dell EMC Online Support.

For more information about metadata volumes for VPLEX with GeoSynchrony v6.x, see the *Dell EMC VPLEX with GeoSynchrony 6.0 Product Guide* available on Dell EMC Online Support.

Logging volumes

Logging volumes are created during initial system setup and are required in each cluster to keep track of any blocks written during a loss of connectivity between clusters. After an inter-cluster link is restored, the logging volume is used to synchronize distributed devices by sending only changed blocks over the inter-cluster link.

For more information about logging volumes for VPLEX with GeoSynchrony v4.x, see the *Dell EMC VPLEX CLI Guide*, available on Dell EMC Online Support.

For more information about logging volumes for VPLEX with GeoSynchrony v5.x, see the *Dell EMC VPLEX with GeoSynchrony 5.0 and Point Releases Product Guide*, available on Dell EMC Online Support.

For more information about metadata volumes for VPLEX with GeoSynchrony v6.x, see the *Dell EMC VPLEX with GeoSynchrony 6.0 Product Guide*, available on Dell EMC Online Support.

Required storage system setup

See product documentation for details and installation procedures for connecting a Symmetrix, VNX series, VNXe series or XtremIO storage system to a VPLEX instance, available on Dell EMC Online Support..

Required Symmetrix FA bit settings

For Symmetrix to VPLEX connections, configure the Symmetrix Fibre Channel directors (FAs) as shown in the following table:

ⓘ **Note:** It is recommended to download the latest information before installing any server.

Set[a]	Do not set	Optional
SPC-2 Compliance (SPC2) SCSI-3 Compliance (SC3) Enable Point-to-Point (PP) Unique Worldwide Name (UWN) Common Serial Number (C)	Disable Queue Reset on Unit Attention (D) AS/400 Ports Only (AS4) Avoid Reset Broadcast (ARB) Environment Reports to Host (E) Soft Reset (S) Open VMS (OVMS) Return Busy (B) Enable Sunapee (SCL) Sequent Bit (SEQ) Non Participant (N) OS-2007 (OS compliance)	Enable Auto-Negotiation (EAN) Linkspeed VCM/ACLX[b]

    a. For the Symmetrix 8000 series, the flags should be Unique Worldwide Name (UWN), CommonSerial Number, and Enable Point-to-Point (PP).
    b. Must be set if VPLEX is sharing Symmetrix directors with hosts that require conflicting bit settings. For any other configuration, the VCM/ACLX bit can be either set or not set.

ⓘ **Note:** When setting up a VPLEX-attach version 4.x or earlier with a VNX series, VNXe series or Unity series, the initiator type must be set to CLARiiON Open and the Failover Mode set to 1. ALUA is not supported.

When setting up a VPLEX-attach version 5.0 or later with a VNX series, VNXe series or Unity series, the initiator type can be set to CLARiiON Open and the Failover Mode set to 1 or Failover Mode 4 since ALUA is supported.

If you are using the LUN masking, you will need to set the VCM/ACLX flag. If sharing array directors with hosts which require conflicting flag settings, VCM/ACLX must be used.

ⓘ **Note:** The FA bit settings listed in previous table are for connectivity of VPLEX to a Symmetrix array only. For Host to Symmetrix FA bit settings, see the Dell EMC Simple Support Matrix.

Supported storage arrays

The Dell EMC Simple Support Matrix lists the storage arrays that have been qualified for use with VPLEX.

See the *VPLEX Procedure Generator*, available on Dell EMC Online Support, to verify supported storage arrays.

VPLEX automatically discovers storage arrays that are connected to the back-end ports. All arrays connected to each director in the cluster are listed in the storage array view.

Initiator settings on back-end arrays

See the *VPLEX Procedure Generator*, available on Dell EMC Online Support, to verify the initiator settings for storage arrays when configuring the arrays for use with VPLEX.

Host connectivity

For the most up-to-date information about qualified switches, hosts, host bus adapters, and software, see the Dell EMC Simple Support Matrix or contact your Dell EMC Customer Representative.

The latest Dell EMC-approved HBA drivers and software are available for download on the Dell EMC page of the Broadcom, QLogic, and Brocade websites.

The Dell EMC HBA installation and configurations guides are available at the Dell EMC-specific download pages of these websites.

(i) **Note:** Direct connect from a host bus adapter to a VPLEX engine is not supported.

Exporting virtual volumes to hosts

A virtual volume can be added to more than one storage view. All hosts that are included in the storage view can access the virtual volume.

About this task

The virtual volumes that are created on a device or distributed device are not visible to hosts (or initiators) until you add them to a storage view. For failover purposes, two or more front-end VPLEX ports can be grouped to export the same volumes.

A volume is exported to an initiator as a LUN on one or more front-end port WWNs. Typically, initiators are grouped into initiator groups; all initiators in such a group share the same view on the exported storage (they can see the same volumes by the same LUN numbers on the same WWNs).

An initiator must be registered with VPLEX to see any exported storage. The initiator must also be able to communicate with the front-end ports over a Fibre Channel switch fabric. Direct connect is not supported. Registering an initiator attaches a meaningful name to the WWN, typically the server's DNS name. This allows you to audit the storage view settings to determine which virtual volumes a specific server can access.

(i) **Note:** When performing encapsulation with Citrix XenServer, apply Citrix Hot Fix XS62E003 to XenCenter.

Perform the following task to export a virtual volume:

Procedure

1. Create a storage view, as shown in the following figure:

Figure 23 Creating storage view

2. Register the initiators.

Figure 24 Registering initiators

Note: When initiators are registered, you can set their type as indicated in Table 29: Supported hosts and initiator types.

3. Add the ports to the storage view.

Figure 25 Adding ports to storage view

4. Add virtual volumes to the storage view.

Figure 26 Adding virtual volumes to storage view

Front-end paths

This section contains the following information:

- Viewing the World Wide Name for an HBA port

- VPLEX ports

- Initiators

Viewing the World Wide Name for an HBA port

Each HBA port has a World Wide Name (WWN) associated with it. WWNs are unique identifiers that the VPLEX engine uses to identify its ports and host initiators. You can use one of the following ways to view WWNs:

- Switch's name server output

- Dell EMC Ionix Control Center or Solutions Enabler

- `syminq` command (Symmetrix users)

- The Linux sysfs file system; for example

```
# cat /sys/class/fc_host/host1/port_name 0x10000000c93dc171 #
```

VPLEX ports

The virtual volumes that are created on a device are not visible to hosts (initiators) until you export them. Virtual volumes are exported to a host through front-end ports on the VPLEX directors and HBA ports on the host/server. For failover purposes, two or more front-end VPLEX ports can be used to export the same volumes. Typically, to provide maximum redundancy, a storage view will have two VPLEX ports assigned to it, preferably from two different VPLEX directors. When volumes are added to a view, they are exported on all VPLEX ports in the view, using the same LUN numbers.

Initiators

For an initiator to see the virtual volumes in a storage view, it must be registered and included in the storage view's registered initiator list. The initiator must also be able to communicate with the front-end ports over Fibre Channel connections, through a fabric.

A volume is exported to an initiator as a LUN on one or more front-end port WWNs. Typically, initiators are grouped so that all initiators in a group share the same view of the exported storage(they can see the same volumes by the same LUN numbers on the same WWN host types).

Ensure that you specify the correct host type in the **Host Type** column as this attribute cannot be changed in the **Initiator Properties** dialog box once the registration is complete. To change the host type after registration, you must unregister the initiator and then register it again using the correct host type.

VPLEX supports the host types that are listed in following table. When initiators are registered, you can set their type, also indicated in the following table.

Table 30 Supported hosts and initiator types

Host	Initiator type
Windows, MSCS, Linux	default
AIX	Aix
HP-UX	Hp-ux
Solaris, VCS	Sun-vcs

Configuring Linux hosts to recognize VPLEX volumes

The VPLEX presents SCSI devices to the Linux operating system like other Dell EMC storage products. See LUN scanning mechanisms for information about how to make the virtual volumes be recognized by the operating system.

Linux native cluster support

VPLEX in a clustered storage configuration may present devices that are distributed across the two VPLEX sites. Coherency is provided through a WAN link between the two sites such that if one site should go down, the data is available from the other site. Further high availability for the data center is achievable through server clustering. Red Hat's Red Hat Cluster Suite (RHCS) and SUSE High Availability Extension (HAE) are supported in such an application. The following figure shows an example of RCHS in a noncross-cluster connected configuration.

Figure 27 Red Hat RCHS in a non-cross-cluster connected configuration

The following figure shows an example of HAE in a cross-cluster connected configuration.

Figure 28 SUSE HAE in a cross-cluster connected configuration

Supported Red Hat RHCS configurations and best practices

Consider the following support information and best practices.

Supported

- Red Hat Enterprise Linux RHCS 5.5 and later, and Red Hat Enterprise Linux RHCS 6.x
- Minimum GeoSynchrony 5.0.1
- Active/passive or active/active cluster
- Stretched storage cluster without cross-connect
- Stretched storage clusters with cross-cluster connect
  - < 1 ms round-trip time latency for connections between the two sites.
  - PowerPath only
- PowerPath or Linux native DM-MPIO
- Red Hat Enterprise Linux GFS and GFS2 clustered file system
- RHCS quorum disk
- RHCS fence_scsi() and sg3util sg_persist()
- Stretched storage cluster without cross-connect

Best practices

- For Red Hat's stretch cluster support, see the Red Hat Knowledgebase ArticleID: 58412,Support for Red Hat Enterprise Linux High Availability Cluster Stretch and Multi-site Architectures.
- Server cluster heartbeat limitation per Red Hat specification (< 2.0 ms RTT)
  - See the Red Hat Knowledgebase Article ID: 58412,Support for Red Hat Enterprise LinuxHigh Availability Cluster Stretch and Multi-Site Architectures.
- Consistency groups should be set for one site wins for all distributed devices in the VPLEX metro cluster that is attached to the RHCS cluster. The winning site should be aligned to the winning nodes in the RHCS cluster.
- An RHCS quorum disk was qualified on a distributed device. If an RHCS quorum disk is configured under an PowerPath pseudo device, it is recommended that the TOTEM TOKEN should be set at 330,000 milliseconds. This is to protect against a worst case failure scenario regarding paths to the quorum disk. It is further recommended that if the recommended setting is not used, test your desired settings in a nonproduction environment before implementing them. If not, an unstable cluster may result.
Find further guidance about RHCS quorum disks inHow to Optimally Configure a Quorum Disk in Red Hat Enterprise Linux Clustering and High-Availability Environments in the Red Hat Customer Portal.
- fence_scsi() and sg_persist() were qualified on a distributed device under both PowerPath and Linux DM-MPIO.
- Patches for fence_scsi() and fence_scsi_test() to support PowerPath pseudo devices may be required for Red Hat Enterprise Linux 5.x configurations in order for them to detect the PowerPath pseudo device.

```
- Patches to fence_scsi()
# sub get_scsi_devices [This is the original macro] #{
#    my ($in, $out, $err);
#    my $cmd = "lvs --noheadings --separator : -o vg_attr,devices"; #    my
$pid = open3($in, $out, $err, $cmd) or die "$!\n";
```

```
#
#     waitpid($pid, 0); #
#     die "Unable to execute lvs.\n" if ($?>>8); #
#     while (<$out>) #     {
# chomp;
# print "OUT: $_\n" if $opt_v; #
# my ($vg_attrs, $device) = split(/:/, $_); #
# if ($vg_attrs =~ /.*c$/) # {
#     $device =~ s/\(.*\)//; #     push(@volumes, $device); # }
#     }
#
#     close($in); #     close($out); #     close($err); #}
sub get_scsi_devices [This is the change]
{
open(FILE, "/var/run/scsi_reserve") or die "$!\n";

while (<FILE>)
{ chomp; push(@volumes, $_);
}
close FILE;
}LEX
```

Patches to fence_scsi_test()

```
- Patches to fence_scsi_test()

sub get_block_devices
{
my $block_dir = "/sys/block";
opendir(DIR, $block_dir) or die "Error: $! $block_dir\n";
my @block_devices = grep { /^sd*/ | /^emcpower*/ } readdir(DIR);

closedir(DIR);
for $dev (@block_devices)
{
push @devices, "/dev/" . $dev;
}
```

Patches for /usr/sbin/fence_scsi to support pseudo devices may be required for Red Hat Enterprise Linux 6.x configurations in order for it to detect the PowerPath pseudo device.

```
- Patches to /usr/sbin/fence_scsi sub get_devices_scsi ()
{
my $self = (caller(0))[3]; my @devices;
opendir (\*DIR, "/sys/block/") or die "$!\n";
#     @devices = grep { /^sd/ } readdir (DIR);     #[This is the original
macro]
@devices = grep { /^sd/ | /^emcpower/ } readdir (DIR);#[This is the change]
closedir (DIR);

return (@devices);
```

Supported SUSE HAE configurations and best practices

Consider the following support information and best practices.

Supported

- SLES 11 and later, and the SUSE High Availability Extension (HAE)
- Minimum GeoSynchrony 5.0.1
- Active/passive or active/active cluster

- Stretched storage cluster without cross-connect
- Stretched storage clusters with cross-cluster connect
  - < 1ms round-trip time latency for connections between the two sites.
  - PowerPath only
- PowerPath or Linux native DM-MPIO
- OCFS2 clustered file system
- STONITH block device (sbd)

Best practices

- Reference *SUSE Linux Enterprise High Availability Extension* documentation
- It is recommended that SUSE technical support is engaged to design and tune your cluster configuration files.
- Stretched storage cluster without cross-connect
- Server cluster heartbeat limitation < 2.0ms RTT
- Consistency groups should be set for one site wins for all distributed devices in the VPLEX metro cluster attached to the HAE cluster. The winning site should be aligned to the winning nodes in the HAE cluster.
- The following changes from the default CRM settings were made to the CRM Config policy engine under advisement of SuSE engineering for the configuration tested:
  - Default Action Timeout = 100 s
  - Stonith Timeout = 300 seconds(This was made based on PowerPath characteristics. For Linux native DM-MPIO, follow the SuSE HAE documentation for recommended settings.)
- If a STONITH block device (sbd) is used (http://www.linux-ha.org/wiki/SBD_Fencing)
  - Recommended settings for a multipathed device (PowerPath or Linux native DM-MPIO):
    - Timeout (watchdog) : 90
    - Timeout (allocate) : 2
    - Timeout (loop) : 1
    - Timeout (msgwait) : 180
  - Create the sbd on a distributed data volume from the VPLEX and place it in a consistency group with a *site 1 wins policy*.
- Identically configured servers at both sites per SuSE recommendations.
- All LUNs used for the OCFS2 file system should be under CLVM control, use distributed data volumes from the VPLEX, and be placed in a consistency group with a *site 1 wins* policy.

MPIO configuration for Dell EMC VPLEX virtualized storage

This section discusses methods for configuring VPLEX virtualized storage on the following platforms:

Red Hat Enterprise Linux (RHEL)

This topic provides a brief overview if you are planning to modify the default behavior of the Linux native MPIO.

The Linux native MPIO since RHEL 5.5 already contains default configuration parameters for VPLEX virtualized storage to provide optimal performance in most environments. There is no need to create a device stanza for these arrays unless you want to modify the default behavior.

(i) **Note:**

The following code syntax is only an example. Consult Red Hat documentation to ensure the correct syntax is followed for your release. MPIO continues to evolve with each release of RHEL.

```
## Use user friendly names, instead of using WWIDs as names. defaults {
user_friendly_names yes
}
devices {
# Device attributes for EMC VPLEX device {
vendor"EMC"
product"Invista"
path_checkertur
no_path_retry5
product_blacklist"LUNZ"
}
}
```

This is a sample output from `multipath -ll`:

```
# multipath -v2 -d
create: mpath15 (36006048000019010196553303042374 4) EMC,INVISTA
[size=8.4G][features=0][hwhandler=0]
\_ round-robin 0 [prio=2][undef]
\_ 2:0:0:49 sdp 8:240 [undef][ready]
\_ 3:0:0:49 sds 65:32 [undef][ready]
```

Oracle Linux and VM Server

All the Oracle Linux versions using the stock Red Hat kernel, Oracle enhanced Red Hat kernel or the Unbreakable Enterprise Kernel (UEK) use the same configurations as the Red Hat Enterprise Linux.

For more information, see

SuSE Linux Enterprise Server (SLES)

This topic provides a brief overview if you are planning to modify the default behavior of the SLES Linux native MPIO.

Since SLES10 SP3, the Linux native MPIO already contains default configuration parameters for VPLEX virtualized storage to provide optimal performance in most environments. There is no need to create a device stanza for these arrays unless you want to modify the default behavior.

(i) **Note:** The following code syntax is only an example. Consult SUSE documentation to ensure that the correct syntax is followed for your release. MPIO continues to evolve with each release of SLES.

```
##
defaults { user_friendly_namesyes
```

```
}
devices {
##Device attributes for EMC SYMMETRIX device {
vendor"EMC"
product"INVISTA"
path_grouping_policymultibus
path_checkertur
no_path_retry5
product_blacklist"LUNZ"
}
}
```

This is a sample output from `multipath.conf`:

```
mpath27 (36006048000019010050153303135381) EMC,INVISTA
[size=468M][features=0][hwhandler=0]
\_ round-robin 0 [prio=4][undef]
\_ 11:0:0:39 sdbr 68:80 [undef][ready]
\_ 11:0:1:39 sdcc 69:0[undef][ready]
\_ 10:0:0:39 sdl8:176[undef][ready]
\_ 10:0:1:39 sdw65:96[undef][ready]
```

Optimal Path Management

Optimal Path Management (OPM) was introduced with VPLEX 5.5 to improve VPLEX performance. OPM uses the ALUA mechanism to spread the I/O load across VPLEX directors while gaining cache locality.

VPLEX OPM feature overview

The OPM feature in VPLEX FE uses the SCSI Asymmetric Logical Unit Access (ALUA) mechanism to indicate to a host which path is optimal and which is non-optimal. This increases cache locality by directing LUN I/O only through the local director, and improves overall performance.

The following figures show the cache coherency overhead reductions on the local site, first without OPM, and then with OPM.

Figure 29 Cache coherency overhead reduction without OPM multiple volumes

Figure 30   Cache coherency overhead reduction with OPM multiple volumes

OPM is enabled on VPLEX per initiator base. A host should have all its initiators enabled or disabled for OPM. Mixed mode is not supported.

After OPM is enabled for a particular volume that is provisioned to a host, you can check out its optimal director using VPLEX CLI. The following figure shows an example of a volume in a metro cluster.

Figure 31   Volume in a metro cluster example

```
VPlexcli:/> opm virtual-volume show-state DD_Xtremio_LUN771_Symm0052_427C-HP_vol
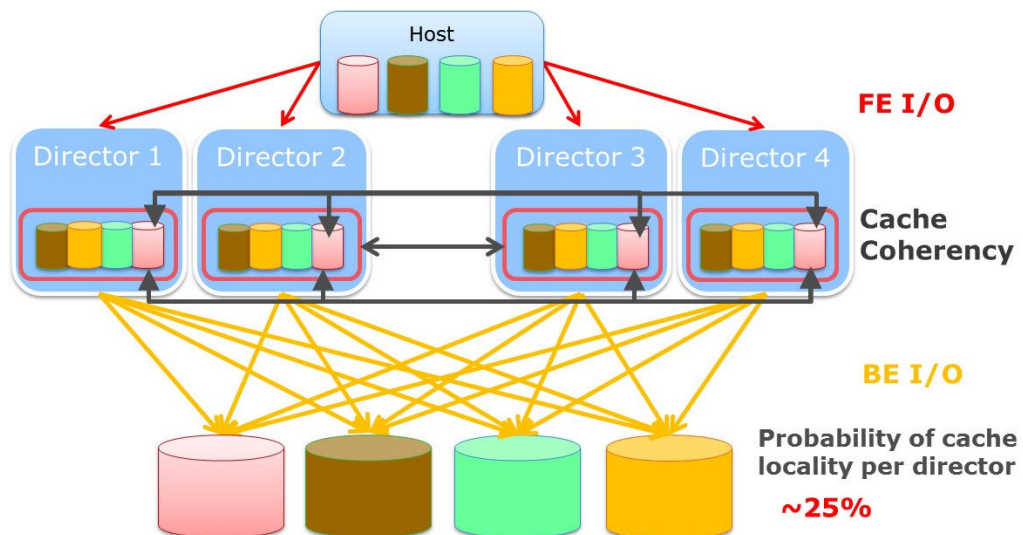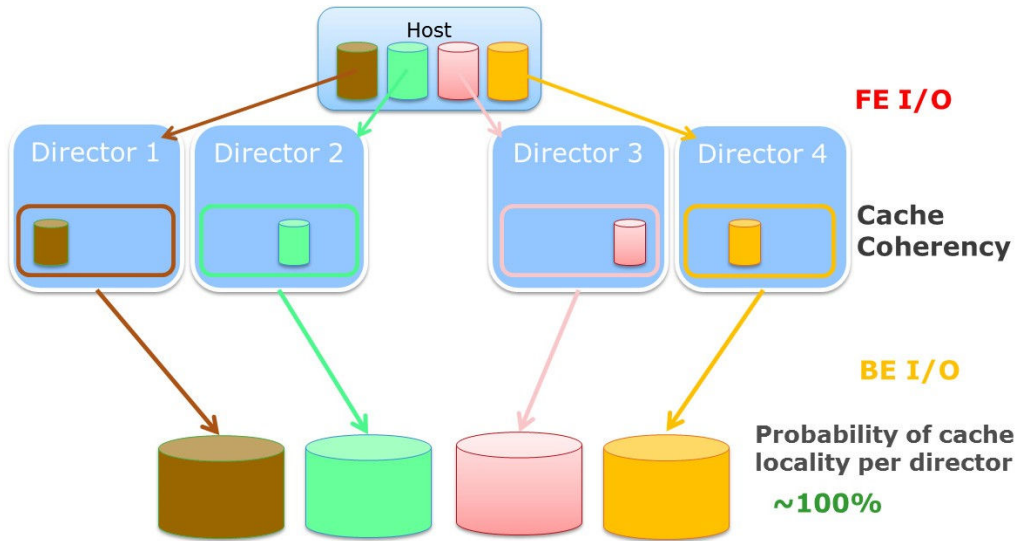
cluster-1
Virtual Volumes                         Optimal Directors   Override Directors
------------------------------------    -----------------   ------------------
DD_Xtremio_LUN771_Symm0052_427C-HP_vol  [director-1-1-A]    []

cluster-2
Virtual Volumes                         Optimal Directors   Override Directors
------------------------------------    -----------------   ------------------
DD_Xtremio_LUN771_Symm0052_427C-HP_vol  [director-2-1-A]    []
```

Host connectivity best practices while using OPM

These are the best practices while using OPM:

- A Storage View should include target ports from all directors if possible, to get the most benefit from OPM.

- OPM is not supported in a cross-connected configuration.

- If a host is cross connected to two VPLEX clusters and OPM is enabled, the RTPG responses may be inconsistent and the behavior is undefined.

The following figure shows the best practice host connectivity for dual VPLEX engines:

Figure 32 Host connectivity across dual VPLEX engines

The following figure shows the best practice host connectivity quad engines in a VPLEX cluster.

Figure 33 Host connectivity across quad engines in a VPLEX cluster

Host multipathing software configuration while using OPM

Host multipathing software must be ALUA-aware to detect the optimal path for each LUN, and send I/O down only the optimal path. Multipathing software, such as PowerPath, can automatically detect the OPM ALUA mode with a host reboot, but for some host native multipathing software, additional configuration may be required.

(i) **Note:** Veritas DMP does not support OPM ALUA.

PowerPath

PowerPath versions that support VPLEX OPM are summarized in the following table.

Table 31 PowerPath versions that support VPLEX OPM

	PowerPath 6.0 and later	PowerPath 5.9.x	PowerPath 5.8.x	PowerPath 5.7.x	PowerPath 5.5.x
Linux	Yes	N/A	N/A	N/A	N/A

When first enabling OPM on the VPLEX side, the host must be rebooted for PowerPath to detect the change and handle the LUNs in the OPM ALUA mode. After it is in ALUA mode, you can check it by typing the `powermt display alua device=[device name | all]` command, as shown in the following figure:

Figure 34 Output from the powermt command

```
Pseudo name=emcpower6a
VPLEX ID=PMTELABA000001
Logical device ID=6000144000000010F00333BB862AC0D8 [DD_XtremIO_LUN502_CX4809_LUN473-Solaris_vol]
state=alive; policy=ADaptive; queued-IOs=0
==============================================================================
---- Host ----   - Stor -  ------------ I/O Path ------------    -- Stats ---
### I/O Paths   Interf.   ALUA State            Mode    State   Errors
==============================================================================
3074 c6t500014426004D501d0s0 CL2-09   Active/non-optimized   active  alive    1
3072 c5t500014426004D603d0s0 CL2-03   Active/optimized        active  alive    1
3072 c5t500014427004D503d0s0 CL2-0F   Active/non-optimized   active  alive    1
3074 c6t500014427004D601d0s0 CL2-05   Active/non-optimized   active  alive    1
```

LUNs from the VPLEX array are displayed, and optimal/non optimal paths are displayed in the command output. I/O runs only on the optimized path when compared to the previous result of all four paths.

You can view the active path match with the optimal director path for that LUN shown in the VPLEX CLI. Verify that the host PowerPath device ALUA paths match the OPM configuration for a given LUN

In the host, run the `powermt display alua dev=[device name | all]` command, as shown in the following figure:

Figure 35 Output from the powermt command

```
Administrator@A50T3234 ~
$ powermt display alua dev=all
Pseudo name=harddisk1
VPLEX ID=FNM00141000019
Logical device ID=6000144000000000107043F71D5C566D82
state=alive; policy=ADaptive; queued-IOs=0
==============================================================================
---- Host ----   - Stor -  ------------ I/O Path ------------    -- Stats ---
### I/O Paths   Interf.   ALUA State            Mode    State   Errors
==============================================================================
    3 c3t3d0    CL1-08   Active/non-optimized   active  alive    0
    3 c3t2d0    CL1-06   Active/non-optimized   active  alive    0
    3 c3t1d0    CL1-05   Active/non-optimized   active  alive    0
    3 c3t0d0    CL1-03   Active/optimized        active  alive    0
    2 c2t3d0    CL1-04   Active/non-optimized   active  alive    0
    2 c2t2d0    CL1-02   Active/non-optimized   active  alive    0
    2 c2t1d0    CL1-01   Active/non-optimized   active  alive    0
    2 c2t0d0    CL1-00   Active/optimized        active  alive    0
```

Note the *Storage Interface* values for each of the paths of the device.

In the VPLEX CLI, run the export storage-view `show-powerpath-interfaces` command, as shown in the following figure:

Figure 36 Output from the export storage-view command

```
VPlexcli:/> export storage-view show-powerpath-interfaces -c cluster-1/
PowerPath Interface    Target Port                    Director Name
-------------------    -----------------------        -------------
CL1-0E                 P00000000476043F7-A0-FC03.0    director-1-2-A
CL1-0F                 P000000000477043F7-B0-FC03.0   director-1-2-B
CL1-0C                 P0000000000477043F7-B0-FC02.0  director-1-2-B
CL1-0D                 P000000000047704419-B0-FC03.0  director-1-1-B
CL1-0A                 P00000000476043F7-A0-FC02.0    director-1-2-A
CL1-0B                 P0000000000047604419-A0-FC03.0 director-1-1-A
CL1-05                 P000000000047704419-B0-FC01.0  director-1-1-B
CL1-06                 P0000000000476043F7-A0-FC01.0  director-1-2-A
CL1-03                 P0000000000047604419-A0-FC01.0 director-1-1-A
CL1-04                 P0000000000477043F7-B0-FC00.0  director-1-2-B
CL1-09                 P000000000047704419-B0-FC02.0  director-1-1-B
CL1-07                 P0000000000047604419-A0-FC02.0 director-1-1-A
CL1-08                 P0000000000477043F7-B0-FC01.0  director-1-2-B
CL1-01                 P000000000047704419-B0-FC00.0  director-1-1-B
CL1-02                 P0000000000476043F7-A0-FC00.0  director-1-2-A
CL1-00                 P0000000000047604419-A0-FC00.0 director-1-1-A
```

Map each `Storage Interface` value to its corresponding director name, as shown in the following table. The optimal paths detected in the host lead to director-1-1-A.

Table 32 Mapping Storage Interface values to corresponding director name

Interface	Director	Path ALUA state
CL1-08	Director-1-1-B	Non-optimized
CL1-06	Director-1-2-A	Non-optimized
CL1-05	Director-1-1-B	Non-optimized
CL1-03	Director-1-1-A	Optimized
CL1-04	Director-1-2-B	Non-optimized
CL1-02	Director-1-2-A	Non-optimized
CL1-01	Director-1-1-B	Non-optimized
CL1-00	Director-1-1-A	Optimized

To map the *Logical Device ID* from the host output, go to the virtual-volumes context in the VPlexCli, and run the `ll` command. Look for the Logical Device ID (which is prefixed with VPD83T3) and then match it to its corresponding volume name in VPLEX. In the example shown in the following figure, the ID matches the volume name *rc1_lr0_0000_vol*.

Figure 37 Output from the ll command

```
VPlexcli:/clusters/cluster-1/virtual-volumes> ll
Name              Operational  Health  Service    Block    Block  Capacity  Locality  Supporting    Cache Mode   Expandable  Expandable  Consisten
cy  VPD ID                                                                                                                                
                  Status       State   Status     Count    Size   --------  --------  Device        ----------   ----------  Capacity    Group
--------------    -----------  ------  ---------  -------  -----  --------  --------  ----------    -----------  ----------  ----------  ---------
--
rc1_lr0_0000_vol  ok           ok      running    4194304  4K     16G       local     rc1_lr0_0000  synchronous  true        0B          -
     VPD83T3:6000144000000010107043f71d5c566d82
rc1_lr0_0001_vol  ok           ok      unexported 4194304  4K     16G       local     rc1_lr0_0001  synchronous  true        0B          -
     VPD83T3:6000144000000010107043f71d5c566d83
```

In the VPLEX CLI, run the `opm virtual-volume show-state [volume name]` command, as shown in the example in the following figure:

Figure 38 Output from the opm command

As can be seen, the optimal director configured for the volume is director-1-1-A, so the OPM configuration is verified to match the host detected ALUA paths.

Verify that I/O is running to the optimal paths in PowerPath:

- In the host, configure PowerPath to collect perf tracing by running the `powermt set perfmon=on interval=60` command.

- Start running I/O from the host on the selected target devices, and wait for 60 s.

- In the host, run the `powermt display perf dev=[ dev name | all]` command. For example, as shown in the example in the following figure:

Figure 39 Output from the powermt command

The I/O paths that show I/O being formed and match them to the original output of `powermt display alua dev=[dev name | all]` to verify that the optimal ALUA paths are the only paths with I/O

Native MPIO

The following examples provide some general guidance for configuring native MPIO with the ALUA mode. Whether the particular operating system native multipathing software is ALUA-aware and what the parameters are to configure can vary from operating system to operating system, even version to version.

(i) **Note:** See the operating system distributor's documentation for complete details.

Linux DM-MPIO
For Linux using native DM-MPIO, with a default setting, all VPLEX LUN paths will have same `prio` value and all active for I/O, even if OPM is already enabled from array side for that host initiators.

```
#multipath -ll
36000144000000010f00333bb8640ace9 dm-5 EMC,Invista
size=5.0G features='1 queue_if_no_path' hwhandler='0' wp=rw
`-+- policy='round-robin 0' prio=1 status=active
  |- 3:0:0:1 sdb    8:16    active ready running
  |- 3:0:1:1 sdf    8:80    active ready running
  |- 4:0:1:1 sdo    8:224 active ready running
  `- 4:0:2:1 sds    65:32 active ready running
```

For a host to detect OPM, modify the `/etc/multipath.conf` file with the following ALUA stanza, and then restart the multipath service for the change to take effect. Verify the syntax with your Red Hat release.

RHEL 6, 7 & 8

```
device {
  vendor "EMC"
  product "Invista"
  product_blacklist "LUNZ"
  path_grouping_policy "group_by_prio"
  path_checker "tur"
  features "0"
  hardware_handler "1 alua"
  prio "alua"
  rr_weight "uniform"
  no_path_retry 5
  failback immediate
}
```

SLES 11 & 12

```
device {
  vendor "EMC"
  product "Invista"
  product_blacklist "LUNZ"
  path_grouping_policy "group_by_prio"
  path_checker "tur"
  features "0"

  hardware_handler "1 alua"
  prio "alua"
```

```
  rr_weight "uniform"
  no_path_retry 5
  path_selector "round-robin 0"
  failback immediate
}
```

Oracle Linux 6.x, 7.x, and equivalent VM server

```
device {
vendor "EMC"
product "Invista"
product_blacklist "LUNZ"
path_grouping_policy group_by_prio getuid_callout "/lib/udev/scsi_id --
whitelisted
--device=/dev/%n" (OVM/UEK6 only)

path_selector "round-robin 0" features " 1 queue_if_no_path" hardware_handler
" 1 alua" prio tpg_pref
rr_weight uniform no_path_retry 5
rr_min_io_rq 1 failback immediate
}
```

```
[root@lxh9ser5 ~]# multipath -ll 36000144000000010f00333bb862ac384 dm-1
EMC,Invista
size=100G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='round-robin 0' prio=130 status=active
| `- 4:0:1:1 sdj 8:144 active ready running
`-+- policy='round-robin 0' prio=10 status=enabled
|- 4:0:0:1 sdd 8:48 active ready running
|- 3:0:0:1 sdq 65:0 active ready running
`- 3:0:1:1 sdw 65:96 active ready running
```

XenServer 6.5 or 7.x

```
device {
vendor     "EMC"
product    "Invista"
detect_prio    yes

retain_attached_hw_handler yes path_grouping_policy group_by_prio failback
immediate
}
36000144000000010f00333bb862ac38c dm-4 EMC,Invista
size=200G features='2 queue_if_no_path retain_attached_hw_handler'
hwhandler='1 alua' wp=rw
|-+- policy='round-robin 0' prio=50 status=active
| `- 1:0:1:1 sdt 65:48 active ready running
`-+- policy='round-robin 0' prio=10 status=enabled
|- 1:0:0:1 sde 8:64    active ready running
|- 8:0:1:1 sdj 8:144 active ready running
`- 8:0:0:1 sdo 8:224 active ready running
```

(i) | **Note:**

> For XenServer, starting with version 6.5, the previous setting will automatically detect a Dell EMC array in an ALUA mode, and no additional ALUA parameter is needed in the `/etc/multipath.conf` file.

After the `multipath.conf` file is correctly configured, and the multipath service is restarted, verify that you can see that the handler is changed to `1 alua`, and paths are grouped into two.

One group should contain one or more optimal paths while the other contains the rest of the non-optimal paths. The optimal group will have a higher priority listed than the non-optimal group.

The `multipath -ll` output after ALUA takes effect is shown in the following example:

```
mpathb (36000144000000010f00333bb8640ac48) dm-2 EMC    ,Invista
size=10G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- 0:0:4:1 sdb    8:16     active ready running
`-+- policy='service-time 0' prio=10 status=enabled
|- 0:0:0:1 sdx    65:112 active ready running
|- 5:0:3:1 sdm    8:192    active ready running
`- 5:0:4:1 sdq    65:0     active ready running
```

You can confirm this output by observing that I/O activity. It will be only happening on the one optimal path:

```
[root@lxh9ser5 ~]# iostat -t 2 | grep sd[dwjq]
sdd            0.00            0.00            0.00            0
0
sdj            89.00           0.00            548.00          0             1096
sdq            0.00            0.00            0.00            0
0
sdw      0.50        4.00            0.00            8             0
```

This path will match the optimal director path of the LUN shown in the VPLEX CLI.

Confirming optimal paths with Linux DM-MPIO
To confirm if the optimal path showing from DM-MPIO matches with the optimal path set at the VPLEX side, you can track from the multipath subpath ID to its connected VPLEX director port WWN as shown in the following example from Red Hat Enterprise Linux 7.

```
mpatha (36000144000000010f00333bb862ac37d) dm-2 EMC    ,Invista
size=20G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='service-time 0' prio=50 status=active
| `- 5:0:5:0 sdn    8:208    active ready running
`-+- policy='service-time 0' prio=10 status=enabled
   |- 0:0:2:0 sda    8:0      active ready running
   |- 0:0:8:0 sde    8:64     active ready running
   `- 5:0:6:0 sdr    65:16    active ready running

[root@lxh5ser3 ~]# cat /sys/class/fc_transport/target5\:0\:5/port_name
0x500014427004d600
}
```

With the above VPLEX director port WWN, you can compare it with the optimal director set at the VPLEX side, as shown in the following figure:

Figure 40 Storage view in VPLEX GUI manager


```
VPlexcli:/>opmvirtual-volumeshow-state
DD-Symm3050_0587 Symm0292_1714-Linux_vol cluster-2
Virtual Volumes
Optimal            Directors               Override Directors
DD-Symm3050_0587 Symm0292_1714-Linux_vol   [director-2-1-B]     []
```

```
VPlexcli:/>ll/engines/engine-2-1/directors/director-2-1-B/hardware/ports/
/engines/engine-2-1/directors/director-2-1-B/hardware/ports:
Name            Address                    Role          Port Status
-------         ------------------         ---------     -----------
B0-FC00         0x500014427004d600         front-end     up
B0-FC01         0x500014427004d601         front-end     up
B0-FC02         0x500014427004d602         front-end     up
B0-FC03         0x500014427004d603         front-end     up
```

Veritas DMP

All current GA versions of Veritas DMP do not support VPLEX OPM ALUA now. Dell EMC and Veritas are collaborating on a solution.

After OPM is enabled on the VPLEX for a host initiator and the host is configured as described, OPM provides the benefit of cache locality, which will yield an improvement in your performance. The following restrictions apply to all supported host environments:

- OPM is not supported in cross-connected configurations.

- Mixed mode configurations are not allowed, where one or more host initiators are in OPM mode and other initiators of the same host are in legacy (active/active) mode. The results of such a configuration are indeterminate and should never be attempted.

Consult the latest Dell EMC VPLEX Simple Support Matrix, available on Dell EMC E-Lab Navigator, for the current list of supported configurations.

XtremIO

This section includes the following topics:

- Recommended Configuration Values Summary

- Fibre Channel HBA Configuration

- I/O Path Behavior

- Multipathing Software Configuration

- Post-Configuration Steps - Using the XtremIO Cluster

(i) Note: In hosts running a hypervisor, such as VMwareESX, Microsoft Hyper-V, or any clustering software, it is important to ensure that the logical unit numbers (LUN) of XtremIO volumes are consistent across all hosts in the hypervisor cluster. Inconsistent LUNs may affect operations such as VM online migration or VM power-up.

Recommended Configuration Values Summary

The following table summarizes all used and recommended variables and their values when configuring hosts for Linux.

Validation	Impact	Severity
Timeout: Minimum I/O of 30 s is set on HBAs of all hosts.	Stability	Mandatory
LUN Queue Depth recommended values: • QLogic: 32 • Emulex: 30	Performance	Recommendation
HBA Queue depth recommended values:	Performance	Warning

Validation	Impact	Severity
• QLogic: 65535 • Emulex: 8192 (lpfc_hba_queue_depth)		
I/O Elevators: Should be configured to *noop* or *deadline*	Performance	Warning
Native Multipathing: To configure the XtremIO disk Device, modify `/etc/multipath.conf` file: • **vendor** - XtremIO • **product** - XtremApp • **path_selector** - "queue-length 0" (FOR RHEL>=6) • **path_selector** - "round-robin 0" (FOR RHEL<6) • **rr_min_io** - 1 (for kernels older than 2.6.31) • **rr_min_io_rq** - 1 (for kernels 2.6.31 and above) • **path_grouping_policy** - multibus	Performance	Recommendation
Native Multipathing: To configure the XtremIO disk Device, modify `/etc/multipath.conf`: • **vendor** - XtremIO • **path_checker** - tur • **failback** - immediate • **fast_io_fail_tmo** - 15	Performance	Mandatory
iSCSI Configuration: To configure the XtremIO disk Device, modify `/etc/iscsi/iscsid.conf`: • node.session.timeo.replacement_timeout = 5 • Set the FirstBurstLength and MaxBurstLength parameters to 512 KB With Red Hat Enterprise Linux 4 series, to configure the XtremIO disk device, modify the `/etc/iscsi/iscsi.conf` file instead of the `/etc/iscsi/iscsid.conf` file. For Veritas Storage Foundation, keep the `replacement_timeout` parameter default value of `120`. For Red Hat Enterprise Linux 4 series, instead of `node.session.timeo.replacem`	Performance	Recommendation

Validation	Impact	Severity
`ent_timeout`, set the value of `ConnFailTimeout` to 5.		
Disk logical block size: XtremIO setting is LB=512 for volumes presented to ESX.	Storage efficiency and Performance	Mandatory
Validation: With XtremIO version 4.0 (or higher), specify Linux as the operating system for each defined initiator.	Serviceability	Mandatory
Alignment: Starting sector is 2048 (default value) or a multiple of 16.	Storage efficiency and Performance	Warning
LVM stripe size: Use a value greater than the application block size. Recommended value is 1 MB and above.	Storage efficiency	Mandatory
File system configuration: use EXT4.	Storage efficiency and Performance	Recommendation

iSCSI Configuration

This section describes the issues that should be addressed when using iSCSI with XtremIO, for optimal performance (with or without an iSCSI HBA).

(i) **Note:** This section applies only for iSCSI. If you are using only Fibre Channel with Linux and XtremIO, go to Fibre Channel HBA configuration.

(i) **Note:** Before you proceed, make sure to review *iSCSI SAN Guidelines section* of *Dell EMC XtremIO Host Configuration Guide* available on Dell EMC Online Support.

Prerequisites

Follow the recommended installation and setup of the appropriate NIC/iSCSI HBA for your system. It is recommended to install the latest driver version (patch), as described in the operating system vendor's site for each specific NIC/iSCSI HBA.

See the Dell EMC Simple Support Matrix, available on Dell EMC E-Lab Navigator for supported NIC/iSCSI HBA models and drivers.

Configuring the XtremIO Disk Device with iSCSI

The configuration file for the multipath daemon is `multipath.conf`. It is used to overwrite the integrated configuration table of the multipath daemon. When iSCSI is used with XtremIO, the `iscsi.conf` file is used to overwrite iSCSI specific multipathing related settings.

To configure the XtremIO disk device, modify the `/etc/iscsi/iscsid.conf` file, using the following parameters:

Parameter	Value	Description
`node.session.timeo.replacement_timeout`	5	Specifies the number of seconds the iSCSI layer waits for a timed-out path/session to re-establish before failing any commands on that path/session. The default value is 120.

Parameter	Value	Description
`FirstBurstLength/Ma xBurstLength`	Adjust these parameters to 512 `KB`	*FirstBurstLength* - Specifies the maximum amount of unsolicited data (in bytes) which an iSCSI initiator can send to the target during the execution of a single SCSI command. *MaxBurstLength* - Specifies the maximum SCSI data payload per a single SCSI command. Adjust these parameters to the same value when the used block size is larger than the default setting for this parameter (256 KB).

ⓘ **Note:** With Red Hat Enterprise Linux4 series, to configure the XtremIO disk device, modify the `/etc/iscsi/iscsi.conf` file instead of the `/etc/iscsi/iscsid.conf` file.

ⓘ **Note:** With Veritas Storage Foundation, keep the `replacement_timeout` parameter default value of `120`.

ⓘ **Note:** With Red Hat Enterprise Linux 4 series, instead of `node.session.timeo.replacement_timeout`, set the value of `ConnFailTimeout` to `5`.

ⓘ **Note:** With Citrix XenServer 6 series, for existing Storage Repositories (SR), after editing the `iscsid.conf` file per the above table, detach and then reattach SRs from XenCenter or by using XenServer xe CLI. This will reflect the new iSCSI timeout for the existing SRs.

ⓘ **Note:** With Red Hat Enterprise Virtualization Hypervisor 6.2 and later, after editing the `iscsid.conf` file per the previous table, execute the following command to ensure that your `iscsid.conf` file is persistent across reboots:
`persist /etc/iscsi/iscsid.conf`

Using these settings prevents commands from being split by the iSCSI initiator and enables instantaneous mapping from the host to the volume.

To apply the adjusted `iscsid.conf` settings, run the following command on the Linux host:

```
service iscsi restart
```

ⓘ **Note:** If a previous iSCSI target is discovered on the Linux host, delete the iSCSI DB and rerun the iSCSI target discovery procedure with the `iscsid.conf` setting described above.

Fibre Channel HBA Configuration

This section describes the FC Host Bus Adapters (HBA) issues that must be addressed when using Fibre Channel with XtremIO for optimal performance.

Prerequisites

To install one or more Dell EMC-approved HBAs into a Linux host, see the procedure in the following documents according to the FC HBA type:

- For QLogic HBAs - *Dell EMC Host Connectivity with Qlogic FIbre Channel iSCSI* and *HBAs and Converged Network Adapters (CNAs) for the Linux Environment*

- For Emulex HBAs - *Dell EMCHost Connectivity with Emulex Fibre Channel and iSCSI* and *HBAs and Converged Network Adapters (CNAs) for the Linux Environment*

These documents provide the procedure to configure the host for connection to the Dell EMC Storage Arrays over a Fibre Channel including any needed HBA BIOS settings.

These documents are available in the Dell EMC OEM section of the QLogic site at http://www.qlogic.com, or the Emulex site at http://www.broadcom.com. They can also be found at http://support.EMC.com.

See the Dell EMC Simple Support Matrix, available on Dell EMC E-Lab Navigator for supported FC HBA models and drivers.

Queue depth and I/O elevators

This section describes queue depth and I/O elevators, and provides the recommended settings.

(i) Note: The FC HBA recommendations in this section are applicable to the following FC HBAs:

- QLogic - adapters with names that start with *ql*

- Emulex - adapters with names that start with *lpfc*

See the Dell EMC Simple Support Matrix, available on Dell EMC E-Lab Navigator for all supported FC HBA models and drivers.

Queue depth

(i) Note: Changing the queue depth settings is designed for advanced users. Increasing the queue depth may cause the host to overstress other arrays that are connected to the Linux host, resulting in performance degradation while communicating with them. Therefore, especially in mixed environments with multiple array types that are connected to the Linux host, compare the XtremIOrecommendations for queue depth with those of other platforms before applying them.

Queue depth is the number of SCSI commands (including I/O requests) that can be handled by a storage device at a given time. A queue depth can be set on either of the following:

- Initiator level - HBA queue depth
- LUN level - LUN queue depth

The HBA queue depth (also known as execution throttle) setting controls the number of outstanding I/O requests per HBA port. The HBA queue depth should be set to the maximum value. This can be done on the HBA firmware level, using the HBA BIOS or CLI utility, provided by the HBA vendor:

- QLogic - HBA queue depth - Change the default value (32) to 65535.

- Emulex - lpfc_hba_queue_depth - No need to change the default (and maximum) value (8192).

HBA queue depth (execution throttle) does not apply to QLE2600 and QLE8300 Series QLogic adapters, and is read-only for 10 GbE adapters. For further information, see http://qlogic.force.com/SupportCenter/articles/FAQ/What-are-the-HBA-Execution-Throttle-settings-best-practices?retURL=%2FSupportCenter%2Fapex%2FCustomer_Support_main&popup=true.

The LUN queue depth setting controls the amount of outstanding I/O requests per a single path. These settings are controlled in the driver module for the card at the operating system level.

With XtremIO and Linux, LUN queue depth setting should retain its default values.

If required, LUN queue depth setting for QLogic and Emulex HBAs can be adjusted, using Linux udev rules (see Setting the LUN queue depth using Linux Udev).

Alternatively, it is possible to adjust the LUN queue depth setting for QLogic and Emulex HBAs, using the CLI. See Setting the LUN queue depth with QLogic HBA using the CLI and Setting the LUN queue depth with Emulex HBA using the CLI, respectively. Follow the instructions in either of these sections according to the HBA type.

(i) **Note:** If the HBA queue depth is set to a value lower than the LUN queue depth, it may limit the LUN queue depth to a lower value than set.

The following table summarizes the default and recommended queue depth settings for Linux:

	LUN Queue Depth	HBA Queue Depth
Default Value	QLogic: 32 Emulex: 30	QLogic: 32 Emulex: 8192
Recommended Value	QLogic: Keep the default value. Emulex: Keep the default value.	QLogic: 65535 (maximum) Emulex: 8192 (maximum)

I/O Elevators

I/O scheduling controls how I/O operations are submitted to storage. Linux offers various I/O algorithms (also known as "Elevators") to accommodate for different workloads. Elevators mainly help in reducing seek operations.

Linux sets a default I/O scheduler at boot-time that can be changed on the fly per block device.

When connecting a Linux host to XtremIO storage, set the I/O elevator to either *noop* or *deadline*. It is not recommended to use the *cfq* I/O elevator setting, as it is less optimal for XtremIO storage.

For optimal operation with XtremIO, it is recommended to set the I/O elevators, using Linux udev rules (see Setting the I/O Elevators using Linux Udev.

Alternatively, it is possible to set the I/O elevators using CLI (see Setting the I/O Elevators using the CLI.

Setting the LUN queue depth and I/O elevators, using Linux Udev

Setting the LUN queue depth using Linux Udev

About this task

(i) **Note:** The procedures for setting the LUN queue depth with QLogic HBA and with Emulex HBA are identical, except for the set queue depth value.

To set the LUN queue depth, using Linux udev:

Procedure

1. Create or edit the following file:

   ```
   /etc/udev/rules.d/99-XtremIO.rules
   ```

2. Append the following rule to the file:

   ```
   #increase queue depth on the volume
   ACTION=="add|change", SUBSYSTEM=="scsi",
   ATTR{vendor}=="XtremIO", ATTR{model}=="XtremApp    ",
   ATTR{queue_depth}="<selected LUN queue depth value>"
   ```

   (i) **Note:** There should be eight (8) spaces between 'XtremApp' and the closing quotation mark.

   (i) **Note:** Be sure to use the appropriate queue depth value according to the HBA vendor (QLogic or Emulex). To determine the LUN queue depth value, see the corresponding guidance (see Queue depth).

3. Save the changes.

4. Run the following command to apply the changes:

   `udevadm trigger`

   ⓘ **Note:** Older kernels may use the `udevtrigger` command instead.

Setting the I/O elevators using Linux Udev

There are two options for disabling any I/O elevator:

About this task

For Linux running with any multipath program other than Linux native multipath:

To disable any I/O elevator, follow one of these procedures according to the used GRUB version:

- GRUB (versions prior to GRUB2):

  - Modify the `/boot/grub/menu.lst` file and append `elevator=noop` as an option at the end of the kernel line.

- GRUB2:

  1. Modify the `/etc/default/grub` file and append *elevator=noop* to the list of *GRUB_CMDLINE_LINUX_DEFAULT* values.

  2. Rebuild GRUB2 configuration files, using the following shell command:
     `sudo update-grub`

For Linux running with native multipath:

Procedure

1. Create or edit the following file:

   `/etc/udev/rules.d/99-XtremIO.rules`

2. Copy and paste the following rules to the file:

   `# Use noop scheduler`

   ```
   ACTION=="add|change", SUBSYSTEM=="block",
   ENV{ID_VENDOR}=="XtremIO", ENV{ID_MODEL}=="XtremApp",
   ATTR{queue/scheduler}="noop"

   ACTION=="add|change", SUBSYSTEM=="block", KERNEL=="dm*",
   ENV{DM_NAME}=="??14f0c5*", ATTR{queue/Scheduler}="noop"
   ```

   ⓘ **Note:** These rules are general recommendations. Some Linux operating systems may benefit from using deadline elevator configuration.

3. Save the changes that are made to the file.

4. Run the following command to apply the changes:

   `udevadm trigger`

   ⓘ **Note:** Older kernels may use the *udevtrigger* command instead.

   ⓘ **Note:** Following this procedure, if new devices are configured, udev discovers them and applies the rules automatically.

Setting the LUN queue depth and I/O elevators, using the CLI

Setting the LUN queue depth with QLogic HBA, using the CLI

About this task

ⓘ **Note:** This procedure applies to other Linux environments, which are supported by XtremIO, as well. For details, see the XtremIO Simple Support Matrix, available on Dell EMC E-Lab Navigator.

For QLogic cards controlled by the qla2xxx module, if required, set the `ql2xmaxqdepth` parameter (default value 32) to the appropriate queue depth value.

ⓘ **Note:** Be sure to use the appropriate queue depth value according to the HBA vendor (QLogic). To determine the LUN queue depth value, see the corresponding guidance (see Queue depth).

To check the QLogic LUN queue depth:

Run the following command to display the current QLogic LUN queue depth:

`cat /sys/module/qla2xxx/parameters/ql2xmaxqdepth`

Example:

`# cat /sys/module/qla2xxx/parameters/ql2xmaxqdepth 32`
If the value displayed is per the required queue depth, no further modifications are required for the QLogic LUN queue depth.

To modify the LUN queue depth:

Procedure

1. Select one of the following options according to the version:

    - For SuSE and Red Hat version 5.x:
      Edit the file `/etc/modprobe.conf` file, and add or modify the following parameter:

      `options qla2xxx ql2xmaxqdepth=<new_queue_depth>`

    - For later SuSE and Red Hat releases:
      Create or edit the`/etc/modprobe.d/qla2xxx.conf` file, and add or modify the following parameter:

      `options qla2xxx ql2xmaxqdepth=<new_queue_depth>`

2. Run the following command to rebuild the RAMdisk:

    `mkinitrd -v --with=qla2xxx -f /boot/initramfs-w-qla2xxx-'uname -r'.img 'uname -r'`

    ⓘ **Note:** After copying and pasting this command, ensure that the back quote sign appears as `. Otherwise, edit it manually before running this command.

3. Add a new entry to the GRUB file with the RAMdisk that was rebuilt in step 2.

    ⓘ **Note:** To find the name of the new RAMdisk, run the following command: `ls /boot/ | grep qla`

    ⓘ **Note:** For details on the procedure for editing the GRUB file, see Linux documentation.

4. After the parameter is set, reboot the host.

5. Run the following command to verify that the parameter is set:

```
cat /sys/module/qla2xxx/parameters/ql2xmaxqdepth
```

Setting the LUN queue depth with Emulex HBA, using the CLI

About this task

(i) **Note:** This procedure applies to the other Linux environments, which are supported by XtremIO, as well. For details, see the XtremIO Simple Support Matrix available on Dell EMC E-Lab Navigator.

For Emulex cards there are two parameters:

- `lpfc_lun_queue_depth` (default value 30)
- `lpfc_hba_queue_depth` (default value 8192)

(i) **Note:** Be sure to use the appropriate queue depth value according to the HBA vendor (Emulex). To determine the LUN queue depth value, see the corresponding guidance (see Queue depth).

To check the Emulex LUN queue depth:

`cat /sys/class/scsi_host/host*/lpfc_lun_queue_depth`

Example:

```
# cat /sys/class/scsi_host/host*/lpfc_lun_queue_depth 30
```

If the value displayed is per the required queue depth, no further modifications are required for the Emulex LUN queue depth.

To modify the LUN queue depth value:

Procedure

1. Select one of the following options according to the version:

   - For SuSE and Red Hat 5.x and earlier:
     Edit the `/etc/modprobe.conf` file, and add or modify the following parameter:

     ```
     options lpfc lpfc_lun_queue_depth=<new_queue_depth>
     ```

   - For later SuSE and Red Hat releases:
     Create or edit the `/etc/modprobe.d/lpfc.conf` file, and add or modify the following parameter:

     ```
     options lpfc lpfc_lun_queue_depth=<new_queue_depth>
     ```

2. Run the following command to rebuild the RAMdisk:

   ```
   mkinitrd -v --with=lpfc -f /boot/initramfs-w-lpfc-'uname -r'.img
   'uname -r'
   ```

   (i) **Note:** After copying and pasting this command, ensure that the back quote sign appears as `. Otherwise, edit it manually before running this command.

3. Add a new entry to the GRUB file with the RAMdisk that was rebuilt in step 2.

   (i) **Note:** To find the name of the new RAMdisk, run the following command: `ls /boot/ | grep lpfc`.

ⓘ **Note:** For details on the procedure for editing the GRUB file, see the Linux documentation.

4. After the parameter is set, reboot the host.

5. Run the following command to verify that the parameter is set:

```
cat /sys/class/scsi_host/host*/lpfc_lun_queue_depth
```

Setting the I/O Elevators using the CLI

Follow one of the following procedures according to the GRUB version and Red Hat Enterprise Linux version:

Setting the I/O Elevators using the CLI (for versions prior to GRUB2)

Procedure

1. Check if I/O elevator is already disabled by searching for the `elevator=noop` or `elevator=deadline` string in the /etc/grub.conf file.

   If it is disabled, skip the rest of this procedure.

2. Modify the /boot/grub/menu.lst file, and append "elevator=noop" or "elevator=deadline" as an option to the end of the kernel line.

```
default=0
timeout=5
title CentOS (2.6.32-358.el6.x86_64)
root (hd0,0)
kernel /vmlinuz-2.6.32-358.el6.x86_64 ro ...
elevator=noop
initrd /initramfs-2.6.32-358.el6.x86_64.img
```

Setting the I/O Elevators using the CLI (for GRUB2 - RHEL versions prior to 7)

Procedure

1. Check if I/O elevator is already disabled by searching for the `elevator=noop` or `elevator=deadline` string in the /etc/default/grub file.

   If it is disabled, skip the rest of this procedure.

2. Modify the /etc/default/grub file, and append/write "elevator=noop" or "elevator=deadline" to the list of GRUB_CMDLINE_LINUX_DEFAULT values.

```
GRUB_CMDLINE_LINUX_DEFAULT=" ... elevator=noop"
```

3. Run the following shell command to rebuild the GRUB2 configuration files:

```
sudo update-grub
```

Setting the I/O Elevators using the CLI (for GRUB2 - RHEL versions 7 and later)

Procedure

1. Check if I/O elevator is already disabled by searching for the `elevator=noop' or `elevator=deadline` string in the /etc/default/grub file.

   If it is disabled, skip the rest of this procedure.

2. Modify the `/etc/default/grub` file, and append/write "`elevator=noop`" or "`elevator=deadline`" to the list of `GRUB_CMDLINE_LINUX_DEFAULT` values.

```
GRUB_CMDLINE_LINUX_DEFAULT="elevator=noop"
```

3. Run the following shell command to rebuild the GRUB2 configuration files:

```
grub2-mkconfig -o /boot/grub2/grub.cfg
```

4. Run the following command to verify that the grub file is rebuilt:

```
grep elevator /etc/grub2.cfg
```
The output should include lines containing "`elevator=noop`" or "`elevator=deadline`".

```
#grep elevator /etc/grub2.cfg
linux16 /vmlinuz-3.10.0-229.e17.x86_64
root=/dev/mapper/centos-root ro
rd.lvm.lv=centos/swap crashkernel=auto
ipv6.disable=1 rd.lvm.lv=centos/root
net.ifnames=0 rhgb quiet elevator=noop
linux16 /vmlinuz-0-rescue-f6c987899a1d456ea1fa9b97fe3
5bcf7 root=/dev/mapper/centos-root ro
rd.lvm.lv=centos/swap crashkernel=auto
ipv6.disable=1 rd.lvm.lv=centos/root
net.ifnames=0 rhgb quiet elevator=noop
```

I/O Path Behavior

The XtremIO presents active, accessible LUNs on all paths that are configured to see the LUN. Multipath handles this by using the policy multibus, which is essentially a round-robin policy that distributes the I/O operations over the available channels.

Multipathing software configuration

Configuring Linux native multipathing

Even when the default multipath configuration appears to be functioning and provides path failover, key values must be set to enable the cluster to survive controller failover. It is recommended to configure a volume as a multipath device whenever possible. Even in situations in which a single path is initially used, configuring the volume as a multipath device is advantageous.

The Linux native multipathing requires installing the device-mapper-multipath version 0.4.9-72 (or later) RPM.

To configure the XtremIO disk device, modify the `/etc/multipath.conf` file with the following parameters:

```
devices {
      device {
              vendor                  XtremIO
              product                 XtremApp
              path_selector           "queue-length 0" (FOR RHEL>=6)
              path_selector           "round-robin 0" (FOR RHEL<6)
              rr_min_io               1 (for kernels older than 2.6.31)
```

```
                    rr_min_io_rq           1 (for kernels 2.6.31 and above)
                    path_grouping_policy   multibus
                    path_checker           tur
                    failback               immediate
                    fast_io_fail_tmo       15
            }
    }
```

Parameter	Value	Description
path_selector	**round-robin 0**	Loops through every path in the path group, sending the same amount of I/O to each.
path_selector	**queue-length 0**	Sends the next bunch of I/O down the path with the least number of outstanding I/O requests.
rr_min_io	**1**	Specifies the number of I/O requests to route to a path before switching to the next path in the current path group (for kernels older than 2.6.31).
rr_min_io_rq	**1**	Specifies the number of I/O requests to route to a path before switching the next path in the current path group (for kernels 2.6.31 and above).
path_grouping_policy	**multibus**	All paths are in a single group (all paths have the same priority).
path_checker	**tur**	Specifies TEST UNIT READY as the default method used to determine the state of the paths.
failback	**immediate**	Manages the path group failback. immediate refers to immediate failback to the highest priority path group that contains active paths.
fast_io_fail_tmo	**15**	Specifies the number of seconds between detection of a problem on an FC remote port and failing I/O to devices on the remote port by the SCSI layer.

PowerPath multipathing with XtremIO

XtremIO supports multipathing using Dell EMC PowerPath on Linux. PowerPath versions 5.7 SP3 and above provide Loadable Array Module (LAM) for XtremIO Array devices. With this support, XtremIO devices running versions 2.2 SP1 and above are managed under the XtremIO class.

PowerPath provides enhanced path management capabilities for up to 32 paths per logical device, as well as intelligent dynamic I/O load-balancing functions. Having multiple paths enables the host to access a storage device even if a specific path is unavailable. Multiplepaths share the I/O traffic to a storage device, using intelligent load-balancing policies that enhance I/O performance and increase application availability. Dell EMC PowerPath is the recommended multipathing choice.

(i) Note: For the most updated information about PowerPath support with XtremIO storage, see the XtremIO Simple Support Matrix available on Dell EMC E-Lab Navigator.

PowerPath features include:

- Multiple paths - enables higher availability and I/O performance.
- Path management insight capabilities - PowerPath characterizes I/O patterns and aides in diagnosing I/O problems due to flaky paths or unexpected latency values. Metrics are provided on:
  - Read and write - in MB/seconds per LUN
  - Latency distribution - the high and low watermarks per path
  - Retries - the number of failed I/Os on a specific path
- Autostandby - automatically detects intermittent I/O failures and places paths into autostandby (also known as flaky paths).
- PowerPath Migration Enabler - a host-based migration tool that allows migrating data between storage systems. PowerPath Migration Enabler works along with the host operating system (also referred to as Host Copy) and other underlying technologies, such as Open Replicator (OR). Starting with PowerPath for Linux version 6.1, PowerPath Migration Enabler is supported from VMAX2 to XtremIO storage.
- Remote monitoring:
  - SNMP management daemon
  - PowerPath Management Appliance 2.2 (PPMA 2.2)
- Automated kernel upgrade support in Red Hat Enterprise Linux and Oracle Linux kernels (available in PowerPath 6.0 and later).

Further PowerPath related information:

- For details on the PowerPath releases supported for your Linux host, see the XtremIO *Simple Support Matrix* on Dell EMC E-Lab Navigator.
- For details on class support with XtremIO for your host, see the *Dell EMC PowerPath release notes* document for the PowerPath version you are installing.
- For details on installing and configuring PowerPath with XtremIO class support on your host, see the *Dell EMC PowerPath on Linux Installation and Administration Guide* for the PowerPath version you are installing. This guide provides the required information for placing XtremIO volumes under PowerPath control.

(i) **Note:** The PowerPath with XtremIO class support installation procedure is fully storage-aware. All required PowerPath settings with XtremIO storage are automatically done when PowerPath is installed on your host. This includes settings such as the PowerPath multipathing policy that does not require manual setting

Post-Configuration Steps - Using the XtremIO Cluster

After the host configuration is completed, you can access the XtremIO cluster from the host. For details on creating, presenting, and managing volumes that are accessed from the host using either the GUI or the CLI, see the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

Disk Formatting

Generally, a 512 B logical block size should be used for a new XtremIO volume. The following figure demonstrates formatting an XtremIO Volume using the WebUI.

Figure 41 Formatting an XtremIO Volume using the WebUI

ⓘ Note: For details on formatting a newly created Volume (using either the WebUI or the GUI interfaces), see the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

Presenting XtremIO Volumes to the Linux Host

ⓘ Note: It is recommended to avoid using host-level software for deduplication, compression, and encryption of data stored in XtremIO. These operations are performed natively on the XtremIO cluster, thus releasing host computing resources. Furthermore, using these operations may severely impact data efficiency on the XtremIO cluster as they can lead to a higher physical or logical capacity utilization.

ⓘ Note: The following information in this section applies only to XtremIO version 4.0 and later.

When adding Initiator Groups and Initiators to allow Linux hosts to access XtremIO volumes, specify `Linux` as the operating system for the newly created Initiators.

The following figure demonstrates setting the Operating System field for an Initiator using the WebUI.

Figure 42 Setting the operating system field

> (i) **Note:** Setting the Initiator's Operating System is required for optimal interoperability and stability of the host with XtremIO storage. You can adjust the setting while the host is online and connected to the XtremIO cluster with no I/O impact.

See the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster for details.

Following a cluster upgrade from XtremIO version 3.0.x to version 4.0 (or later), ensure to modify the operating system for each initiator that is connected to a Linux host.

Partition alignment in Linux

To align partitions on XtremIO volumes that are presented to Linux hosts, use the default value (2048), but create a partition using the `fdisk` command to ensure that the file system is aligned.

When performing partition alignment, the logical device (`/dev/mapper/<naa-name>`) should be used rather than the physical device (`/dev/<device>`). When multipathing is not used (for example in a virtual machine), the physical device should be used.

The following example demonstrates using the `fdisk` command to create an aligned partition on an XtremIO volume.

```
[root@lg114 ~]# fdisk -cu /dev/mapper/3514f0c5b12a00004
Device contains neither a valid DOS partition table, nor Sun, SGI or OSF
disklabel
Building a new DOS disklabel with disk identifier 0x12d4e90c
Changes will remain in memory only, until you decide to write them.
After that, of cource, the previous content won't be recoverable.

Warning: invalid flag 0x0000 of partition table 4 will be corrected by w(rite)

Command (m for help): p

Disk /dev/mapper/3514f0c5b12a00004: 1649.3 GB, 1649267441664 bytes
255 heads, 63 sectors/track, 200512 cylinders, total 3221225472 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 16384 bytes / 65536 bytes
Disk identifier: 0x12d4e90c
            Device Boot Start End Blocks Id System
```

In this mode, rather than using *cylinders* for creating partitions, the `fdisk` command uses sectors, which are a direct mapping to the array's LBA space. Thus, verifying that the partition is aligned is simply a case of verifying that the starting sector number is a multiple of 16 (16sectors, at 512 bytes each, is 8 KB). The `fdisk` command will default to a starting sector for the first partition of 2048, which is divisible by 16, and thus is correctly aligned.

Logical volume manager

> (i) **Note:** Logical Volume Manager (LVM) configuration and management are out of the scope of this document .

When configuring LVM with a Linux host that is connected to XtremIO storage, follow these recommendations:

- LVM stripe size - The LVM stripe size should be set to a value greater than that of the application block size.

- XtremIO storage supports a physical extent size of 8 KB and above. the recommended physical extent size is 1 MB or above.

Creating a file system

It is recommended to create the file system using its default block size (using a non-default block size may lead to unexpected behavior). See the operating system and file system documentation.

For a more efficient data utilization and better performance, use Ext4 file system with XtremIO storage instead of Ext3. For details about converting to Ext4 file system (from either Ext3 or Ext2), see Upgrade to Ext4.

XtremIO Optimal Transfer Length

XtremIO Storage Array is reporting an optimal transfer length of 64 KB to the Linux host, while maximum transfer size reported is 4 MB in all XtremIO releases (for XtremIO version 4 or later).

Linux kernels running versions earlier than 4.3 (and sometimes earlier than 3.1) do not check the optimal transfer length and use a default value of 512 KB.

With Linux kernel version 4.3 or later, and sometimes with version 3.1 or later, the kernel is using the array reported optimal sizes. This means that for XtremIO array, any block size larger than the optimal size (64 KB) would be split into 64 KB chunks, and the maximum size of an I/O will be 4 MB. This change should not cause performance issues in most environments.

However, if there is a need to change the settings to match the previous kernel settings, use `udev` to override the maximum block size setting.

To do this, **create a text file named** `etc/udev/rules.d/99-xtremio.rules` **with the** following content:

```
ACTION=="add|change", SUBSYSTEM=="block",
ATTRS{vendor}=="XtremIO ", ATTR{queue/max_sectors_kb}="512",
ATTR{queue/scheduler}="<scheduler>"
```

Confirm the following:

- There should be a space after the string 'XtremIO' and before the quotation mark.

- All entries are == except for the last two which are =.

- The scheduler option should be either *deadline* or *noop*.

To apply the `{{99-xtremio.rules}}` configuration file, run the command `udevadm` trigger or reboot the host.

To confirm that the setting was applied, run the following command for one of the XtremIO Volumes presented to the Linux host:

```
$ cat /sys/block/sdX/queue/max_sectors_kb 512
```

(i) **Note:** In the above command, change `sdX` to match the selected XtremIO Volume.

Using LUN 0 with XtremIO storage

This section details considerations and steps that should be performed when using LUN0 with Linux.

Notes on the use of LUN numbering:

- In XtremIO version 4.0.0 (or later), volumes are numbered by default starting from LUN ID 1 (and not 0 as was the case in previous XtremIO versions).

- Although possible, it is not recommended to manually adjust the LUN ID to 0, as it may lead to issues with some operating systems.
  (i) **Note:** In case LUN 0 is required for an XtremIO volume (for example boot from SAN, PowerPath), restart the Linux host if a rescan fails to locate this volume.

- When a cluster is updated from XtremIO version 3.0.x to 4.0.x, an XtremIO volume with a LUN ID 0 remains accessible following the upgrade.

- With XtremIO version 4.0.0 (or above), no further action is required if volumes are numbered starting from LUN ID 1.

The following output to the `lsscsi` command indicates that the device at 1:0:0:0 is the XtremIO cluster controller:

```
[1:0:0:0] storage XtremIO XtremApp 3000 –
```

In that case, an XtremIO volume with LUN 0 will be inaccessible to the host.

To access a volume with a LUN 0 on a Linux host:

1. Perform one of the following commands to remove the controller device:

   - ```
     # /usr/bin/rescan-scsi-bus.sh -r
     ```
     (i) **Note:** -r enables the device removal.

   - ```
     # echo 1 > /sys/class/scsi_device/1:0:0:0/device/delete
     ```

2. Run the following command:
   ```
   # /usr/bin/rescan-scsi-bus.sh
   ```

   (i) **Note:** In older Linux versions a host reboot is required instead of a rescan.

Space reclamation

This section provides a comprehensive list of capacity management steps for achieving optimal capacity utilization on the XtremIO array, when connected to a Linux host.

Data space reclamation helps to achieve optimal XtremIO capacity utilization. Space reclamation is a Linux operating system function, enabling to reclaim used space by sending zeros to a specific address of the volume after being notified by the file system that the address space was deleted. While some file system scan performs this action automatically, others require a user-initiated operation.

The following sections present steps for performing space reclamation with:

- EXT3 file system
- EXT4 file system
- XFS file system

A suggested script for initiating space reclamation is also provided.

EXT3 file system

Manual Space Reclamation - EXT3 space reclamation can be performed, using the `fstrim` utility. Using `fstrim` to reclaim unused space does not create a balloon file. Therefore, there is no need to review the free space.

The following example shows running `fstrim` on `/mnt/ext3` mount point.

```
# fstrim /mnt/ext3
```

Sometimes, `fstrim` may fail due to T10 commands blocking (VPLEX/VMDK). In such scenarios, it is required to apply a manual copy of zeroes to the relevant free space.

The following example shows running a manual script on /mnt/ext3 mount point (see Linux space reclamation script).

```
# ./reclaim_space.sh /mnt/ext3
```

EXT4 file system

- Automatic space reclamation - EXT4 supports automatic space reclamation, using the discard option. This option should be enabled when mounting the file system (either manually or using `fstab`).
  The following example shows `discard` usage when mounting a file system.

  ```
  /dev/vg_ext4_test/lv_ext4_test /mnt/ext4 ext4 defaults,discard 0 1
  ```

  If automatic space reclamation is not enabled immediately after file system creation, it is recommended to run a manual initial space reclamation prior to automatic mounting.

- Manual space reclamation - EXT4 space reclamation can be performed, using the `fstrim` utility.
  Using the `fstrim` utility does not create a balloon file. Therefore, it is not necessary to review the free space.

  The following example shows running `fstrim` on **/mnt/ext4** mount point.

  ```
  # fstrim /mnt/ext4
  ```

  The `fstrim` utility may fail due to T10 commands blocking (VPLEX/VMDK). In such cases, it is required to apply a manual copy of zeroes to the relevant free space.

  The following example shows running a manual script on /mnt/ext4 mount point (see Linux space reclamation script).

  ```
  # ./reclaim_space.sh /mnt/ext4
  ```

XFS and BTRFS file systems

- Automatic space reclamation - XFS and BTRFS support automatic space reclamation, using the `discard` option. This option should be enabled when mounting the file system (either manually or using `fstab`).
  The following examples show discard usage when mounting a file system:

  - For BTRFS file system:
    ```
    # mount -o discard /dev/mapper/mpathax /mnt/btrFS
    ```

  - For XFS file system
    ```
    # mount -o discard /dev/vg_xfs_test/lv_xfs_test /mnt/xfs
    ```

  The following examples show `fstab` usage:

- ■ For BTRFS file system:
  ```
  /dev/mapper/mpathax /mnt/btrFS btrfs defaults,discard 0 1
  ```
- ■ For XFS file system:
  ```
  /dev/vg_xfs_test/lv_xfs_test /mnt/xfs xfs defaults,discard 0 1
  ```

When automatic space reclamation is not enabled immediately after the file system creation, it is recommended to run a manual initial space reclamation prior to automatic mounting.

- • Manual space reclamation - XFS and BTRFS space reclamation can be performed, using the `fstrim` utility.

  Using the `fstrim` utility does not create a balloon file. Therefore, it is not necessary to review the free space.

  The following example shows running `fstrim` on an XFS or BTRFS file system that is mounted on the Linux host.

  ```
  # fstrim <mount point of file system>
  ```

  The `fstrim` utility may fail due to T10 commands blocking (VPLEX/VMDK). In such cases, it is required to apply a manual copy of zeroes to the relevant free space.

  The following example shows running a manual script on the XFS or BTRFS mount point (see Linux space reclamation script).

  ```
  # python linux-reclaim.py -m <mount point of file system>
  ```

Linux space reclamation script

The following example shows the space reclamation script usage, followed by the space reclamation script for the Linux shell.

```
# python linux-reclaim.py --help
usage: linux-reclaim.py [-c <cluster_name>]

optional arguments:
   -h, --help            show this help message and exit
   -v, --version         show program's version number and exit
   -t, --thread_number   Threads Amount
   -m, --mount_point     Mount Point Name

# python linux-reclaim.py --thread_number 4 --mount_point /mnt/btrFS
#!/usr/bin/env python
from __future__ import print_function
import time
import threading
import subprocess
import argparse
import re


class myThread (threading.Thread):
   def __init__(self, name, count, running):
      threading.Thread.__init__(self)
      self.name = name
      self.count = count
      self.running = running

   def run(self):
      dd_something(self.name, self.count, self.running)

def dd_something(name, count, running):
   execute('/bin/dd count={1} bs=131072 if=/dev/zero
of={0}/xtremio_file-{2}.zf oflag=direct'.format(name, count,
running))
```

```
def execute(command):
    return subprocess.Popen(command, shell=True,
stdout=subprocess.PIPE,
stderr=subprocess.PIPE).communicate()[0].splitlines()


def get_mount_points():
    return execute('df -m')

threadLock = threading.Lock()

def main():
    parser = argparse.ArgumentParser(usage='%(prog)s [-c
<cluster_name>]')
    parser.add_argument('-v', '--version', action='version',
version='%(prog)s 1.02')
    parser.add_argument('-t', '--thread_number', action='store',
dest='thread_number', default='1', required=False, help='Threads
Amount', metavar='')
    parser.add_argument('-m', '--mount_point', action='store',
dest='mount_point', default=None, required=True, help='Mount Point
Name', metavar='')
    args = parser.parse_args()
    thread_number = int(args.thread_number)
    mount_point = args.mount_point
    print('Starting Main Thread {0}'.format(time.ctime(time.time())))
    threads = []

for entry in get_mount_points():
    if mount_point in entry:
        pring(mount_point)
        filesystem, mblocks, used, available, usep, mounted =
map(str.strip, re.split(" +", entry))

for i in xrange(thread_number):
    i = myThread(mounted, int(int(available) * 0.95 /
thread_number) * 8, str(i))
    i.start()
    threads.append(i)
    for t in threads:
        t.join()

    execute('/bin/rm -rf {0}/xtremio_file-*.zf'.format(mounted))
    print('Exiting Main Thread {0}'.format(time.ctime(time.time())))

if __name__ == "__main__":

    exit(main())
```

(i) Note: While Increasing the percentage causes elevated precision, it may increase the probability of receiving a "no free space" SCSI error during the reclamation process.

Configuring boot from SAN with an XtremIO volume

For guidelines on configuring the host to boot from an XtremIO volume, see the *Configuring Boot from SAN with XtremIO* section of *Dell EMC XtremIO Host Configuration Guide*, available on Dell EMC Online Support.

Executing the Linux host validation script on Linux hosts

After all Linux hosts connected to XtremIO are configured according to the recommendations provided in this document, execute the Linux Host Validation Script (HVS) on these hosts to

validate that the host configuration is aligned with these recommendations. For details about using the Linux HVS, see *Dell EMC KB# 499814* available on Dell EMC Online Support.

MPIO configuring for XtremIO storage

This section discusses methods for configuring XtremIO storage. To configure the XtremIO disk device, modify the `/etc/multipath.conf` file depending on your operating system, as described in the following sections:

- RedHat Enterprise Linux (RHEL)
- Oracle Linux and VM server
- SuSE Linux Enterprise server

Red Hat Enterprise Linux

Red Hat Enterprise Linux device-mapper-multipath configuration.

(i) Note: Device Mapper Multipath support and default configuration for an XtremIO storage array are provided with Red Hat Enterprise Linux 7 in the following version:
Device-mapper-multipath version 0.4.9-77.el7.x86_64 or later

All previous versions of Red Hat Enterprise Linux device-mapper-multipath require the following configuration:

```
device {
vendor     XtremIO
product    XtremApp
path_selector    "queue-length 0" (FOR RHEL>=6) path_selector        "round-
robin 0" (FOR RHEL<6) rr_min_io            1000 (FOR RHEL<6)
rr_min_io_rq        1 (FOR RHEL>=6) path_grouping_policy    multibus
path_checker    tur
failback    immediate fast_io_fail_tmo 15
}
```

The sample output of *multipath -ll* is as follows:

```
mpathg (3514f0c5548c004ba) dm-39 XtremIO ,XtremApp size=1.0G features='0'
hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=1 status=active
|- 10:0:0:5    sdj    8:144    active ready running
|- 10:0:1:5    sdaz 67:48    active ready running
|- 8:0:0:5    sdk    8:160    active ready running
`- 8:0:1:5    sdba 67:64    active ready running
```

Oracle Linux and VM Server

All the Oracle Linux versions using the stock Red Hat kernel or Oracle enhanced Red Hat kernel use the same configurations as Red Hat Enterprise Linux.

See the Red Hat Enterprise Linux (RHEL) section for more information.

SuSE Linux Enterprise Server (SLES)

SuSE Linux Enterprise Server sample configuration.

(i) Note: See the SuSE documentation to ensure that the correct syntax is followed for your release. MPIO continues to evolve with each release of SLES. The following is only an example:

```
device {
vendor XtremIO product XtremApp
```

```
path_selector "queue-length 0" rr_min_io 1 (older than SLE11 SP1)
rr_min_io_rq 1 (SLE11 SP1 and above) path_grouping_policy multibus
path_checker tur
failback immediate fast_io_fail_tmo 15
}
```

The output of *multipath -ll* is as follows:

```
mpathy (3514f0c5548c004cc) dm-26 XtremIO ,XtremApp size=2.0G features='0'
hwhandler='0' wp=rw
`-+- policy='queue-length 0' prio=1 status=active
- 10:0:0:13 sdy    65:128    active ready running
|- 10:0:1:13 sdbo 68:32     active ready running
|- 8:0:0:13    sdab 65:176    active ready running
`- 8:0:1:13    sdbs 68:96    active ready running
```

PowerStore storage

This section provides information about the Dell EMC PowerStore storage array host connectivity.
The following topics are described in this section:

- Fibre Channel (FC) configuration

- iSCSI channel configuration

- I/O path behavior

- Multipathing software configuration

- Post configuration steps - Using the PowerStore Array

- Importing external storage to PowerStore

Fibre Channel configuration

This section describes the best practices for attaching hosts to a PowerStore storage array in a
highly available resilient and optimal Fibre Channel Storage Area Network (SAN).The PowerStore
storage array also supports presenting storage to a direct attached host.

Recommended configuration values

The following table provides the recommended configuration values that are related to Fibre
Channel SAN.

Table 33 Recommended configuration values

Validation	Impact	Severity
Use two separate fabrics.	Redundancy	Mandatory
Each host must be zoned to both nodes of each appliance.	Redundancy	Mandatory
Balance the hosts between the nodes of the appliance to provide a distributed load across all target ports.	Redundancy	Recommended
Maximum number of paths per appliance per volume per host: 8	Redundancy	Warning

Table 33 Recommended configuration values (continued)

Validation	Impact	Severity
Recommended number of paths per volume per host: 4	Redundancy	Warning
Link speed should be consistent across all paths to the PowerStore storage array.	Redundancy	Warning
Maximum ISL Hops: 2		

Recommended zoning configuration

This topic provides recommended zoning configuration setting for a Fibre Channel SAN infrastructure.

Follow these guidelines when setting a Fibre Channel SAN infrastructure:

- Use two separate fabrics. Each fabric must be on a different physical FC switch for resiliency.

- Balance the hosts between the two nodes of the appliance.

- The PowerStore storage array can be shipped with various extension modules for Fibre Channel. If your PowerStore storage array contains more than one extension I/O module per node, distribute the zoning among all I/O modules for highest availability and performance.

- Use single initiator zoning scheme: Utilize single-initiator per multiple-target (1:many) zoning scheme when configuring zoning with a PowerStore storage array. It keeps a consistent link speed and duplex across all paths between the host and the PowerStore storage array.

- Host I/O latency can be severely affected by SAN congestion. Minimize the use of ISLs by placing the host and storage ports on the same physical switch. When this is not possible, ensure that there is a sufficient ISL bandwidth, and both the host and PowerStore storage array interfaces are separated by no more than two ISL hops.

- The optimal number of paths depend on the operating system and server information. To avoid multipathing performance degradation, do not use more than eight paths per device per host. It is recommended to use four paths.

- The PowerStore storage array supports direct attachment of hosts to the appliances. PowerStore only support FC Point to Point connection.

- With Federation, it is recommended to zone the host to as many appliances as possible to achieve the best load distribution across the clusters. Ensure that you keep the minimum or optimal zoning recommendations for each appliance.

(i) Note: A federation is designed to provide a better load balancing, and not for better resiliency. To perform volume migration between appliances, a host must be zoned to both appliances.

The following figure shows a simple connectivity

1. Host
2. PowerStore Appliance
3. SAN Switch
4. Node
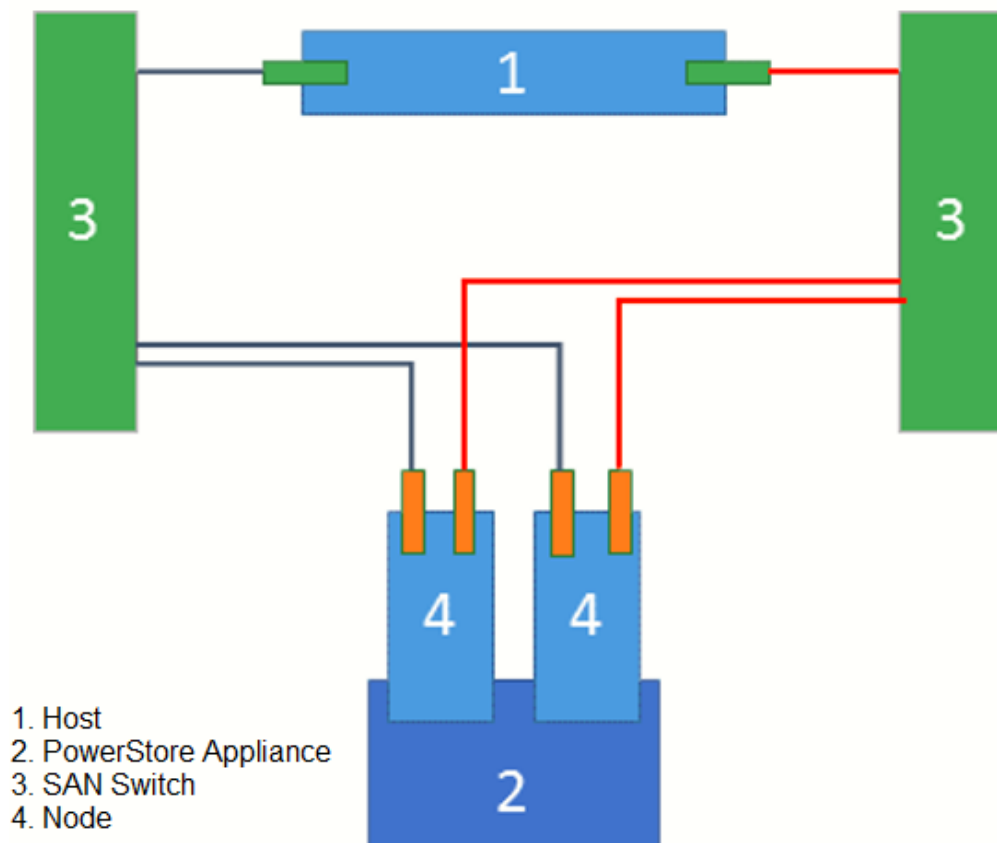
ⓘ **Note:** For implementation instructions, see *User Manual* of the Fibre Channel switch .

iSCSI channel configuration

This section provides best practices for attaching the hosts to a PowerStore storage array in a highly available resilient and optimal iSCSI channel SAN. The PowerStore storage array also supports presenting storage to a direct attached host.

Recommended configuration values summary

The following table summarizes the recommended variables that are related to iSCSI channel:

Table 34 Recommended configuration values for iSCSI SAN

Validation	Impact	Severity
Each host must be connected to both nodes of each appliance.	Redundancy	Mandatory
Balance the hosts between the nodes of the appliance to provide a distributed load across all target ports.	Performance	Recommended
Maximum number of paths per volume per host: 8	Performance	Warning
Recommended number of paths per volume: 4	Performance	Warning
Link speed must be consistent across all paths to the PowerStore storage array.	Performance	Warning

Table 34 Recommended configuration values for iSCSI SAN (continued)

Validation	Impact	Severity
Duplex setting should be consistent across all paths to the PowerStore storage array.	Performance	Warning
Use dedicated NICs or iSCSI HBAs for PowerStore array iSCSI and do not partition the iSCSI interface.	Performance	Recommended
Configuration should not allow traffic between different iSCSI IP subnets.	Performance	Normal
If jumbo frames are required for iSCSI traffic, all ports (server, switches, and storage) must be configured with the correct MTU value.	Performance	Warning
Flow control must be disabled on all ports (server, switches, and storage).	Performance	Warning
Enable the TCP Offloading Engine (TOE) on the host iSCSI interfaces.	Performance	Warning

Recommended configuration

This topic provides recommended configuration for setting an iSCSI SAN infrastructure.

- iSCSI switch recommendations

    - Use nonblocking switches.

    - Use enterprise grade switch.

    - Use minimum 10 GbE interfaces.

- Set configure network port speed to a specific value rather than allowing speed negotiation.

- Use dedicated iSCSI switches. If not possible, use dedicated iSCSI VLANs on the switches.

- Ensure that the different iSCSI IP subnets cannot transmit traffic between them.

- Disable flow control features on the servers, switches, and array ports.

- It is recommended to use dedicated NICs or iSCSI HBAs for PowerStore storage array iSCSI, and not to partition the iSCSI interface (that is disable NIC Partitioning - NPAR).

- Enable the TCP Offloading Engine (TOE) on the host iSCSI interfaces to offload the TCP packet encapsulation from the host's CPU to the NIC or iSCSI HBA, and free up CPU cycles.

- When using PowerStore iSCSI, it is recommended to use interfaces individually rather than using NIC Teaming (Link Aggregation), to combine multiple interfaces into a single virtual interface.

- If Jumbo Frames are required for iSCSI traffic, ensure that all ports (servers, switches, and array) are configured with the correct MTU value.

- Avoid connecting target ports of different speed which could bring I/O issue to host side.

(i) Note: For implementation instructions, see iSCSI switch *User Manual*.

iSCSI host configuration

The first release of PowerStore storage array supports only one network subnet for iSCSI connection. To ensure path redundancy, add static route for each NIC and bind initiator target

pairs together. NIC teaming method is not recommended as it might have some potential issue with iSCSI.

The following figure shows two types of connection:

Figure 43 SP Connection types

Cross-SP Connection   Dedicated-SP Connection
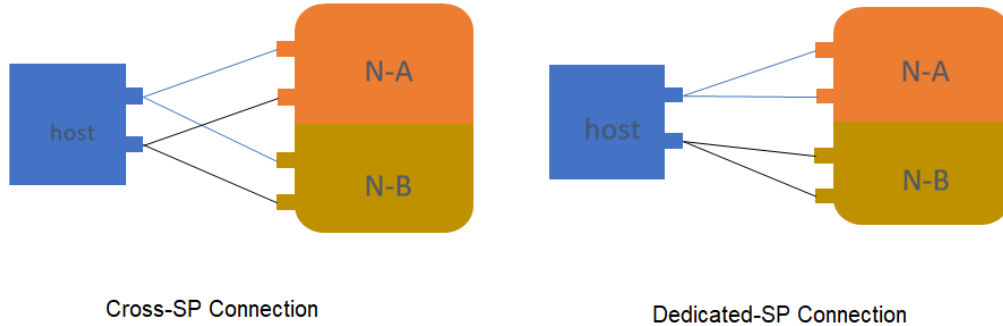
The following table provides the difference between cross-SP connection and dedicated-SP connection:

Table 35 Difference between cross-SP connection and dedicated-SP connection

I/O active path after cable pull	Host cable pull	Target cable pull	Influence
Cross-SP connection	Initially two active I/O paths from two NIC, it only leaves 1 active path for I/O	Totally three active paths left. With different LUN owner of node, some LUN has one active I/O path and some have two paths.	Better performance in normal state. Two NIC bandwidth used.
Dedicated-SP connection	Initially two active I/O paths from one NIC, then two active I/O path failovers to other NIC.	Totally three paths left. With different LUN owner of node, some LUN has one active I/O path and some have two paths.	Better Redundancy

Path failure examples

- Cross-SP connection:

```
mpathaa (368ccf09800aed8f3c64b27040cd586e8) dm-5 DellEMC,PowerStore
size=10G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='queue-length 0' prio=50 status=active
| |- 15:0:0:2  sdz  65:144 active ready running
| `- 10:0:0:2  sds  65:32  active ready running
`-+- policy='queue-length 0' prio=10 status=enabled
  |- 18:0:0:2  sdal 66:80 active ready running
  `- 9:0:0:2   sdo  8:224  active ready running
```

- Disable one host port:

```
mpathaa (368ccf09800aed8f3c64b27040cd586e8) dm-5 DellEMC,PowerStore
size=10G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='queue-length 0' prio=50 status=active
| |- 15:0:0:2  sdz  65:144 active ready running
| `- 10:0:0:2  sds  65:32  failed faulty running
`-+- policy='queue-length 0' prio=10 status=enabled
  |- 18:0:0:2  sdal 66:80  failed faulty running
  `- 9:0:0:2   sdo  8:224  active ready running
```

- Dedicated-SP connection:

```
mpathaa (368ccf0980038737d4c196bf464458913) dm-4 DellEMC,PowerStore
size=10G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='queue-length 0' prio=50 status=active
| |- 17:0:0:4   sdx  65:112 active ready running
| `- 18:0:0:4   sdab 65:176 active ready running
`-+- policy='queue-length 0' prio=0 status=enabled
  |- 15:0:0:4   sdp  8:240  active ready running
  `- 16:0:0:4   sdt  65:48  active ready running
```

- Disable one host port:

```
mpathaa (368ccf0980038737d4c196bf464458913) dm-4 DellEMC,PowerStore
size=10G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='queue-length 0' prio=50 status=active
| |- 17:0:0:4   sdx  65:112 failed faulty running
| `- 18:0:0:4   sdab 65:176 failed faulty running
`-+- policy='queue-length 0' prio=0 status=enabled
  |- 15:0:0:4   sdp  8:240  active ready running
  `- 16:0:0:4   sdt  65:48  active ready running
```

iSCSI initiator configuration

Procedure

1. Add static route into network configuration file and reboot host. Run the following command:

   `/etc/sysconfig/network-scripts/route-enp3s0f1`

   `172.16.11.42 dev enp3s0f1`

   Host local route table

   ```
   default via 10.228.100.1 dev eno3 proto static metric 102
   10.228.100.0/24 dev eno3 proto kernel scope link src 10.228.100.26
   metric 102
   172.16.10.0/23 dev enp3s0f0 proto kernel scope link src 172.16.10.169
   metric 104
   172.16.10.0/23 dev enp3s0f1 proto kernel scope link src 172.16.10.163
   metric 105
   172.16.11.41 dev enp3s0f0 proto static scope link metric 104
   172.16.11.42 dev enp3s0f1 proto static scope link metric 105
   172.16.11.45 dev enp3s0f0 proto static scope link metric 104
   172.16.11.46 dev enp3s0f1 proto static scope link metric 105
   ```

2. Discovery iSCSI target. It is recommended to use `iface` name to configure NIC separately.

   ```
   iscsiadm --m iface -I iscsi-enp3s0f1 --o new
   iscsiadm --mode iface -I iscsi-enp3s0f1 -o update -n
   ```

```
iface.net_ifacename -v enp3s0f1
iscsiadm –m discovery -t sendtargets -p 172.16.11.42:3260 -I iscsi-
enp3s0f1
```

3. Delete unusable target (discovery will list all iscsi target portal), which makes session numbers consistent after reboot. Run the following command:

```
iscsiadm -m node -p 172.16.11.48:3260 -I iscsi-enp3s0f1 -o delete
```
Or

```
iscsiadm -m node -T iqn.2015-10.com.dell:dellemc-powerstore-
fnm00185000718-a-0cc9e809 -I iscsi-enp3s0f1 -o delete
```

Expected result of iSCSI node (each target portal only have one `Iface` that is connected):

```
iscsiadm -m node  -P 1
Target: iqn.2015-10.com.dell:dellemc-powerstore-fnm00185000060-a-2795b26
        Portal: 172.16.11.41:3260,1
                Iface Name: iscsi-enp3s0f0
Target: iqn.2015-10.com.dell:dellemc-powerstore-fnm00185000060-
a-7fbbd45d
        Portal: 172.16.11.45:3260,1
        Iface Name: iscsi-enp3s0f0
Target: iqn.2015-10.com.dell:dellemc-powerstore-fnm00185000060-
b-0d707232
Portal: 172.16.11.42:3260,1
        Iface Name: iscsi-enp3s0f1
Target: iqn.2015-10.com.dell:dellemc-powerstore-fnm00185000060-
b-271cd430
Portal: 172.16.11.46:3260,1
        Iface Name: iscsi-enp3s0f1
```

4. Run iSCSI login.

```
iscsiadm –m node -T  iqn.2015-10.com.dell:dellemc-powerstore-
fnm00185000060-a-7fbbd45d -I iscsi-enp3s0f1 –login

iscsiadm –m session
tcp: [13] 172.16.11.41:3260,1 iqn.2015-10.com.dell:dellemc-powerstore-
fnm00185000060-a-2795b26f (non-flash)
tcp: [14] 172.16.11.45:3260,1 iqn.2015-10.com.dell:dellemc-powerstore-
fnm00185000060-a-7fbbd45d (non-flash)
tcp: [15] 172.16.11.42:3260,1 iqn.2015-10.com.dell:dellemc-powerstore-
fnm00185000060-b-0d707232 (non-flash)
tcp: [16] 172.16.11.46:3260,1 iqn.2015-10.com.dell:dellemc-powerstore-
fnm00185000060-b-271cd430 (non-flash)
```

5. Map LUN to host initiator.

I/O path behavior

The PowerStore storage system supports only implicit ALUA access as described in this topic.

- The default LUN owner is assigned during LUN creation. If the default owner of the LUN is not accessible from the host, the LUN is not trespassed to the other storage processor as Unity.

- By default, I/O paths on LUN owner are shown as *Active/Optimized* and paths on LUN non-owner are shown as *Active/Non-Optimized*. I/O only runs to *Active/Optimized* paths when using standard MPIO settings.

- The *Active/Non-optimized* path will only handle I/O during a node failover and all *Active/ Optimized* paths failure.
- If default LUN owner node is down, all ports on it are down, and MPIO sends I/O to the other node. However, on PowerStore, the path state on the other node remains *Non-Optimized*. This is a difference from Unity series.

Multipathing software configuration

MPIO configuration

- **Red Hat Enterprise Linux**
  This section provides information for configuring the PowerStore storage array for RHEL 7.*x* and later.

  The following is an example of `multipath.conf` to support the PowerStore array:

```
## Use user friendly names, instead of using WWIDs as names.
defaults {
        user_friendly_names yes
    }
devices {
    device  {
        vendor "DellEMC"
        product "PowerStore"
        path_selector "queue-length 0"
        path_grouping_policy "group_by_prio"
        path_checker "tur"
        detect_prio "yes"
        failback "immediate"
        no_path_retry 3
        rr_min_io_rq 1
        hardware_handler "1 alua"
        prio "alua"
        fast_io_fail_tmo 15
        max_sectors_kb          1024      #only for RHEL 6.9 and above
        }
# other devices
}
```

> (i) **Note:**
>
> In PowerStore array, the maximum I/O size is 1 MB. PowerStore does not set an optimal transfer size.
>
> To make Trident import successful, the keyword blacklist and curly brace should appear in the same line (`blacklist {`) in the `/etc/multipath.conf` file.
>
> If the `multipath.conf` file has blacklist keyword, such as devices section before product_blacklist section, you must move devices section after the blacklist section for imports to work successfully.

The following output is a sample output from multipath-ll :

```
mpathaa (368ccf09800aed8f3c64b27040cd586e8) dm-5 DellEMC,PowerStore
size=10G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='queue-length 0' prio=50 status=active
| |- 15:0:0:2   sdz  65:144 active ready running
| `- 10:0:0:2   sds  65:32  active ready running
`-+- policy='queue-length 0' prio=10 status=enabled
  |- 18:0:0:2   sdal 66:80 active ready running
  `- 9:0:0:2    sdo  8:224  active ready running
```

- **SuSE Linux Enterprise Server (SLES)**

SLES can apply same MPIO configuration file as Red Hat Enterprise Linux. See *MPIO configuring-Red Hat Enterprise Linux*.

PowerPath configuration

Linux PowerPath 7.1 supports PowerStore. For detail operating system support, see *Dell EMC Simple Support Matrix*, available on E-Lab Navigator.

The following output is a sample output from multipath-ll :

```
Pseudo name=emcpowerw
PowerStore ID=FNM00185000060
Logical device ID=68CCF09800E2B486A23F7B0A56E46EB3
state=alive; policy=ADaptive; queued-IOs=0
========================================================================
---- Host ----   - Stor -  ----------- I/O Path ------------   -- Stats ---
### I/O Paths   Interf.   ALUA State           Mode    State   Errors
========================================================================
   8 sdcd        A0-NA-i11 Active/non-optimized  active  alive     2
   7 sdce        A0-NA-i9  Active/non-optimized  active  alive     8
  10 sdcc        A0-NB-i9  Active/optimized      active  alive     2
   9 sdcb        A0-NB-i11 Active/optimized      active  alive     2
```

Post configuration steps - Using the PowerStore array

After the host configuration is completed, you can access the PowerStore array from the host. For details about creating, presenting, and managing volumes that are accessed from the host using either the GUI or the CLI, see the *PowerStore Array Storage User Guide* that matches the version running on your PowerStore array.

Presenting PowerStore array volumes to the Linux host

This topic provides guidelines for presenting the PowerStore array volumes to the Linux host.

- When adding host groups and hosts to allow Linux hosts to access PowerStore array volumes, specify Linux as the operating system for the newly-created hosts.
  The following figure demonstrates setting the operating system field for a host using the web-based user interface:

  Figure 44 Setting the operating system field

- Setting the host's operating system is required for optimal interoperability and stability of the host with PowerStore array. You can adjust the setting while the host is online and connected to the PowerStore array with no I/O impact.

ⓘ Note: For more information, see the PowerStore storage array *User Guide* that matches the version running on your PowerStore array cluster.

Partition alignment in Linux

Use the parted command to create partition alignment on a PowerStore array volume.

To align partitions on PowerStore array volumes that are presented to Linux hosts, use the default value (2048), and create a partition using the parted command to ensure that the file system is aligned.

When performing partition alignment, the logical device (`/dev/mapper/<naa-name>`) must be used rather than the physical device (`/dev/<device>`). When multipathing is not used (for example in a virtual machine), the physical device should be used.

The following example demonstrates creating an aligned partition on a PowerStore array volume using the parted command:

```
parted /dev/mapper/mpathd
GNU Parted 3.2
Using /dev/mapper/mpathd
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) unit s
(parted) mkpart primary ext4 0% 50%
(parted) p
Model: Linux device-mapper (multipath) (dm)
Disk /dev/mapper/mpathd: 4194304s
Sector size (logical/physical): 512B/512B
Partition Table: gpt
Disk Flags:

Number  Start   End       Size      File system  Name      Flags
 1      2048s  2097151s  2095104s  ext4          primary

(parted)  align-check  opt 1
1 aligned
```

In this mode, rather than using *cylinders* for creating partitions, the parted command uses sector unit, which is a direct mapping to the array's LBA space. Thus, verifying that the partition aligned is simply a case of verifying that the starting sector number is a multiple of 16 (16 sectors, at 512 bytes each, is 8 KB). The parted command will default to a starting sector for the first partition of 2048, which is divisible by 16, and is correctly assigned.

Creating a file system

It is recommended to create the file system using its default block size. Using a non-default block size may lead to unexpected behavior. For detailed procedure, see the operating system and file system documentation.

For a more efficient data utilization and better performance, use Ext4 file system with XtremIO storage instead of Ext3. For details about converting to Ext4 file system (from either Ext3 or Ext2), see Upgrade to Ext4.

Maximum transfer length

The PowerStore Storage Array reports a maximum transfer length of 1 MB to the Linux host.

Linux kernels running versions earlier than 4.3 (sometimes earlier than 3.1) do not check the optimal transfer length and use a default value of 512 KB.

In parallel around RHEL 6.9 another change was introduced, where the kernel would check with the array if it exposes any queue settings and set its internal queue settings accordingly maximum

number of sectors (usually 512 bytes) allowed in a single SCSI command. This value could be overwritten, but it must be smaller than or equal to the maximum size allowed by the hardware.

```
sg_inq -p 0xb0  /dev/mapper/mpathi
VPD INQUIRY: Block limits page (SBC)
  Maximum compare and write length: 1 blocks
  Optimal transfer length granularity: 16 blocks
  Maximum transfer length: 2048 blocks
  Optimal transfer length: 0 blocks [not reported]
  Maximum prefetch transfer length: 0 blocks [ignored]
  Maximum unmap LBA count: 2048
  Maximum unmap block descriptor count: 16
  Optimal unmap granularity: 16 blocks
  Unmap granularity alignment valid: true
  Unmap granularity alignment: 0
  Maximum write same length: 0x800 blocks
```

However, if you need to change the settings to match the previous kernel settings, use the following command to change the maximum block size setting.

```
echo 512 > /sys/block/sda/queue/max_sectors_kb
```

Using LUN 0 with PowerStore storage

This topic provides guidelines and steps you must performed when using LUN 0 with Linux.

About this task

When a LUN ID 0 is not configured and the PowerStore volume is not assigned to an initiator and target pair, the PowerStore array by default exposes a Storage Array Controller Device (SACD) device as LUN ID 0.
Consider the following guidelines when performing LUN numbering:

- Volumes are numbered starting from LUN ID 1 by default.

- It is recommended not to manually adjust the LUN ID to 0, as it can lead to issues with some operating systems.
  - (i) Note: If LUN 0 is required for a PowerStore volume (for example; boot from SAN, PowerPath), restart the Linux host if a rescan fails to locate this volume.

The following output to an `lsscsi` command indicates that the device at 9:0:0:0 is the PowerStore appliance controller:

`[9:0:0:0] storage DellEMC PowerStore 0100 –`

Select a volume and change the logical unit id to `0`.

In that case, a PowerStore volume with LUN 0 will be inaccessible to the host.

To access a volume with a LUN 0 on a Linux host:

Procedure

1. Run one of the following commands to remove the controller device:
   - `# /usr/bin/rescan-scsi-bus.sh -r`
     - (i) Note: -r enables the device removal.

- `# echo 1 > /sys/class/scsi_device/9:0:0:0/device/delete`

2. Run the following command to rescan:

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

   (i) Note: In older Linux versions a host reboot is required instead of a rescan.

```
mpathl (368ccf098004d5860f40b812e1be1216a) dm-21 DellEMC,PowerStore
size=10G features='1 queue_if_no_path' hwhandler='1 alua' wp=rw
|-+- policy='queue-length 0' prio=50 status=active
| `- 7:0:0:0  sdw 65:96  active ready running
`-+- policy='queue-length 0' prio=10 status=enabled
  `- 9:0:0:0  sdx 65:112 active ready running
```

Space reclamation

The PowerStore array is sensitive in the current code level to massive UNMAP operations. This can be caused by any of the following:

- File system creation
- Deleting VMDK on VMFS-6 datastore
- Running manual space reclamation

The impact varies depending on the load, operating system, and several other factors. However, in most cases, we can observe a huge degradation in performance and high latency.

At the moment, it is recommended to disable all UNMAP operations on all hosts. For Linux, add the flag `nodiscard` when creating file system.

`mke2fs -t ext4 -E nodiscard`

Importing external storage to PowerStore

PowerStore provides the abilities of a traditional storage appliance and onboard compute to run embedded workloads.

Importing external storage to PowerStore is a migration solution that imports data from any of the following Dell EMC storage platforms to a PowerStore cluster:

- Dell EqualLogic Peer Storage (PS) Series
- Dell Compellent Storage Center (SC) Series
- Dell EMC Unity Series
- Dell EMC VNX2 Series

Import of the following block storage resources is supported:

- LUNs and volumes
- Consistency groups and volume groups
- Thick and thin clones

(i) Note:
Import of file storage resources and import with PowerPath is not supported.

For PowerStore importing configuration, see *Importing External Storage to PowerStore Guide* available on Dell EMC Online Support.

Known issues

This section lists known issues of PowerStore, along with workarounds.

Table 36 Known issues

Problem description	Environments	Workaround/Fix
When LUN migration between federation appliance is completed, the native block devices from the destination appliance are added in the kernel by `emcp_add`, but they are not configured . Meanwhile, all the source paths become dead. As the source paths are the only paths present for the LUNs, as they have now become dead, I/Os will start failing.	RHEL with PowerPath	1. Run `rescan_scsi_bus.sh` and then `powermt config`. 2. Use the following PowerPath bundled script: `/etc/opt/emcpower/ emcplun_linux scan all - noprompt` It scans the bus and perform a `powermt config` internally.

CHAPTER 5

Operating System Specific Features

This chapter includes the following topics:

Booting from SAN

Installing and booting Linux from a Storage Area Network (SAN) environment is supported on all Dell EMC storage arrays. The Linux host operating system can reside on an external device that is managed by Linux native DM-MPIO utility or PowerPath software. This chapter provides a summary of configuration process and considerations to prevent possible issues and includes the following information:

- Supported environments
- Limitations and guidelines
- Preparing host connectivity
- Configuring a SAN boot for FC attached host
- Configuring SAN boot for iSCSI host
- Configuring SAN boot for FCoE attached host
- Multipath booting from SAN
- PowerPath booting from SAN
- Guidelines for booting from Symmetrix, XtremIO, VNX series, VNXe series or Unity series

Supported environments

Dell EMC storage environments such as PowerMAX and VMAX series, VNX, Unity, VPLEX, XtremIO, and PowerStore are supported.

See Dell EMC Simple Support Matrix for a list of operating system kernels supporting booting from SAN-attached storage.

Dell EMC supported Emulex, QLogic, and Brocade HBAs can be used to boot from SAN. To boot from storage that is attached to the SAN environment, the host bus adapter's Boot BIOS must be installed and enabled on the adapter. See the driver manuals and configuration guides that are found on the Dell EMC section of the Emulex (Broadcom), QLogic, and Brocade websites as well as operating system, HBA, and server vendor documentation.

(i) Note:

- The AX100/100i, AX150/150i are supported only with the low-cost HBAs. See Dell EMC Simple Support Matrix for supported HBAs with these arrays.
- iSCSI booting from SAN is supported in limited configurations. See Dell EMC Simple Support Matrix for supported environments.

Limitations and guidelines

Boot configurations must not deviate from the following limitations that are established by Dell EMC:

- The Dell EMC Storage device must have enough disk space to hold the Linux operating system.
- Space reclamation, available with Enginuity 5874 and later, is prohibited for use on PowerMAX and VMAX Virtual Provisioning (thin) devices which are used for host /boot, / (root), /swap, and /dump volumes.
- Boot from SAN using Linux DM-MPO when configured to VNX series storage is supported with PNR devices and ALUA devices. Boot from a SAN using Linux DM-MPO when configured to

PowerMAX and VMAX series storage that is supported with AA devices and ALUA devices. Boot from a SAN using Linux DM-MPO when configured to Unity series where storage is supported with ALUA devices. See the restrictions and notes in Multipath booting from SAN.

- Boot from a SAN using Linux DM-MPO when configured to PowerStore series where storage is supported with ALUA devices. See the restrictions and notes in Multipath booting from SAN.

- When PowerPath is used in a boot from SAN configuration, you must use the Linux GRUB boot loader. LILO and eLILO are not currently supported in a PowerPath boot from SAN configuration.

Preparing host connectivity

This section contains guidelines to prepare for host connectivity and an example of single and dual path configuration.

Guidelines

The following guidelines should be followed for host connectivity to SAN environments:

- Maintain the simplest connectivity configuration between host server and SAN environment before installing operating system. The configuration can be altered after installation.

- If multiple HBAs are attached to the host, ensure it is the HBA connected to the lowest-numbered PCI slot that is zoned to the array.

- All arrays, except the array where the boot device resides, should be un-zoned from the host server. On the array where the boot device resides, there should only be the boot device that is attached to the host server.

- It is recommended that the boot LUN be assigned Host LUN ID 0. If the boot LUN has taken a Host ID other than 0, there is possibility for HBA BIOS installation failure, hence no visibility to the boot LUN.

- The boot LUN's Host ID on a VNX series can be forced to 0 by removing all the other LUNs from the storage group and adding only the boot LUN back to it.

  ▪ For Symmetrix, the Symmetrix LUN base/offset skip adjustment (symmask set lun offset) capability can be used to assign LUN 0 to the desired boot LUN if necessary.

  ▪ For Unity series or VNXe series, you can modify Host LUN ID directly in Unisphere UI.

  ▪ For the PowerMAX and VMAX series, when you add devices to SG, you can use the switch -lun 0. (symaccess -sid xxx -type stor -name host_s- g_name add devs lun_id -lun 0) command.

- In XtremIO version 4.0.0 or later, volumes are numbered by default starting rom LUN ID 1. it is not recommended to manually adjust the LUN ID to 0, as it may lead to issues with some operating systems. In XtremIO 3.x and earlier versions, LUN ID starts from 0, and remains accessible when XtremIO cluster is updated from 3.0.x to 4.x.

- More LUNs can be added after operating system installation is completed.

- When configuring a RHEL boot LUN under the Linux native DM-MPIO, it may require that either all paths to the boot device be available or only a single path be available during the installation. For details, see the Red Hat RHEL installation documentation.

- When configuring a SLES boot LUN under the Linux native DM-MPIO, it expects only a single path to the boot device during installation with DM-MPIO. For details, see the SuSE SLES installation documentation.

- When configuring a Linux boot LUN under PowerPath control, the boot LUN may be configured using only a single path during the operating system installation; however, you need

to add the additional paths for PowerPath control. For details, see the PowerPath installation documentation.

Single and dual path configuration examples

The following figure shows an example of single path configuration:

Figure 45 Single path configuration
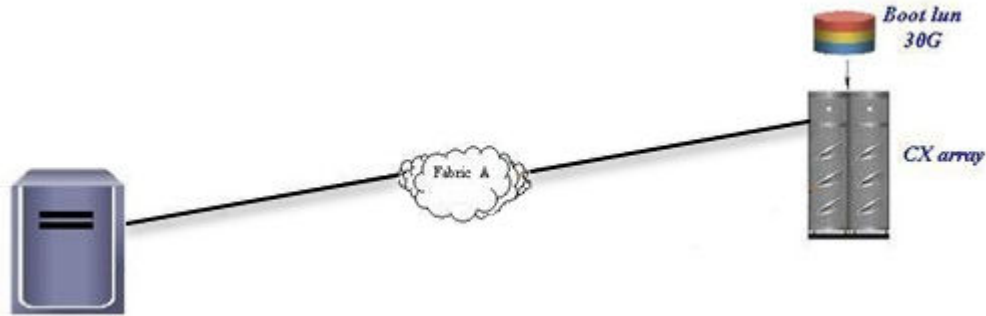
The following figure shows an example of dual path configuration:

Figure 46 Dual path configuration
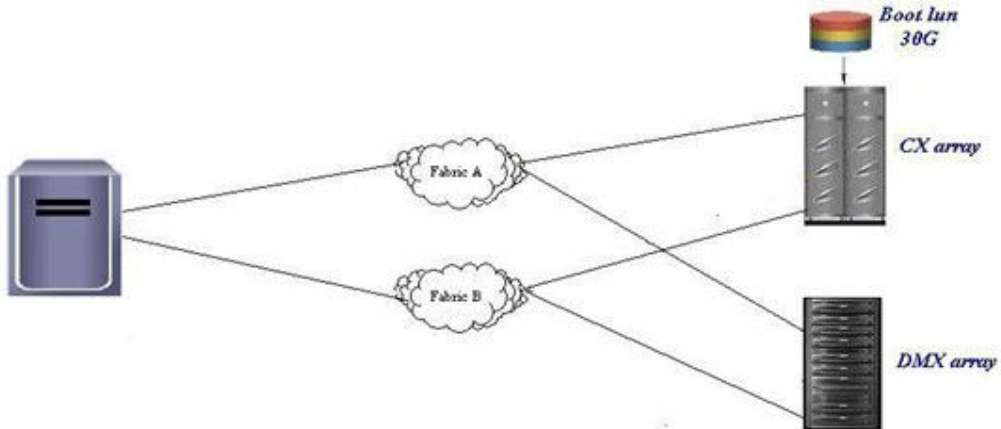
Configuring a SAN boot for FC attached host

Prepare host connectivity

See Fibre Channel Connectivity for instructions on preparing a host FC connection to Dell EMC storage.

Installing and configuring Fibre Channel HBA

After you make a connection between the host and boot LUN, HBA BIOS needs to be enabled and installed to get a boot from a SAN to work. This is because when the host operating system kernel resides on an external device, it will not be loaded during a boot by the host system's hard disk controller. The operating system image can only be fetched by the HBA's BIOS. To facilitate visibility of external boot device by the HBA, therefore, the HBA BIOS must be installed to register the external device as the boot source.

The three major HBA vendors, Emulex, QLogic, and Brocade have embedded HBA configuration utilities in their product BIOS, and can be accessed by a shortcut key during server boot up. For more details, see the vendor's documentation.

Updating HBA BIOS and firmware

Before you configure HBA for boot from SAN, verify that HBA BIOS and firmware have been updated to the Dell EMC supported version that is available in the Dell EMC OEM section of the Broadcom, QLogic, and Brocade websites, and on Dell EMC Online Support.

Enabling HBA port and Selecting boot LUN

The figure in step 1 is an example for setup RHEL 5.3 with Emulex HBA, and is done by single path connection during installation. Different vendor HBA and operating system versions might be similar but require slight modifications. See HBA and operating system vendor websites for more details.

Procedure

1. Boot the server, and press **ALT-E** to enter Emulex BIOS when you see the message as shown in the following figure:

2. Select the adapter port that you want to configure. In a single path configuration, this should be the HBA port zoned to storage.

3. If the link status is shown as **Link UP**, select **Enable/Disable Boot from SAN** to enable the HBA BIOS.
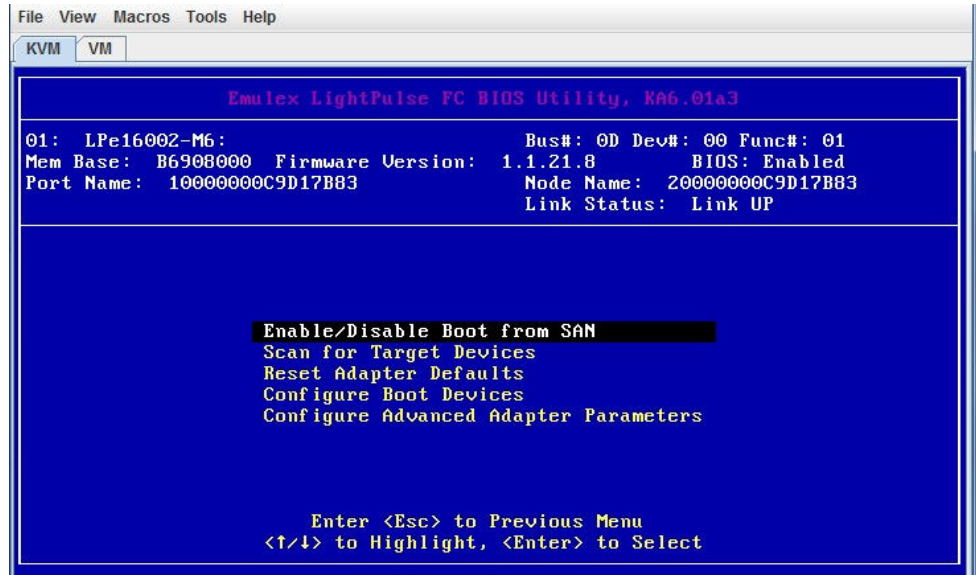
4. To enable the boot BIOS on the adapter port, select **Enable**.
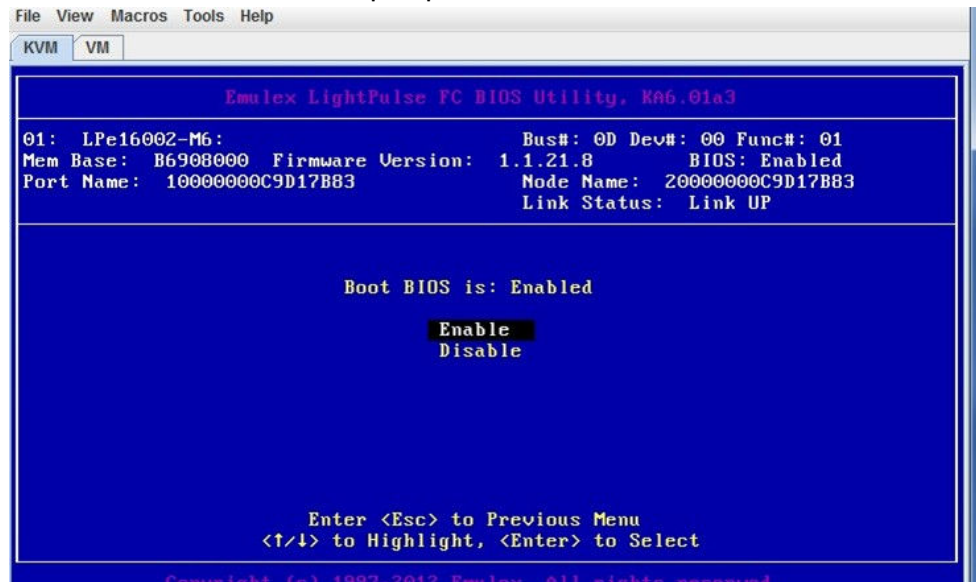
5. After BIOS is enabled, scan for available target devices.
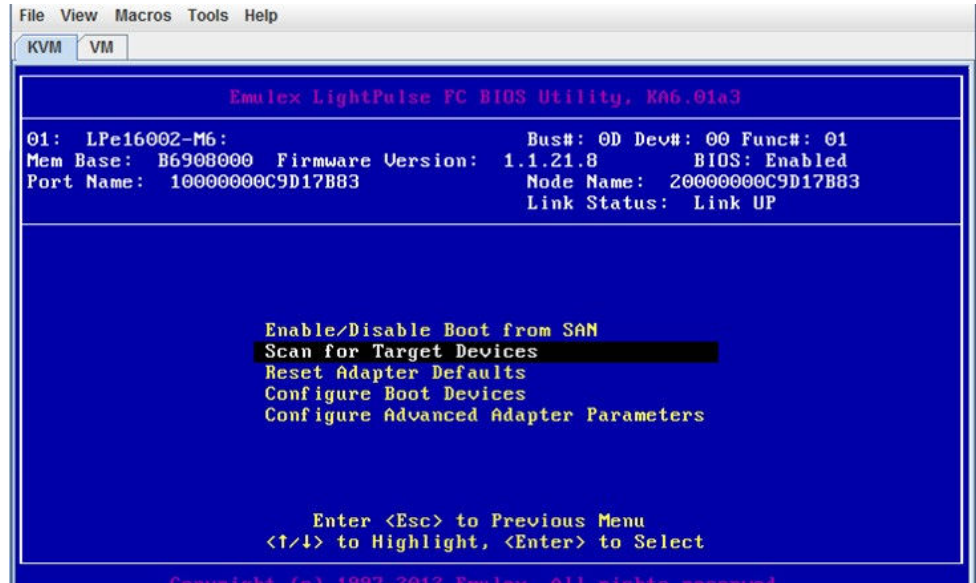
All of the attached LUNs should be listed after scanning. The example in the following figure shows the attached boot LUN, which is a VNX RAID3 LUN.
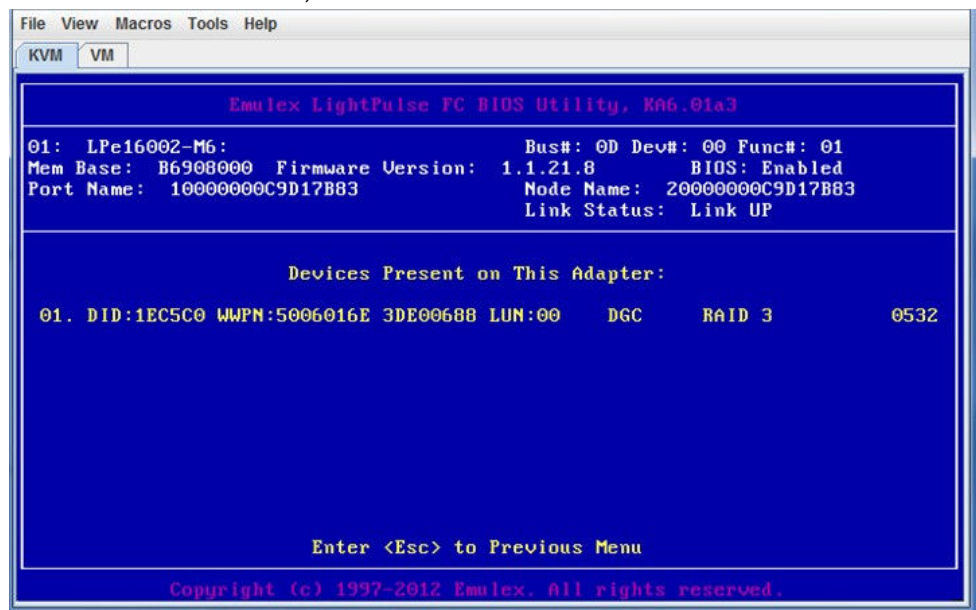
6. When the boot LUN is visible to the HBA, select **Configure Boot Devices**.
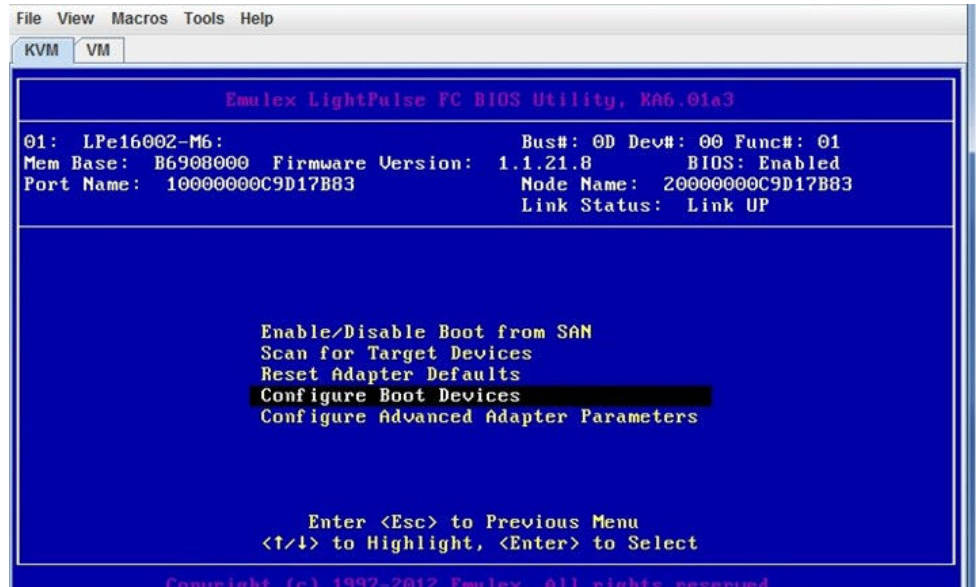
7. Review the list of boot devices, and select the required LUN as the primary boot.
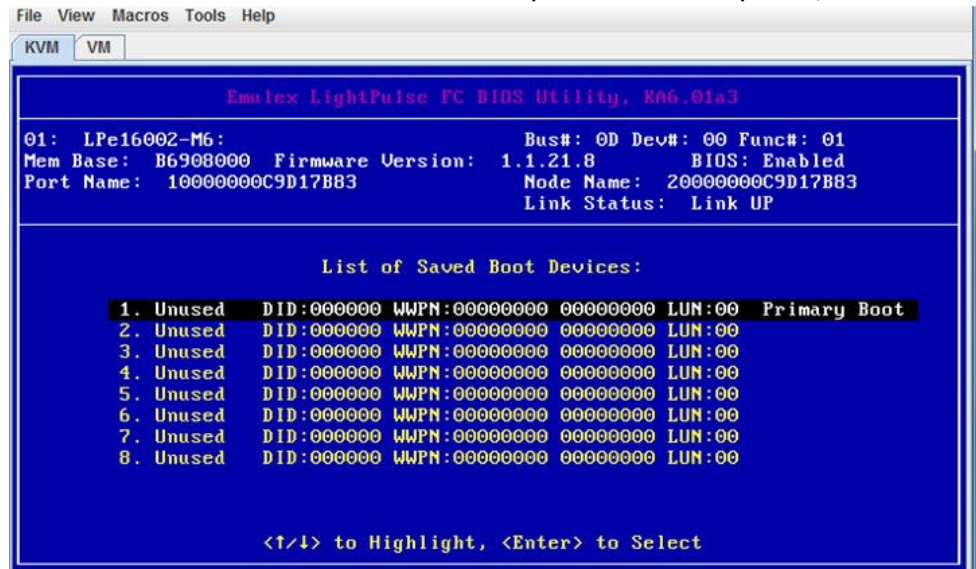
You can view the device for boot details as shown in the following figure:
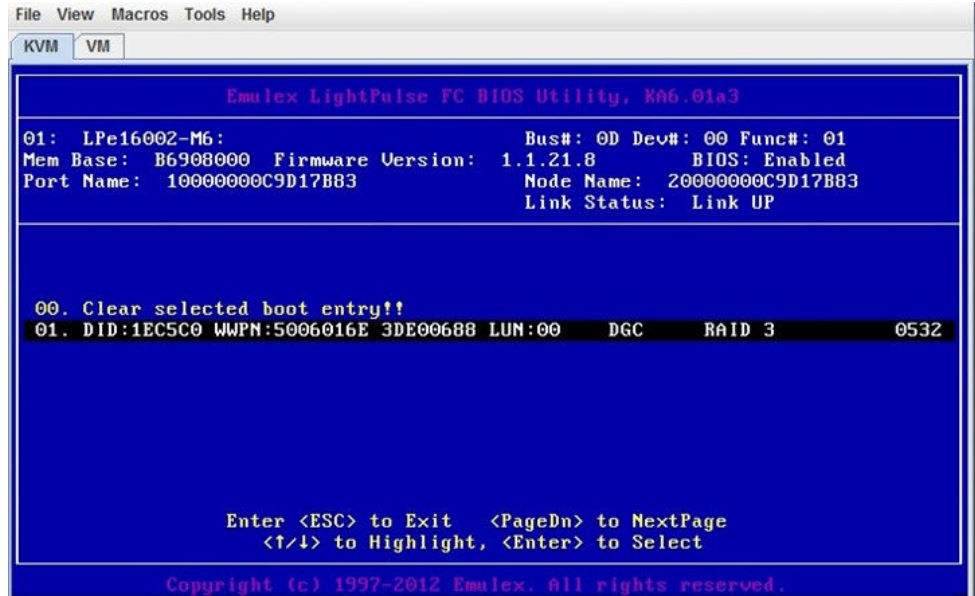
8. Set the boot LUN ID as 00.
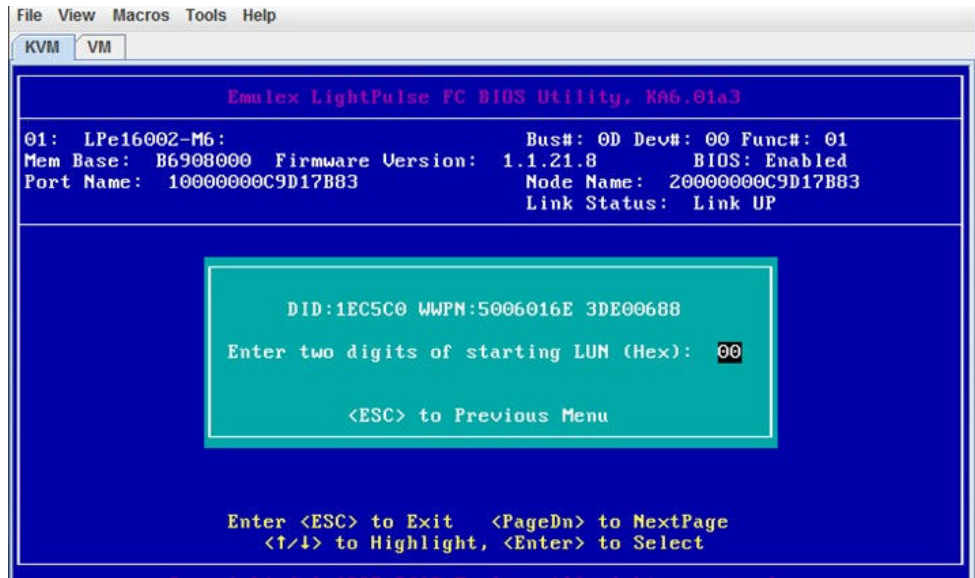
9. Save the changes, and reboot the system.

   If configuration is completed successfully, the information for the selected boot device is displayed after the welcome banner when the server boots, with a BIOS is successfully installed message, as shown in the following figure . The server is then ready for operating system installation.

Configuring SAN boot for iSCSI host

Systems can boot to iSCSI LUNs primarily in two ways:

- A system can use an option ROM during POST to connect to the remote iSCSI LUN and boot to the LUN as if it is a local disk. This is called Hardware iSCSI BFS.
- The other way is Software iSCSI BFS to provide the target system access to a kernel and an *initramfs* image with a *bootloader*, which gets the *initramfs* image to connect to the iSCSI LUN and boot to the iSCSI LUN. The advantage of this option is that the target system does not need any extra hardware (option ROMs or HBAs). The kernel and *initramfs* images that the target system needs can also be served in multiple ways. They can be provided by the *bootloader* on a local hard drive, using an ISO image, or using PXE server.

Setting up the hardware iSCSI SAN boot

Procedure

1. Press **Ctrl + S** when prompted to enter the Broadcom MBA setup.

   (i) Note: The figures in this section show the setup of RHEL 5.4 on a Dell EMC server with Broadcom NICs.

2. Press **Ctrl + K** to enter the iSCSI configuration.

3. Select the primary device.
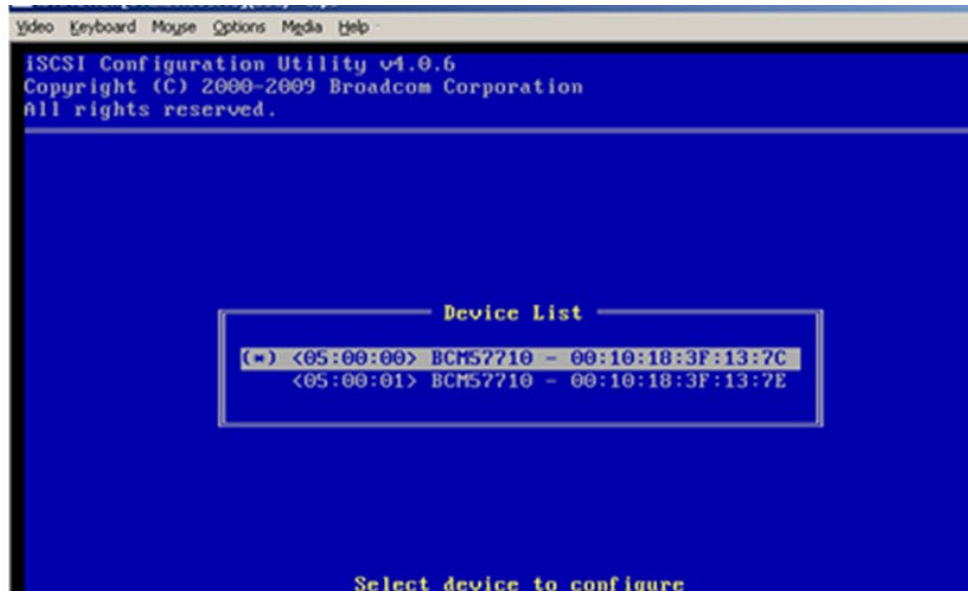
4. Select **General Parameters**.
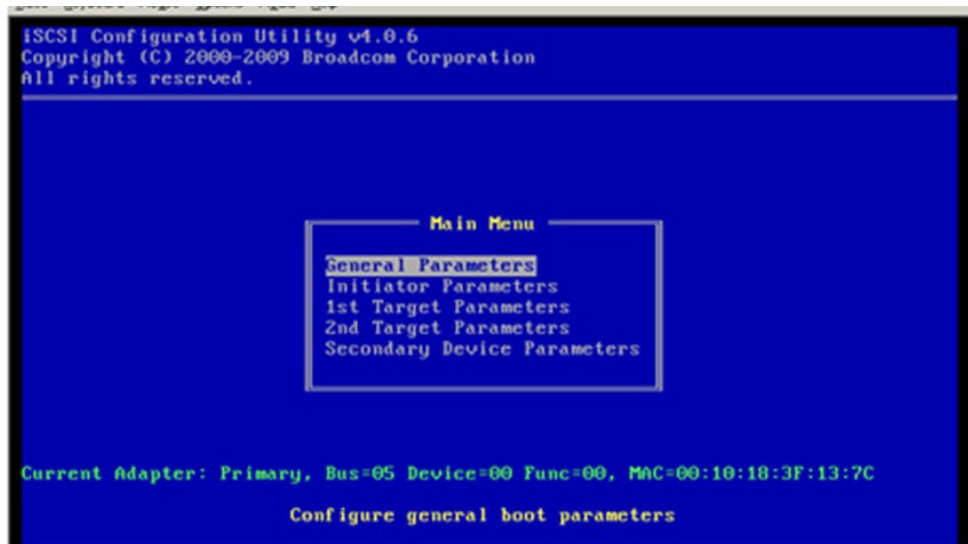
5. In **General Parameters**, change the **Boot to iSCSI target** parameter to `Disabled`.
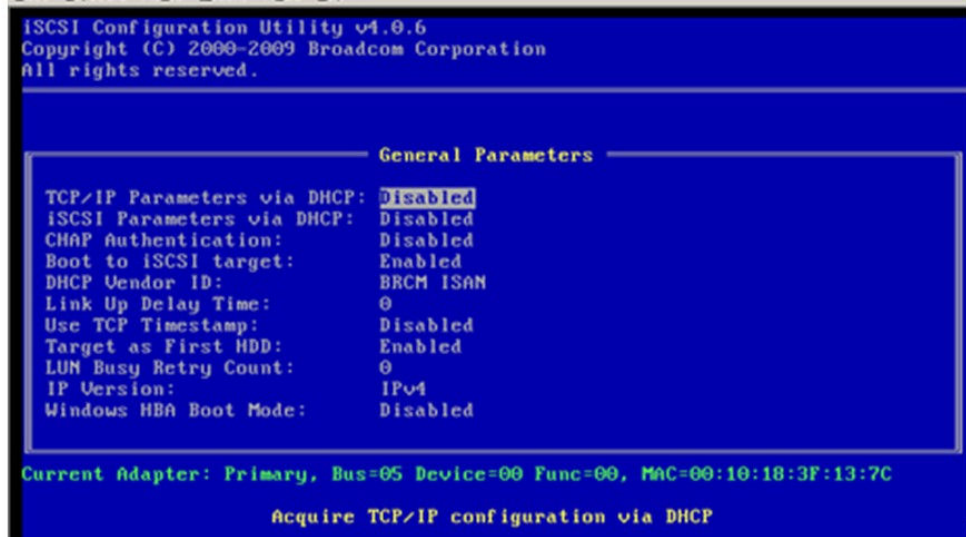
6. Set up the **Initiator Parameters**, as shown in the following figure:
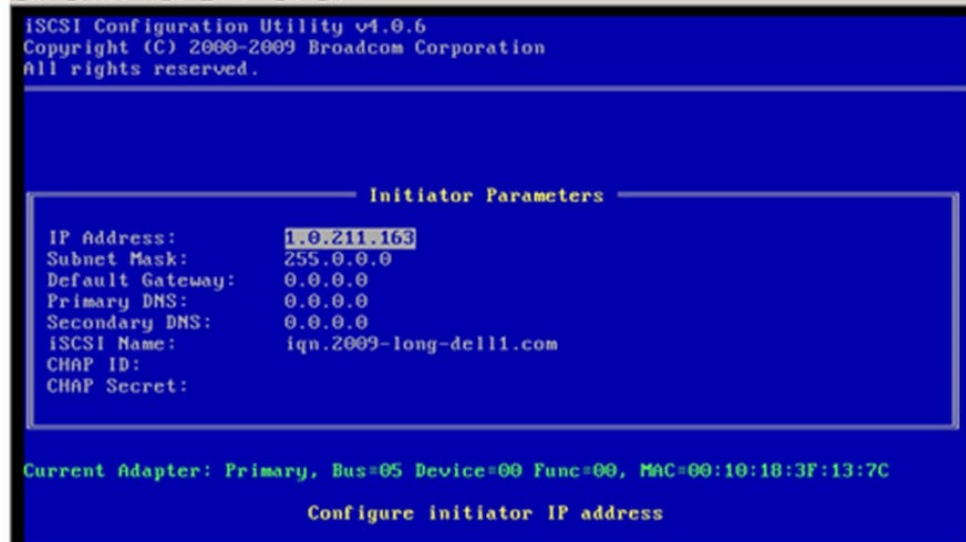
7. Set up the **1st Target Parameters**, as shown in the following figure:
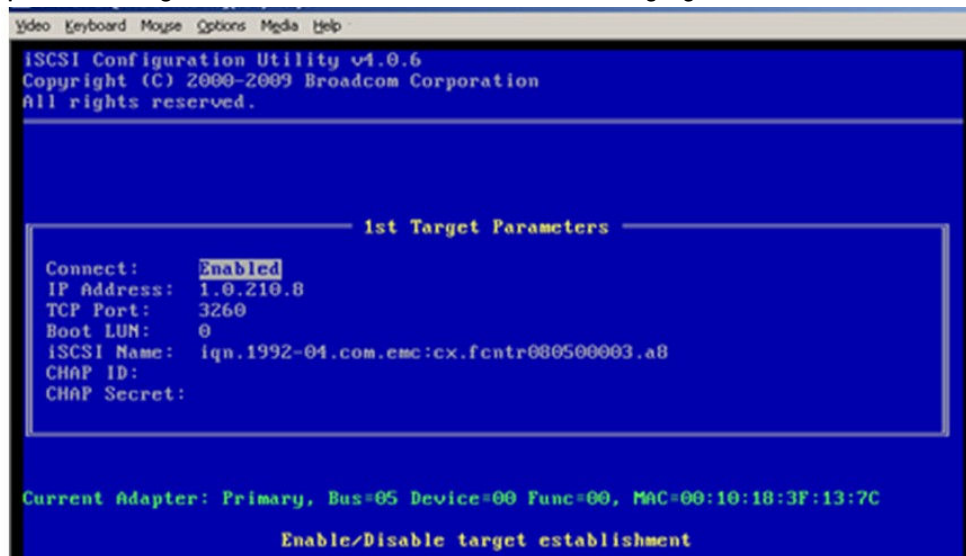
8. Set up the **Secondary Device Parameters**, as shown in the following figure:


```
iSCSI Configuration Utility v4.0.6
Copyright (C) 2000-2009 Broadcom Corporation
All rights reserved.



                          ┌──── Secondary Device Parameters ────┐
                          │ Secondary Device:              00:10:18:3F:13:7E │
                          │ Use Independent Target Portal: Enabled           │
                          │ Use Independent Target Name:   Enabled           │
                          │ Configure Secondary:           Invoke            │
                          └──────────────────────────────────────┘



Current Adapter: Primary, Bus=05 Device=00 Func=00, MAC=00:10:18:3F:13:7C

                        Select secondary device
```
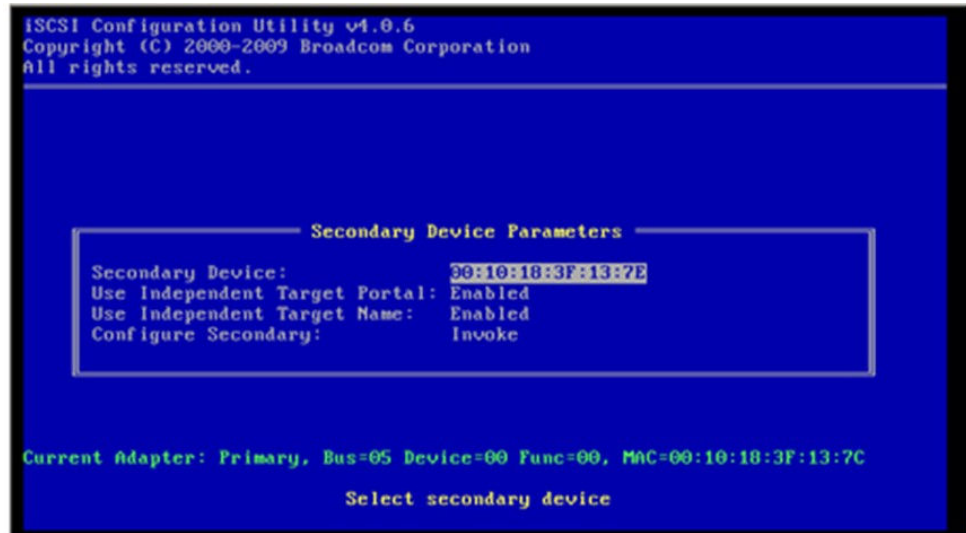
The following steps provide instructions to set up the secondary device, which are same as for the 1st target.

9. For the secondary device, select **General Parameters**, as shown in figure in step 4.

10. Change the **Boot to iSCSI target parameter** to `Disabled`, as shown in figure in step 5.

11. Set up the **Initiator Parameters**, as shown in figure in step 6.

12. Set up the **2nd Target Parameters**, as shown in figure in step 7.

13. Set up the **Secondary Device Parameters**, as shown in figure in step 8.

14. Save and exit the MBA configuration.

15. Reboot the system and enter BIOS to set the boot option to the iSCSI network adapter.

16. Install the operating system on the iSCSI target.

Software iSCSI SAN boot

Many Linux distributions support software iSCSI BFS. RHEL provides native support for iSCSI booting and installing since RHEL 5.1 or later, NOVEL SUSE start to support it since SUSE10. Oracle Linux and Debian operating systems also support it.

About this task

ⓘ **Note:** The figures that are shown in this section display the setup of RHEL 6.4 with Intel 82599EB 10-gigabit controller.

Installing and configuring Intel card for software iSCSI boot:

Procedure

1. Update the latest iSCSI FLB firmware.

   See Intel documents for details.

2. Optionally, you can install an operating system to a local disk and configure the open-iscsi initiator to verify the iscsi network in advance before doing a boot from SAN (BFS).

3. Configure the initiator and target in read-only memory (ROM) in one of the following ways:

   • Intel BIOS to configure (recommended)

   • Boot Intel Utility Tools

4. Create booting LUN and storage group at Array side to make LUN visible to host.

5. Install the operating system.

   Anaconda can discover and then log in to iSCSI disks in the following ways:

   • When Anaconda starts, it checks whether the BIOS or add-on boot ROMs of the system support iSCSI Boot Firmware Table (iBFT), which is a BIOS extension for systems which can boot from iSCSI. If the BIOS supports iBFT, Anaconda will read the iSCSI target information for the configured boot disk from the BIOS and log in to this target, making it available as an installation target.

   • If you select the **Specialized Storage Devices** option during installation, click **Add Advanced Target** to add iSCSI target information like the discovery IP address. Anaconda probes the specified IP address and logs in to any targets that it finds. While Anaconda uses `iscsiadm` to find and log into iSCSI targets, `iscsiadm` automatically stores any information about these targets in the `iscsiadm` iSCSI database. Anaconda then copies this database to the installed system and marks any iSCSI targets that are not used so that the system will automatically log in to them when it starts. If **/** is placed on an iSCSI target, `initrd` will log into this target and Anaconda does not include this target in startup scripts to avoid multiple attempts to log into the same target.

6. Boot the host from the CD-ROM and then select **Specialized Storage Devices** as the type of boot devices for the installation.

   The **Specialized Storage Devices** option installs or upgrades to enterprise devices such as SANs, and enables you to add FCoE, iSCSI, and zFCP disks, and to filter out devices that the installer should ignore.

7. Select a booting LUN.

   Please select the drives you'd like to install the operating system on, as well as any drives you'd like to automatically mount to your system, below:

   | Basic Devices | Firmware RAID | Multipath Devices | Other SAN Devices | Search |

   Filter By: [   ] ∨ Show Only Devices Using: [   ]

☐	Identifier	Capacity (MB)
☐	ip-11.10.10.20:3260-iscsi-iqn.1992-04.com.emc:5000097208243158-lun-0	20625
☐	ip-11.10.10.20:3260-iscsi-iqn.1992-04.com.emc:5000097208243158-lun-1	20625

   After you finish the installation and before you reboot the host, you can open the **GRUB** menu and check the **IBFT** boot firmware menu.

8. Press **Ctrl+Alt+F2** to switch to the **tty2** window.

   • Kernel boot options for IBFT appear after installation in `/boot/grub/menu.lst`:
   `iscsi_firmware ip=ibft ifname=eth0:00:15:17:c8:c3:da`

   • BIOS Enhanced Disk Drive Services (EDD) may stop the booting process because it does not support the ISCSI LUN. If this occurs, you may need to disable the process in the grub boot option:
   `iscsi_firmware ip=ibft ifname=eth0:00:15:17:c8:c3:da edd=off`

   • Make sure that the Intel network adapter IBFT firmware menu appears during the operating system boot.

9. Boot the operating system from the iSCSI LUN.

   (i) Note: During the booting process, if you can see `iscsistart` messages about initiator logging into a target, as shown in the following figure, this confirms that your operating system boot from the iSCSI LUN was successful.

```
iscsistart: Logging into iqn.1992-04.com.emc:5000097208243158 11.10.10.20:3260,1
iscsistart: version 6.2.0-873.2.el6
iscsistart: Connection1:0 to [target: iqn.1992-04.com.emc:5000097208243158, portal: 11.10.10.20,3260] through [iface: default]
s operational now
iscsistart: Logging into iqn.1992-04.com.emc:500009720824315c 12.10.10.20:3260,1
iscsistart: Connection2:0 to [target: iqn.1992-04.com.emc:500009720824315c, portal: 12.10.10.20,3260] through [iface: default]
s operational now
RTNETLINK answers: File exists
iscsistart: version 6.2.0-873.2.el6
iscsistart: Logging into iqn.1992-04.com.emc:5000097208243158 11.10.10.20:3260,1
iscsistart: initiator reported error (15 - session exists)
iscsistart: Logging into iqn.1992-04.com.emc:500009720824315c 12.10.10.20:3260,1
iscsistart: initiator reported error (15 - session exists)
               Welcome to Red Hat Enterprise Linux Server
Starting udev:
```

Configuring SAN boot for FCoE attached host

Installing and configuring Intel card for software FCoE boot

Procedure

1. To configure an Intel Ethernet FCoE boot, power-on or reset the system and press the **Ctrl +D** key combination when the message `Press Ctrl+D to run setup...` is displayed. The Intel Ethernet **FCoE Port Selection** setup menu will then open.

   The first screen of the Intel Ethernet **FCoE Port Selection** setup menu displays a list of Intel FCoE Boot-capable adapters. For each adapter port, the associated SAN MAC address, PCI device ID, PCI bus/device/function location, and a field for the FCoE Boot status is displayed, as shown in the following figure:

   ```
   ════════════════════FCoE Port Selection════════════════════


     MAC:001B21572006 Dev:10FB Loc:6:0:0 BOOT:DISABLED

     MAC:001B21572007 Dev:10FB Loc:6:0:1 BOOT:DISABLED

     MAC:001B213CA096 Dev:10FB Loc:7:0:0 BOOT:FCOE

     MAC:001B213CA097 Dev:10FB Loc:7:0:1 BOOT:FCOE
   ```

2. Select the appropriate port and press **Enter**.

   (i) Note: Up to 10 FCoE Boot-capable ports can be displayed within the **Port Selection** menu. If there are more Intel FCoE Boot-capable adapters, these are not listed in the setup menu.

3. After selecting a port, go to the **FCoE Boot Targets Configuration** page. **Discover Targets** is selected by default. Change the **Discover VLAN** value as required.

4. With **Discover Targets** selected, press **Enter** to show targets associated with the **Discover VLAN** value. Under the **Target WWPN** list, if you know the desired WWPN, you can manually enter it or press **Enter** to display a list of previously discovered targets.

```
┌────────────FCoE Boot Targets Configuration────────────┐
│                                                        │
│   [Discover Targets]      Discover VLAN: 0___          │
│                                                        │
│                                                 Boot   │
│          Target WWPN            LUN    VLAN     Order   │
│                                                        │
│     00:00:00:00:00:00:00:00     0__    0___     0__    │
│                                                        │
│     00:00:00:00:00:00:00:00     0__    0___     0__    │
│                                                        │
│     00:00:00:00:00:00:00:00     0__    0___     0__    │
│                                                        │
│     00:00:00:00:00:00:00:00     0__    0___     0__    │
│                                                        │
│                                                        │
│          Press ENTER to discover targets               │
│                                                        │
│          [Save]                    [Cancel]            │
│                                                        │
└────────────────────────────────────────────────────────┘
```

5. Click **Save**.

```
┌──────────────FCoE Boot Targets Configuration──────────────┐
│                                                            │
│     [Discover Targets]     Discover VLAN: 0___             │
│   ┌───────────────Discover Targets───────────────┐        │
│   │ Please wait while targets are discovered...   │        │
│   │ Logged into fabric on VLAN 171 using priority 3│       │
│   │ Initiator WWPN: 20:00:00:1B:21:3C:A0:96  FID: B00001│   │
│   │ Successfully registered with name server.     │        │
│   │ Retrieved list of targets from name server.   │        │
│   │ Discovered 2 target names.                     │        │
│   │ Select targets using ENTER on the Target WWPN boxes.│   │
│   │ Press any key to continue.                     │        │
│   └────────────────────────────────────────────────┘       │
│                                                            │
│          [Save]                    [Cancel]               │
│                                                            │
└────────────────────────────────────────────────────────────┘
```

6. Select the proper device for the operating system deployment by highlighting a target from the list, and then press **Enter**.

```
═══════════════Select Target from List═══════════════
 WWPN                          Fabric ID

 20:16:00:A0:B8:42:13:8C  -  C20100
 50:00:CC:A0:04:41:6B:2E  -  C205D6




                         1 ⁄ 2

```
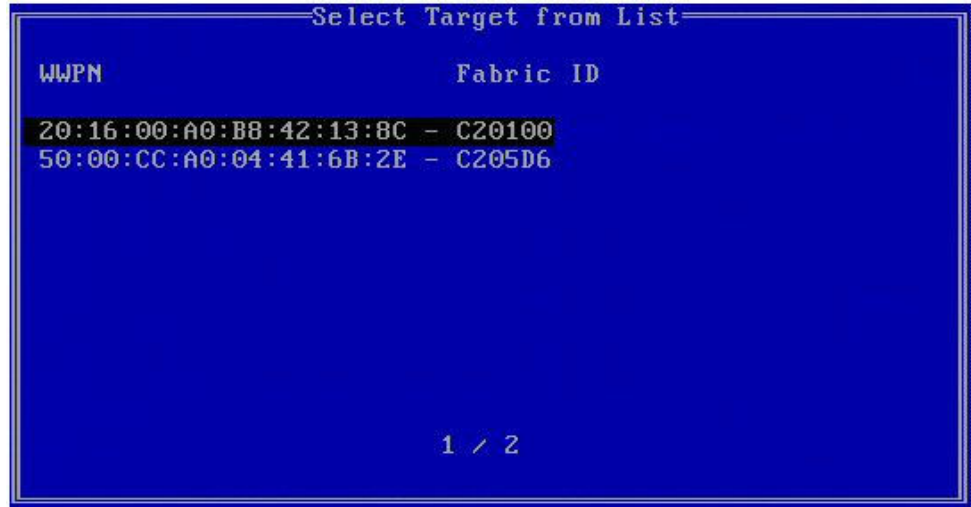
7. Manually enter the LUN and Boot Order values, as shown in the following figure:

```
══════════FCoE Boot Targets Configuration══════════

      [Discover Targets]     Discover VLAN: 171_

                                              Boot
          Target WWPN          LUN   VLAN     Order

      20:16:00:A0:B8:42:13:8C   0__   171_     1__

      50:00:CC:A0:04:41:6B:2E   1__   171_     2__

      00:00:00:00:00:00:00:00   0__   0___     0__

      00:00:00:00:00:00:00:00   0__   0___     0__




            [Save]                  [Cancel]

```
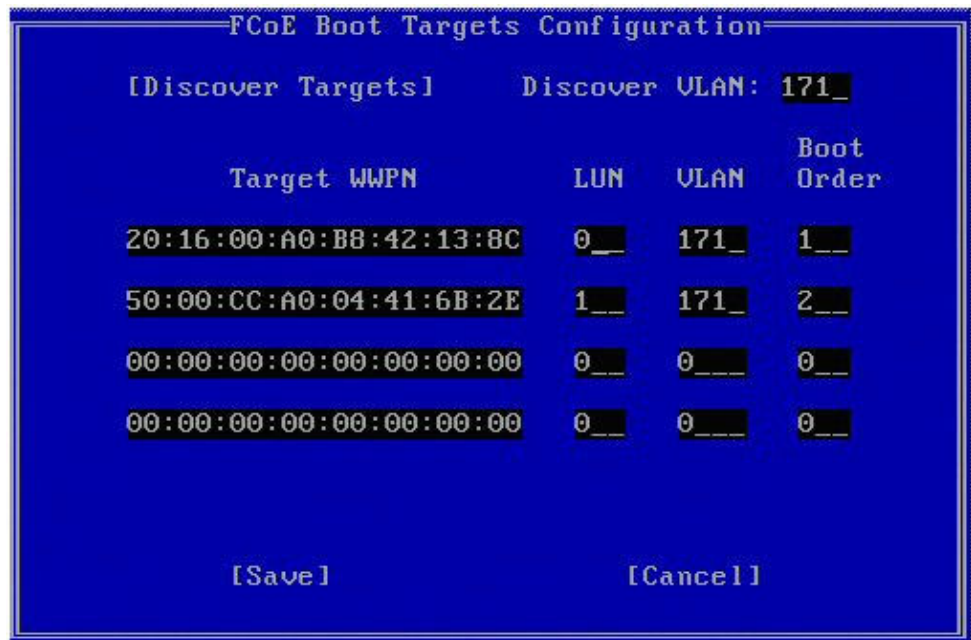
Boot Order valid values are 0-4, where 0 means no boot order or ignore the target. A 0 value also indicates that this port should not be used to connect to the target. Boot order values of 1-4 can only be assigned once to target(s)across all FCoE boot-enabled ports. The VLAN value is 0 by default. You may select **Discover Targets**, which will display a VLAN. If the VLAN displayed is not the one you want, enter the VLAN manually and then select **Discover Targets** on that VLAN.

After rebooting the server, the remote device is displayed in the BIOS, as shown in the following figure:

Installing an operating system on FCoE external devices

Installation of Linux operating system on an external boot device and normal installation on internal hard disks that are attached to the server has one main difference: The partitions where the boot image and operating system packages are installed are hosted on an external device. Complete the following task to install an operating system on FCoE external devices:

Procedure

1. Select the **Specialized Storage Devices** in installation device type prompt.

2. In the **Other SAN Devices** or **Multipath Devices** tab, select the remote device.
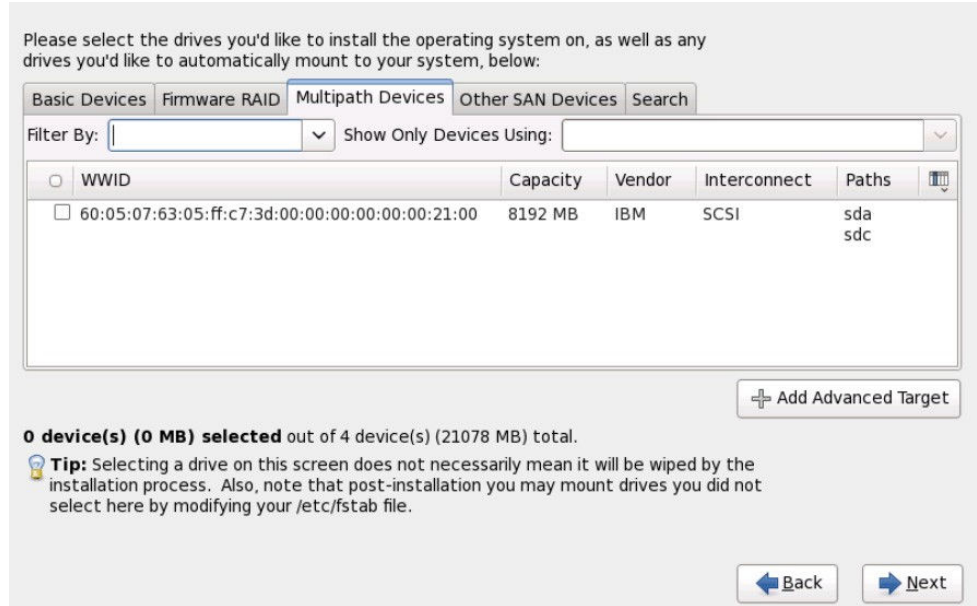
3. With remote device selected on the **Other SAN Devices** tab click **Add Advanced Target**.
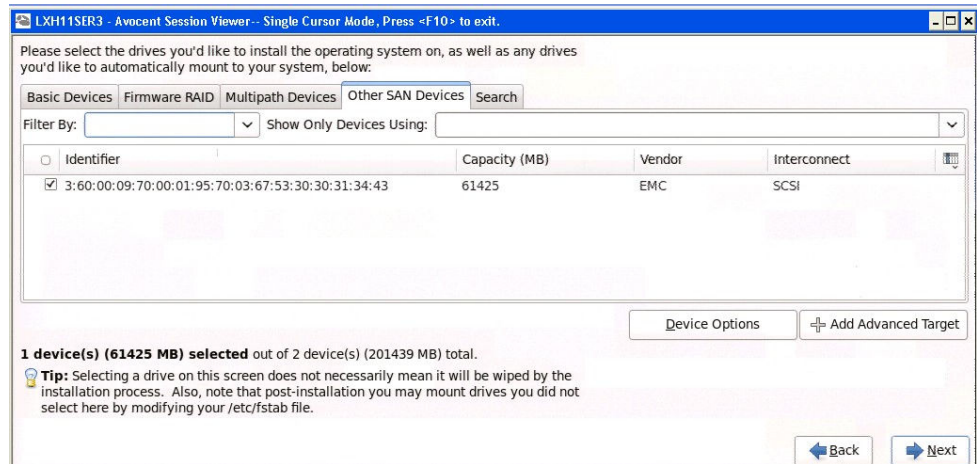
4. In the **Advanced Storage Options** dialog box, select **Add FCoE SAN**, and then click **add drive**.
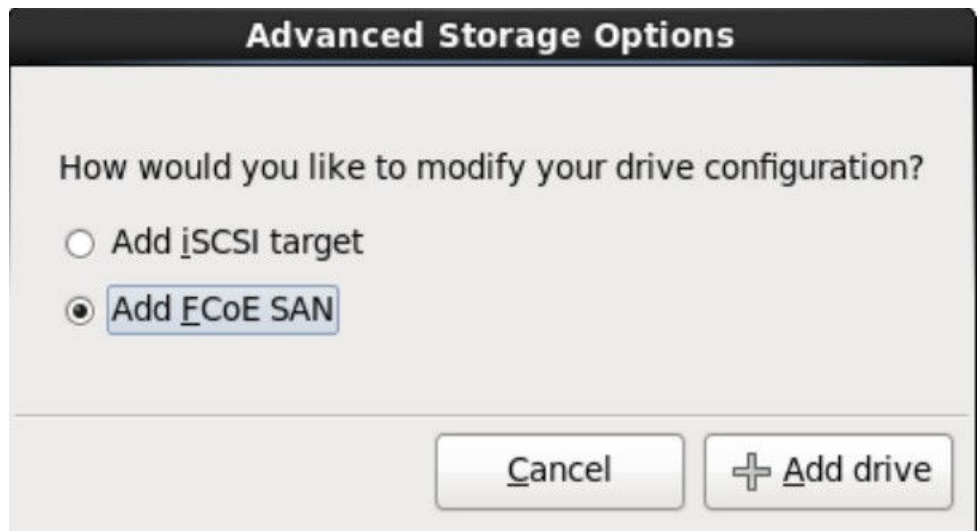
5. In the **Configure FCoE Parameters** dialog box, select the network interface that is connected to the FCOE switch, and then click **Add FCoE Disk(s)**.
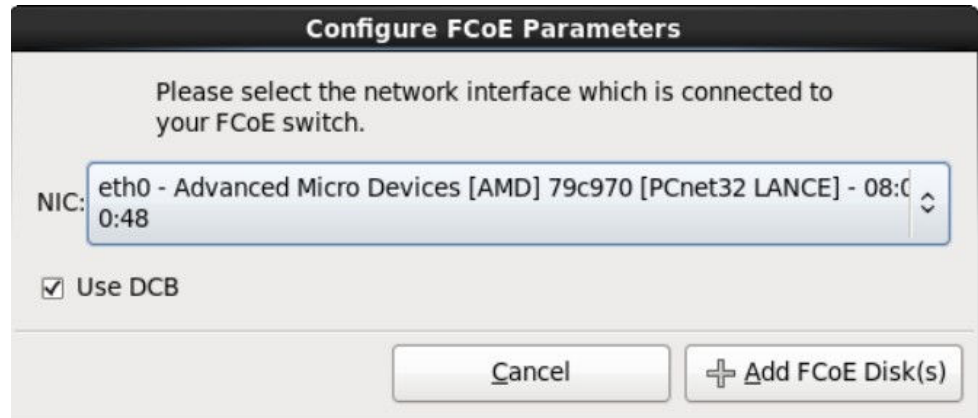
6. After adding the FCoE disks, you can complete the installation. After rebooting, boot from the related remote drive, as shown in the following figure:
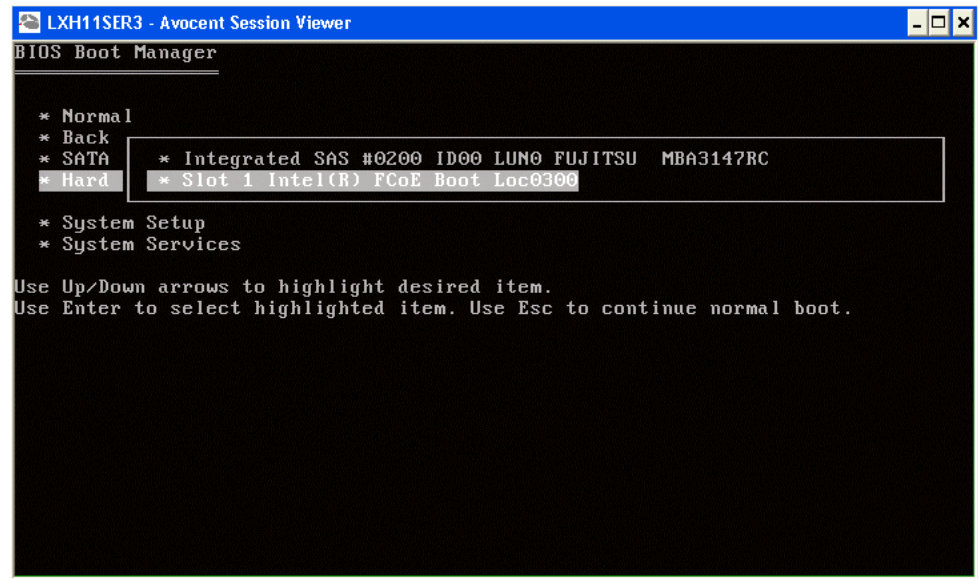
Multipath booting from SAN

> (i) **Note:** Since the Linux operating system will be installed on external boot device, any SCSI hard disks that are attached to the host should be disconnected during and after installation.

Overview

Multipath booting from a SAN is currently available on both PowerPath solutions and Native DM-MPIO.

For PowerPath-This feature was introduced with the release of PowerPath v4.5.0 on Linux and is now a standard offering of PowerPath on Linux.

For Oracle Linux with UEK kernel, PowerPath only supports booting from a SAN on a specific operating system version from PP 5.7.x. Booting from a SAN is not supported on Oracle VM Server 3.x.x. See the PowerPath release notes for more details.

For Native DM-MPIO-This feature was introduced and supported by Dell EMC with the release of RHEL 5.1 and SLES 10 SP2 with Symmetrix and VNX series storage in PNR mode only.

Support for ALUA devices in the VNX series, VNXe series or Unity series storage was introduced and supported by Dell EMC with the release of RHEL 5.8, RHEL 6.3, and SLES 11 SP2.

When using PowerPath to boot from a SAN environment, Dell EMC requires the following:

- Use of PowerPath pseudo (emcpower) device instead of the native sd device when mounting the /boot file system to ensure persistent device naming.

- Use of Logical Volumes LVM2 for creation of all other required partitions.

For instructions about how to configure boot devices with PowerPath, see Dell EMC PowerPath for Linux Installation and Administration Guide.

When using DM-MPIO to boot from a SAN environment, the detail configuration steps may vary depending on Linux distributions and operating system versions. See the operation system guide for the latest information about how to configure DM-MPIO for SAN boot disk.

In General, there are two options to configure DM-MPIO for a SAN boot disk:

- Enabling DM-MPIO at the operating system installation time - To configure the SAN boot LUN with all paths available during installation time, start DM-MPIO and ensure that the boot LUN is detected and managed by DM-MPIO, and then do the Linux installation on the MPIO device.

- Enabling DM-MPIO after operating system installation - To install the operating system with only a single path SAN device (for example /dev/sda), and after the host is booted on this single path device, start DM-MPIO. Note the new DM-MPIO device name for the boot LUN (for example `/dev/mapper/mpatha`) and do the following:

  - Edit the `/etc/fstab` file.

  - Edit the `kernel/grub` file, which depends on different operating system distribution to boot with the new DM-MPIO device name.

  - Remake the initial `ramdisk` image.

  - Reboot the host to boot from the DM-MPIO SAN device.

(i) Note: Installation differs in the following ways for Linux operating system on an external boot device with normal installation on internal hard disks that are attached to the server:

- Partitions where boot image and operating system packages are installed are hosted on an external device.

- The external SAN device with multiple paths must be managed by PowerPath or Linux native DM-MPIO to ensure path failover on path failure.

Configuring DM-MPIO for SAN boot devices

When using DM-MPIO to boot Red Hat Linux from a SAN environment, select from the following options:

(i) Note: The following instructions are for RHEL 6, and the commands are different for other RHEL versions. For more details, see the *Storage/DM Multipath* chapter in the *Red Hat Administration Guide* available on the Redhat website

- To enable DM-MPIO during installation, use the **linux mpath** option for a RHEL installation, which differs from a normal graphical or text mode installation on a local hard disk. Select a multipath device instead of `/dev/sd*` as the target for the operating system to install on. The following figure shows an example of a RHEL installation that uses the **linux mpath** option.

Figure 47 Using the linux mpath option for a RHEL installation
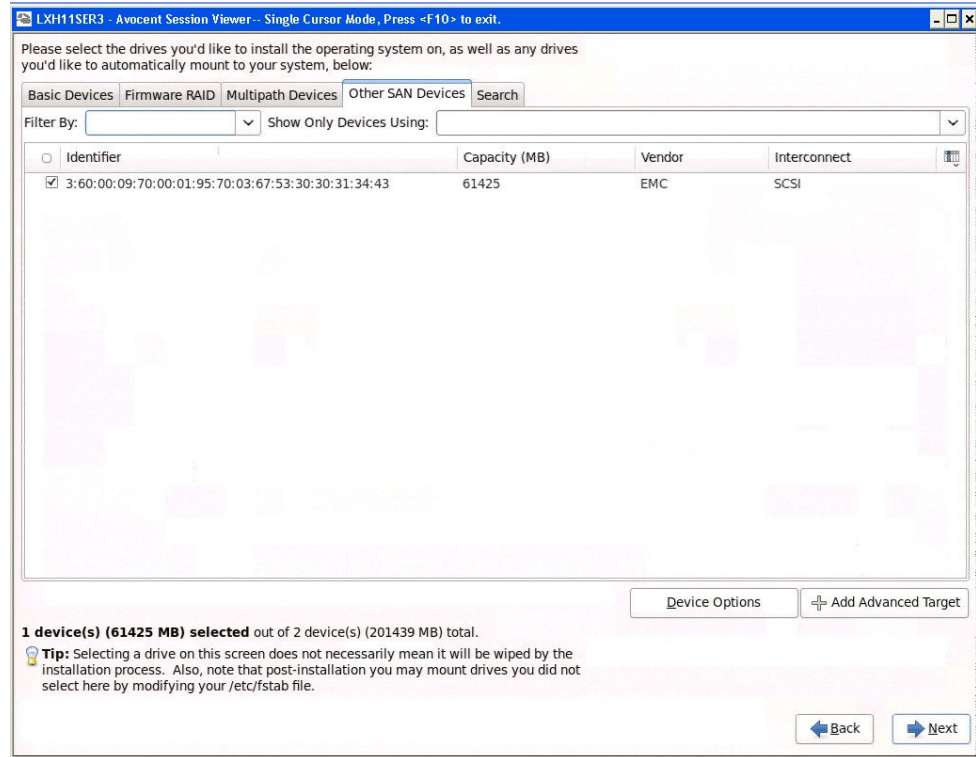
- To enable DM-MPIO after Red Hat Linux installation, install Red Hat with single path device. After you boot Red Hat on a single path device (for example `/dev/sdb`), create `/etc/multipath.conf` file, load multipath, and find the multipath device name for the root LUN.

  - If your root device is not on an LVM volume, and it is mounted by device name, you might need to edit the `fstab` file. For more details, see the *Red Hat Administration Guide* on the Redhat website.

  - After you edit the file, run `'dracut --force --add multipath -include /dev/multipath /etc/multipath` to rebuild the `initramfs` file system, and then reboot the host.

- When using DM-MPIO to boot SLES from a SAN environment:

  (i) **Note:** The following instructions apply to SLES 11. The commands are different for SLES 12 and later. For more details, see *SUSE SLES Storage Administration Guide* on the SUSE website.

  - To enable DM-MPIO during installation, on the **YaST Installation Settings** page, click **Partitioning** to open the **YaST Partitioner**, select **Customer Partitioning**, Select the **Configure Multipath** to start multipath. Ensure that YaST starts to rescan the disks and show boot LUN as multipath device (such as `/dev/disk/by-id/dm-uuid-mpath-3600a0b80000f45893ds`), and use this device to continue installation

  - To enable DM-MPIO after SUSE installation, install SUSE with only a single path device, and ensure that the mount by-id option is checked for the boot partitions by id during installation. After installation, install and start DM-MPIO, add **dm-multipath** to `/etc/sysconfig/kernel:INITRD_MUDULES`. Then re-run `/sbin.mkinitrd` to update the `initrd` image and reboot the server.

Best practices for ALUA mode

This section contains ALUA boot from SAN configuration on:

- Configuring SLES 11 SP2
- Configuring RHEL 5.8
- Configuring RHEL 6.3
- Configuring RHEL 7.0

(i) Note: You can also use instructions in these sections when setting up the boot from SAN configuration on a similar operating system. For example, for RHEL 6.8, see the Configuring RHEL 6.3 section, and for RHEL 7.4, see the Configuring RHEL 7.0 section.

Configuring SLES 11 SP2

Perform the following steps to configuring SLES 11 SP2:

Procedure

1. Install SLES 11 SP2 with a single path to the array and DM-MPIO enabled on the boot LUN.

2. When installation completes, reboot and connect the remaining paths. Run the `multipath -ll` command. The default hwhandler for VNX, VNXe, and Unity are *1 emc*.

3. Copy a default multipath.conf file.

   ```
   #cp /usr/share/doc/packages/multipath-tools/
   multipath.conf.synthetic /etc/multipath.conf
   ```

4. Add the following to `/etc/multipath.conf` file.

   ```
   devices {
                   # Device attributed for EMC VNX series ALUA
                   device {
                           vendor "DGC" product "*"
                           hardware_handler "1 alua"
                   }
   }
   ```

5. Use the `initrd` command to rebuild the `ramdisk` image and reboot:

   ```
   # mkinitrd
   ```
   ```
   # reboot
   ```

6. Check the result of `multipath -ll` command again. Handler should now be `1 alua`.

Configuring Red Hat Enterprise Linux 5.8

Procedure

1. Install Red Hat Enterprise Linux 5.8 with DM-MPIO enabled on the boot LUN.

2. Copy a default `multipath.conf` file to overwrite the existing one.

   ```
   # cp /usr/share/doc/device-mapper-multipath-0.4.7/
   multipath.conf.synthetic /etc/multipath.conf
   ```

3. Edit the `/etc/multipath.conf` file using the example that follows:

   ```
   defaults {
           user_friendly_names     yes
   ```

```
}
blacklist {
        devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*"
        devnode "^hd[a-z][[0-9]*]"
        devnode "^cciss!c[0-9]d[0-9]*"

}
devices {
     device {
                                                          vendor
"DGC"
                                                          product
"*"

prio_callout    "/sbin/mpath_prio_alua /dev/%n"

path_grouping_policy    group_by_prio
                                                          features
"1 queue_if_no_path"
                                                          failback
immediate

hardware_handler    "1 alua"
                                        }
}
```

4. Use the `initrd` command to rebuild the initrd ramdisk image. Reboot the host.

```
#cp /boot/initrd-2.6.18-308.el5.img /boot/
initrd-2.6.18-308.el5.img.bak

# mkinitrd -f --with=scsi_dh_alua /boot/initrd-2.6.18-308.el5.img
2.6.18-308.el5

# reboot
```

Configuring Red Hat Enterprise Linux 6.3

Procedure

1. Install Red Hat Enterprise Linux 6.3 with DM-MPIO enabled on the boot LUN.

2. Copy a default `multipath.conf` file to overwrite the existing one.

   ```
   # cp /usr/share/doc/device-mapper-multipath0.4.9/
   multipath.conf.synthetic /etc/multipath.conf
   ```

3. Edit the `/etc/multipath.conf` file using the example that follows:

```
defaults  {
        user_friendly_names       yes
}
blacklist {
        wwid 26353900f02796769
          devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*"
          devnode "^hd[a-z][[0-9]*]"
          devnode "^cciss!c[0-9]d[0-9]*"
}
devices  {
        device {
                  vendor                    "DGC"
                  product                   "*"
                  prio                      tpg_pref
                  path_grouping_policy      group_by_prio
                  features                  "1 queue_if_no_path"
                  failback                  immediate
```

```
                                 hardware_handler        "1 alua"
              }
}
```

4. Edit the `/boot/grub/menu.lst` file as follows (add the text in bold, exactly as shown):

```
#boot=/dev/mpathb
default=0
timeout=5
splashimage=(hd0,0)/grub/splash.xpm.gz
hiddenmenu
title Red Hat Enterprise Linux (2.6.32-279.el6.x86_64)
root (hd0,0)
kernel /vmlinuz-2.6.32-279.el6.x86_64 ro root=/dev/mapper/vg_lin117107-
lv_root
rd_NO_LUKS LANG=en_US.UTF-8 rd_NO_MD rd_LVM_LV=vg_lin117107/lv_swap
SYSFONT=latarcyrheb-sun16
rd_LVM_LV=vg_lin117107/lv_root crashkernel=128M KEYBOARDTYPE=pc
KEYTABLE=us rd_NO_DM rhgb
quiet rdloaddriver=scsi_dh_alua,scsi_dh_rdac
initrd /initramfs-2.6.32-279.el6.x86_64.img
```

5. Use the `initrd` command to rebuild the *initrd ramdisk* image. Reboot the host:

```
# cp initramfs-2.6.32-279.el6.x86_64.img
initramfs-2.6.32-279.el6.x86_64.img.bak
```

```
# dracut --force initramfs-2.6.32-279.el6.x86_64.img
```

```
# reboot
```

Configuring Red Hat Enterprise Linux 7.0

Procedure

1. Install RHEL 7.0 with DM-MPIO enabled on the boot LUN.

2. Copy a default `multipath.conf` file to overwrite the existing one:

```
# cp /usr/share/doc/device-mapper-multipath0.4.9/
multipath.conf /etc/multipath.conf
```

3. Edit the `/etc/multipath.conf` file using the example that follows:

```
defaults {
          user_friendly_names    yes
}
blacklist {
          wwid 26353900f02796769
          devnode "^(ram|raw|loop|fd|md|dm-|sr|scd|st)[0-9]*"
          devnode "^hd[a-z][[0-9]*]"
          devnode "^cciss!c[0-9]d[0-9]*"
}
devices {
        device {
                vendor                 "DGC"
                product                ".*"
                product_blacklist      "LUNZ"
                path_grouping_policy   group_by_prio
                path_selector          "round-robin 0"
                path_checker           emc_clariion
                features               "1 queue_if_no_path"
                hardware_handler       "1 alua"
```

```
                        prio               alua
                        failback           immediate
                        rr_weight          uniform
                        no_path_retry      60
                        rr_min_io          1
}
```

4. At this point, the procedure is different for RHEL 7 from the other operating system versions as RHEL 7 start to grub2 to manage boot loader. Complete the following steps to continue the configuration:

   a. Build new initram disk to contain `multipath.conf`:

   ```
   # dracut --force /boot/initramfs-alua-3.10.0-123.el7.x86_64.img
   ```

   b. Add a new boot entry with the initramfs file name in `/etc/grub.d/40_custom`:

   ```
   #!/bin/sh

   exec tail -n +3 $0
   ```

   This file provides an easy way to add custom menu entries. Type the menu entries that you want to add after this comment. Do not change the `exec tail` line above.

   ```
   load_video

   insmod gzio

   insmod part_msdos

   insmod xfs

   set root='hd0,msdos1'

   if [ x$feature_platform_search_hint = xy ]; then

   search --no-floppy --fs-uuid --set=root --hint='hd0,msdos1'

   7f898f92-6261-44a6-9d7e-127cc7ec1967

   else

   search --no-floppy --fs-uuid --set=root
   7f898f92-6261-44a6-9d7e-127cc7ec1967fi

   linux16 /vmlinuz-3.10.0-123.el7.x86_64

   root=UUID=8ba76e02-3c10-49da-a699-d44eb913b550 ro selinux=0
   crashkernel=auto

   vconsole.keymap=us rd.lvm.lv=rhel_vmw117176/root
   vconsole.font=latarcyrheb-sun16

   rd.lvm.lv=rhel_vmw117176/swap rhgb quiet LANG=en_US.UTF-8

   rdloaddriver=scsi_dh_alua,scsi_dh_rdac

   initrd16 /initramfs-alua-3.10.0-123.el7.x86_64.img
   ```

   c. Update the new entry in the `grub.cfg` file and make this default entry so that the system boots with the new initrd image automatically:

   ```
   grub2-mkconfig -o /boot/grub2/grub.cfg

   grub2-set-default 'Red Hat Enterprise Linux Server, with Linux
   BFS '
   ```

   d. Reboot the host.

PowerPath booting from SAN

If PowerPath is employed as the multipathing software, the boot partitions must reside on LVM volumes. The coexistence of PowerPath pseudo devices and a native block device may cause duplication problem of LVM2 physical volumes. However, this concern does not apply to configurations with MPIO as multipathing software.

A way to avoid warning messages congesting syslog is to include appropriate filters in the LVM configuration file `/etc/lvm/lvm.conf`.

A boot LUN with PowerPath employed can be replicated. A replication of the boot LUN can use (but is not limited to) one of the following:

- Timefinder/Clone on EMC Symmetrix

- SRDF on Dell EMC Symmetrix

- Snapview/Clone on Dell EMC VNX

- Mirrorview on Dell EMC VNX

- Open Replicator

- SAN Copy

- Vplex Encapsulation of the boot LUN

To replicate the Linux boot LUN with PowerPath installed, see KB article 336485 available on Dell EMC Online Support.

(i) **Note:** See the appropriate *Dell EMC PowerPath for Linux Installation and Administration Guide* for your operating system, available on Dell EMC Online Support. This guide provides the current steps to use when installing PowerPath on your boot device.

Guidelines for booting from Symmetrix, XtremIO, PowerStore, VNX series, VNXe series or Unity series

This section provides booting guidelines for the following:

- Dell EMC Symmetrix-specific guidelines

- VNX series, VNXe series, or Unity series-specific guidelines

- XtremIO-specific guidelines

- PowerStore-specific guidelines

Dell EMC Symmetrix-specific guidelines

- Prior to the installation on a Symmetrix LUN, the Linux host HBA must have successfully logged into the array. Using Solutions Enabler from another host, at least one LUN must be assigned to the host.

- During the installation procedure, it is recommended that only one LUN be allocated to the host for ease of installation. Once the installation has completed, additional LUNs may be assigned to the host.

- When attached to a Symmetrix, the physical-to-logical split must be such that you meet the minimum required disk space that is required to install the Linux operating system. Consult your Linux distribution for these requirements.

- For RHEL 4.5 boot from a LUN with VCM gatekeeper existing on a Symmetrix, you may **receive an** `unhandled exception with ZeroDivisionError` **message when**

partitioning the boot LUN. Remove the VCM gatekeeper from the FA port and the installation will be successful.

- When using the *Basic Storage* method, after choosing any of the installation type options the installation may exit and dis- play the error message `Anaconda died after receiving signal 6` on RHEL 6.1 (and later) with VCM/ACLX set to write enable. This could be the result of the VCM/ACLX device having been labeled as another UNIX device (for example, SUN). Instead of using RHEL *Basic Storage*, select the *Specialized Storage* option and choose your intended boot device. It is recommended that you clear the VCM/ACLX device (1 GB or less) from the lists of devices you would like to automatically mount to your system.

VNX series, VNXe series or Unity series-specific guidelines

- Prior to the installation on a VNX series LUN, the Linux host must have been manually registered on the array and assigned to a Storage Group. At least one LUN must be bound to the host's Storage Group and owned by the SP connected to the HBA being used for the fabric boot. The lowest-numbered path to the boot LUN must be the active path.

- It is recommended that the boot LUN be assigned Host LUN ID 0. During the installation procedure, it is recommended that only one LUN be assigned to the Storage Group for ease of use. Once the installation has completed, additional LUNs may be added to the Storage Group.

- Booting from the SAN requires the use a Unisphere/Navisphere Management station with the Unisphere/Navisphere Manager or Unisphere CLI/Navisphere CLI. The station must be separate from the boot server, but networked to the VNX series, VNXe series or Unity series to manually register the boot server.

XtremIO-specific guidelines

To install the operating system on XtremIO LUN, the Linux host must have been registered on the array. Similar to other array configurations, at least one LUN with enough space for the operating system must be mapped to the host.

During the installation procedure, it is recommended that only one LUN be mapped to the host. Once the installation has completed, additional LUNs may be added the host.

In XtremIO version 4.0.0 or later, volumes are numbered by default starting from the LUN ID 1 (not 0 as was the case in previous XtremIO versions). Although possible, we do not recommend manually adjusting the LUN ID to 0, as it may lead to issues with some operating systems.

PowerStore-specific guidelines

- Prior to the installation on a PowerStore LUN, the Linux host must have been manually registered on the array and assigned to a Storage Group. At least one LUN must be bound to the host's Storage Group and owned by the node connected to the HBA being used for the fabric boot. The lowest-numbered path to the boot LUN must be the active path.

- It is recommended that the boot LUN be assigned Host LUN ID 1. During the installation procedure, it is recommended that only one LUN be assigned to the Storage Group for ease of use. Once the installation has completed, additional LUNs may be added to the Storage Group.

Virtualization

Linux virtualization

The two Linux virtualization offerings are Xen virtualization technology and Kernel Virtual Machine (KVM) virtualization suite.

For all other Linux-based virtualization products not listed in the Dell EMC Simple Support Matrix, contact your local Dell EMC representative.

Benefits

Block-based virtualization products share a common set of features and benefits. These benefits include (but are not limited to) the following:

- Creates an open storage environment capable of easily supporting the introduction of new technologies.

- Decreases the amount of planned and unplanned downtime.

- Increase storage utilization and decrease the amount of *stranded* storage.

- Reduces management complexity through single pane of glass management with simplified policies and procedures for multiple storage arrays.

- Improves application, storage, and network performance (in optimized configurations) through load-balancing and multipathing.

Xen Hypervisor

Xen technology is a customized Linux kernel specializing in providing emulated hardware environments to virtualized guest operating systems.

The Xen kernel can be distributed as a standalone operating system, or embedded into normal enterprise Linux distributions.

Dell EMC only supports a limited set of Xen server implementations. See the Dell EMC Simple Support Matrix for the latest supported configurations.

Virtualization modes

Xen kernel implements virtualization of operating system in two modes:

- Paravirtualization
- Full virtualization

Also included is a brief description of the following:

- Virtual machine-specific drivers
- Cluster or pool environment

Paravirtualization

Paravirtualization transfers the virtualized operating system's burden of interacting with underlying hardware to the Xen kernel. The virtualized operating systems, which are normally referred to as Guests, are partially modified by the Xen kernel, and place most of the hardware handling functionalities on to the Xen kernel. By so doing, the underlying hardware does not need to provide additional support for virtualization. They interact with the Xen kernel in the usual way as with a native Linux kernel, and the Xen kernel will direct the I/O and other information from the hardware to the dedicated guest operating system. As a result of this configuration, the Xen kernel provides modified versions of hardware drivers and block device interfaces to the guest operating systems for access to storage and network equipment.

Full virtualization

In contrast to the reliance on a special hypervisor kernel for special drivers to handle the virtualized hardware, *full virtualization* retains the original code implementation of the guest operating system, and facilitates direct interaction between the guest operating system and underlying hardware. Fully virtualized operating systems use the same hardware drivers and block devices as if they are installed on bare metal hardware.

Special server hardware technologies are essential to support fully virtualized guest operating systems in Xen, such as Intel's VT capable processors or AMD's Pacifica processors. Without these processor technologies, Xen only supports paravirtualized guest operating systems. Full virtualization is also called Hardware Virtual Machine (HVM) in Xen terminology.

Virtual Machine-specific drivers

Both Red Hat and SuSE provide virtual machine-specific drivers instead of operating-system specific drivers to help the guest operating systems interact with hardware. Consult your Linux distributor for details.

Cluster or pool environment

A few Xen technology providers support cluster or pool environment to enable high availability and live migration functionalities of Guest operating systems. But the implementation of a Xen cluster environment is subject to the constraint of a homogeneous hardware environment; meaning that the clustered Xen servers are required to have identical hardware configurations, especially for CPUs.

Virtual machine installation and management

Virtual Machine (VM) can be installed and monitored using both graphic mode and command lines.

Both SuSE and Red Hat offer user interface, GUI, for installation and administration of VMs.

(i) **Note:** Virtualization-related packages need to be manually selected and installed, either during host operating system installation or through package online update option after the host operating system is installed.

Xen kernel hypervisor and virtualization management GUI need to be installed separately using different packages. Ensure that all necessary packages are installed and loaded and the Xen kernel properly boots before installing virtual machines.

Since Xen is a Linux kernel by itself, the server must boot to Xen kernel to access the Xen hypervisor. This can be configured by editing the booting sequence in /boot/grub/menu.lst file.

The virtual machine manager GUI provides an installation wizard to guide the user through the installation of virtual machines, as shown in the figures in this section. The user can configure the operation system type, mode of virtualization, and the parameters for virtualized hardware configuration, such as installation method, installation media, storage, and network.

The following figures provide the instructions for VM installation:

Figure 48 Virtual machine installation wizard interface steps 1 - 3
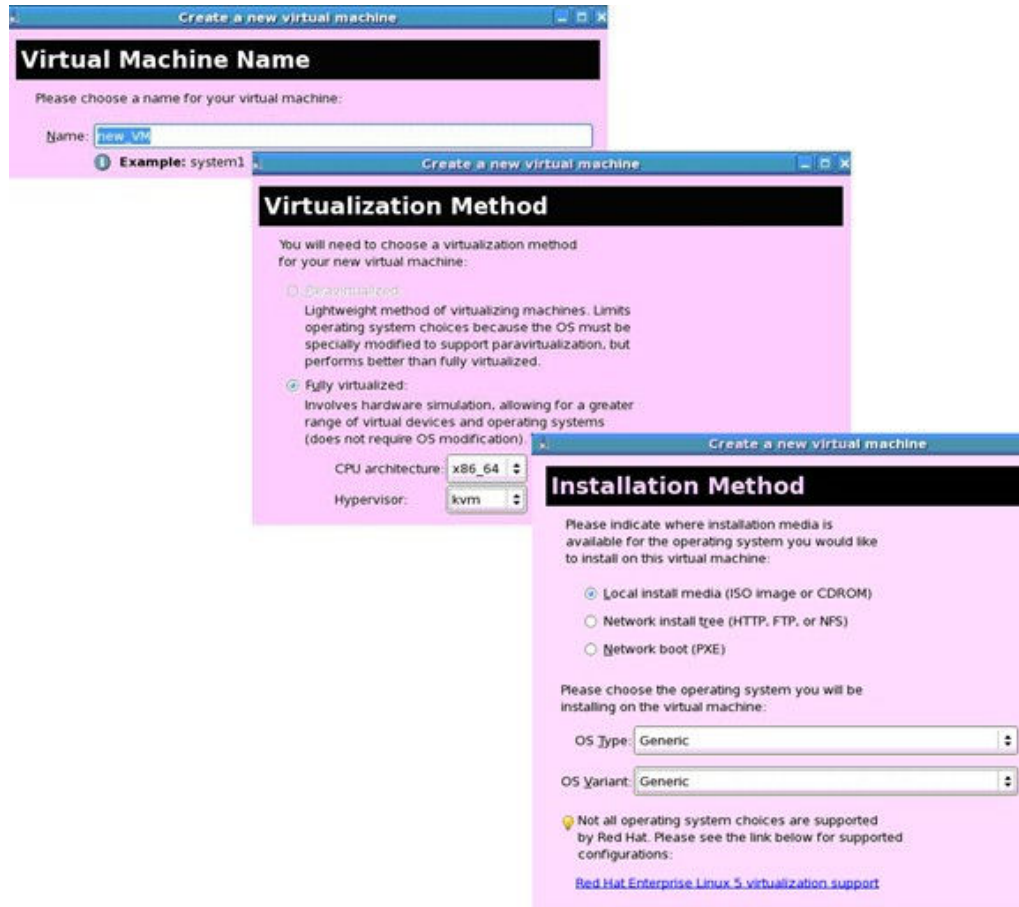
Figure 49 Virtual machine installation wizard interface steps 4 - 6

Figure 50 Virtual machine installation wizard interface step 7
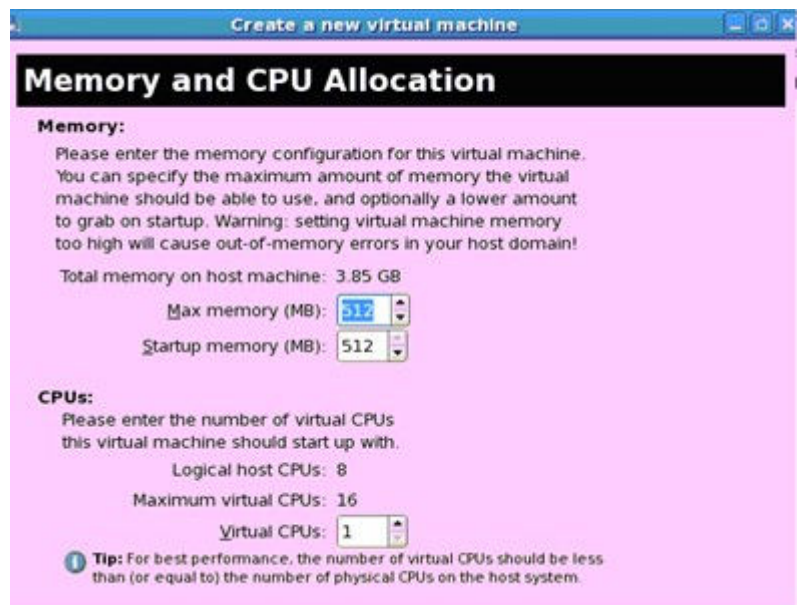
After all parameters are properly indicated, a summary screen will display showing the selected parameters.

Figure 51 Summary screen showing virtual machine installation parameters

When the parameters are confirmed, the installation process can be started.

A **VNC Viewer** window connecting to the virtual machine's console displays and the installation can be continued in the same way as installing a physical machine

Figure 52 VNC Viewer window connecting to the virtual machine's console
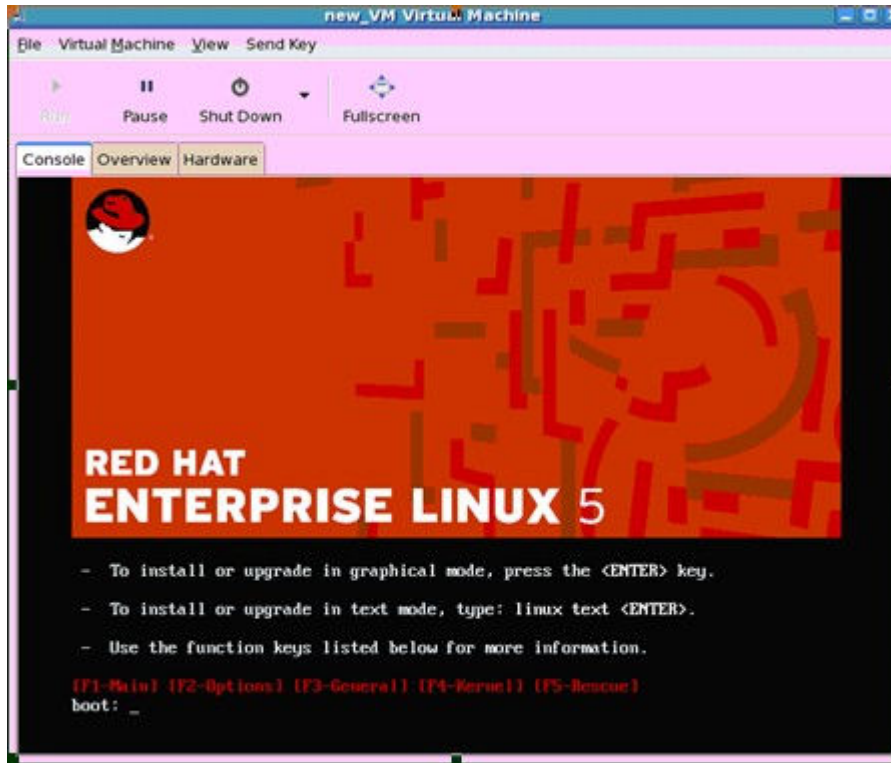
For more information about the virtualization package installation and configuration, and virtual machine installation, see:

- SuSE on the SuSE website
- Red Hat website

After installation, the VMs will appear as new domains under the control of the Xen kernel. Users can access the graphical console of the VMs directly using the VNC Viewer function provided by the virtual machine manager GUI. The virtual machines can be shut down, started, paused, or destroyed using virtualization management utility, or command lines. Their status and hardware utilization status are also monitored by the management utility. An example is shown in following figure:

Figure 53 Virtual Machine Manager interface example
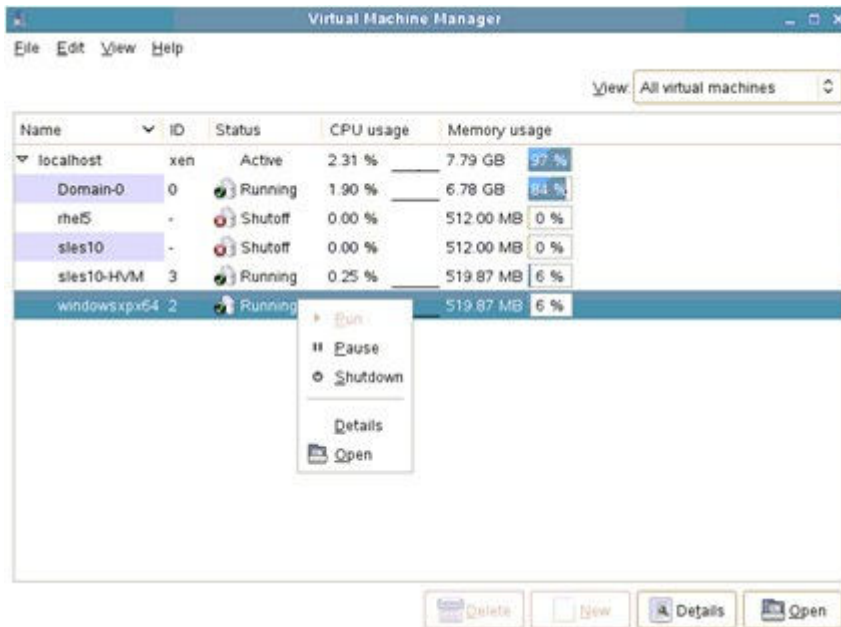
The setting and detailed parameter information is also available graphically. The hardware of the virtual machine can also be removed, and new hardware can be added, using this graphical summary window, as shown in the following figure:

Figure 54 Hardware details window
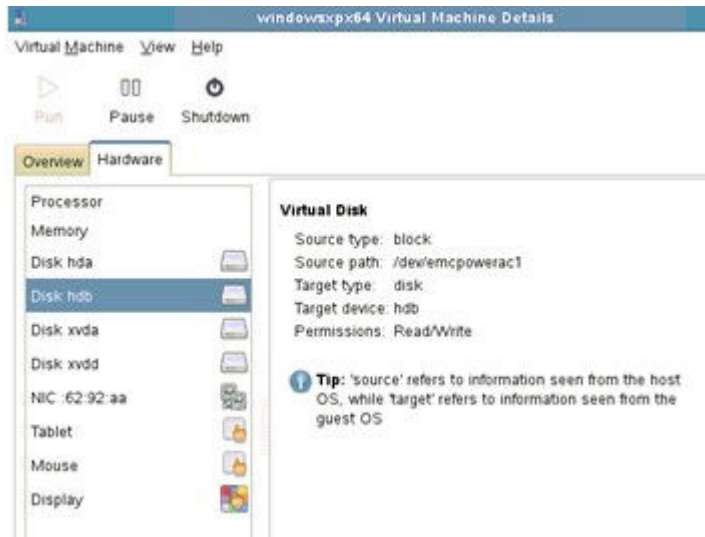
Storage management

The Xen kernel implements virtualized entities to emulate the interface between the guest operating system and hardware components. Storage is one of the key components in hardware emulation. Terminologies that are defined in Xen virtualization for storage management include the following terms.

Domain
Each virtual machine is described as a domain. There are two types of domains, namely Domain 0 (Dom0) and Domain U (DomU).

Domain 0 (Dom0)

Domain 0 is a privileged domain, which is the controlling host environment for all guest operating systems. It is responsible for a few virtualization tasks as instructed by the hypervisor software running directly on the hardware. An especially important management task that is associated with Dom0 is the coordination and distribution of virtual I/O from the physical server to the guest operating system using virtual device drivers.

Domain U (DomU)

Corresponding to the privileged Dom0, there is an unprivileged counterpart, called DomU. The DomU is the virtual machine, guest, that resides on top of the host system. DomU has no access to underlying hardware by default. All the virtual machines created using Xen kernel commands or the virtualization management software interface are considered DomU. The paravirtualized DomUs has no access to the underlying hardware except through the Dom0; whereas, fully virtualized DomUs may have driver components that interact with hardware directly.

Virtual disk
Virtualized storage unit to provide boot function and file storage capacity to the guest operating systems. A virtual disk can be an image file, or a block device, such as a LUN, a partition, a volume, or a physical device like a CD-ROM. So, it is possible to convert a physical operating system installed on a remote device, such as a LUN, to a virtual machine managed by the Xen hypervisor.

An ISO image can also be converted to a virtual disk.

Resource pool
Collection of hardware shared by guest operating systems. Resource pool is an important concept for High Availability and Live Migration functionalities.

Connectivity and path management software

Dell EMC supports storage that is provisioned to hosts running XenServer through the following connectivity:

- Fibre Channel

- iSCSI

- Fibre Channel over Ethernet (FCoE)

XenServer support for multipathing is performed by using the Linux native multipathing (DM-MPIO Multipath). It is the native Linux kernel in the host server that performs the necessary path failover and recovery for the server. Hence, the virtual machines are spared the burden of implementing load balancing and failover utilities. See the Dell EMC Simple Support Matrix Linux operating system specific footnotes in the base connectivity section of Dell EMC E-Lab Navigator for supported configurations.

Currently Fibre Channel, Fibre Channel over Ethernet (FCoE), and iSCSI connectivity are supported as protocols for storage provisioning.

Dell EMC supports both Linux native multipathing (DM-MPIO) and PowerPath as path management software for Xen kernel products.

When the guest operating system is paravirtualized, I/O is not directly handled by the VM, but is redirected by the Xen Dom0. Therefore, multipathing is implemented at Dom0 and path failures are sensed and reported by the Dom0 without generating any abnormalities on the guest operating system.

Kernel-based Virtual Machine

The Kernel-based Virtual Machine (KVM), is a Linux kernel virtualization infrastructure that is implemented as a kernel module, as opposed to a separate integrated kernel, such as Xen.

Introduction

KVM started to be implemented with Linux kernel with kernel version 2.6.20. Similar to Xen, Dell EMC supports the KVM server implementations as provided by SuSE Enterprise Linux and Red Hat Enterprise Linux. Currently, KVM is available on RHEL 5.4 and later, and SLES 11 SP1 and later.

Although KVM is also part of the Linux kernel, unlike Xen it is not a combination of both hypervisor and kernel. Instead, it is a loadable module like most of the other Linux drivers, and is separate from Linux kernel.

With a native Linux kernel running in production, the administrator can install KVM packages and start the service readily without downtime of the host. This one feature has minimized the impact to daily production when implementing Linux virtualization, and enables more applications to run by virtue of a non-modified kernel, which would not be available with a Xen kernel in place.

Prerequisites

You need to install the following packages to run KVM:

- libvirt-0.7.6-1.9.8
- virt-viewer-0.2.0-1.5.65
- libvirt-cim-0.5.8-0.3.69
- virt-manager-0.8.4-1.4.11
- libvirt-python-0.7.6-1.9.8

(i) **Note:** These packages will not be automatically installed during operating system installation unless manually specified during package selection.

SLES 11 and 12

In SLES systems, users are allowed to install additional applications after operating system installation is completed. SLES YaST service serves this purpose. Users can implement KVM technology when the need arises after the operating system is installed, in the same manner as adding other applications to a running SLES system.

Red Hat Enterprise Linux 5.x

In Red Hat Enterprise Linux 5 series systems, users need to provide an installation code that will enable the subscription to KVM service. If the installation code is not specified, KVM-related packages cannot be installed during or after an operating system installation. The following instructions describe the process of installing the KVM package in Red Hat Enterprise Linux 5.x series systems.

1. Enter the installation number.

2. Select the Virtualization package.

3. Select KVM, Virtualization.

Red Hat Enterprise Linux 6, 7 and 8

In the Red Hat Enterprise Linux 6, 7 and 8 distribution, the need of providing an installation code does not exist. Before you can use virtualization, a basic set of virtualization packages must be installed on the target system. The virtualization packages can be installed either during the host installation sequence, or after host installation using the GUI or CLI Linux package management tools.

The following instructions describe the process of installing the KVM package in Red Hat Enterprise Linux 6 series systems:

1. Customize the base system.
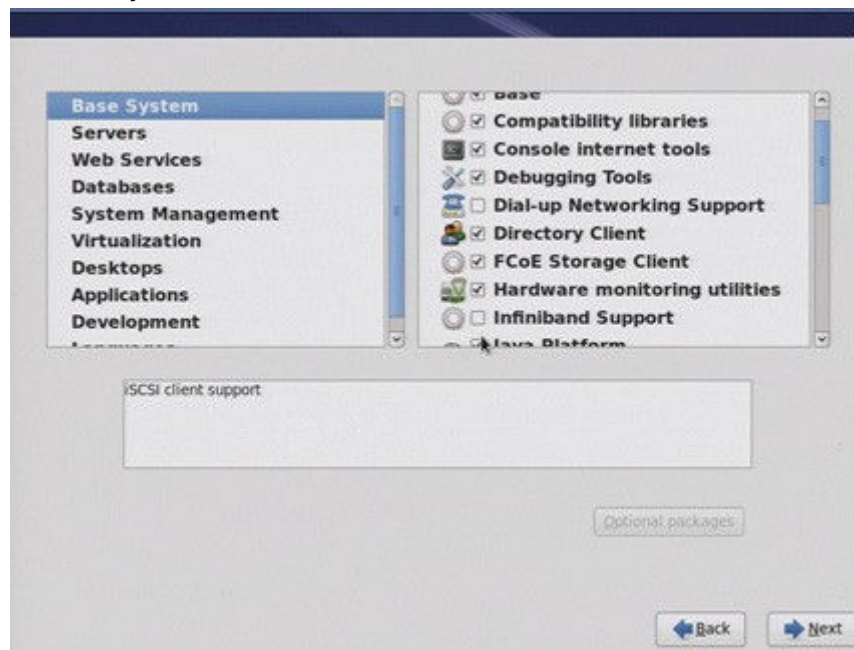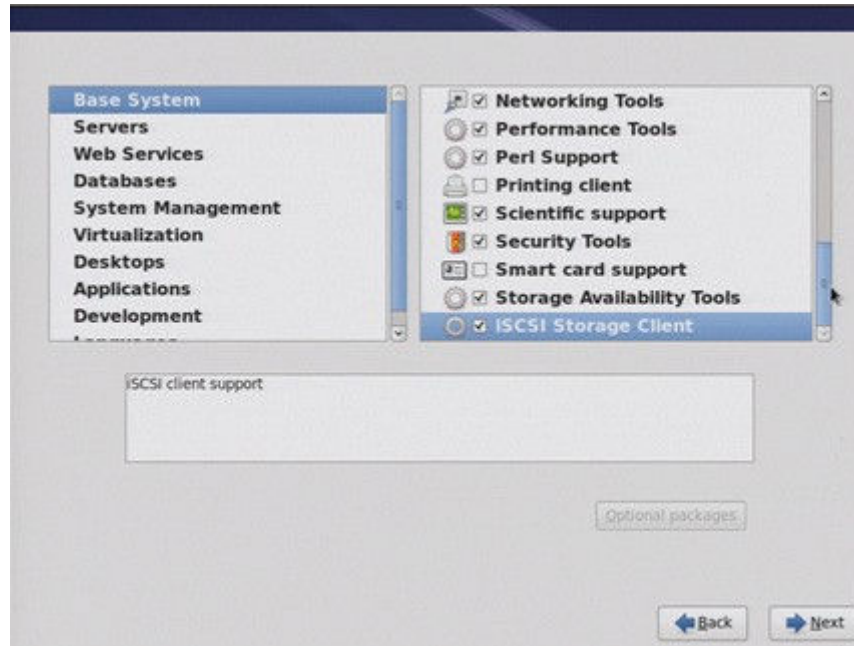
2. Select Virtualization options.
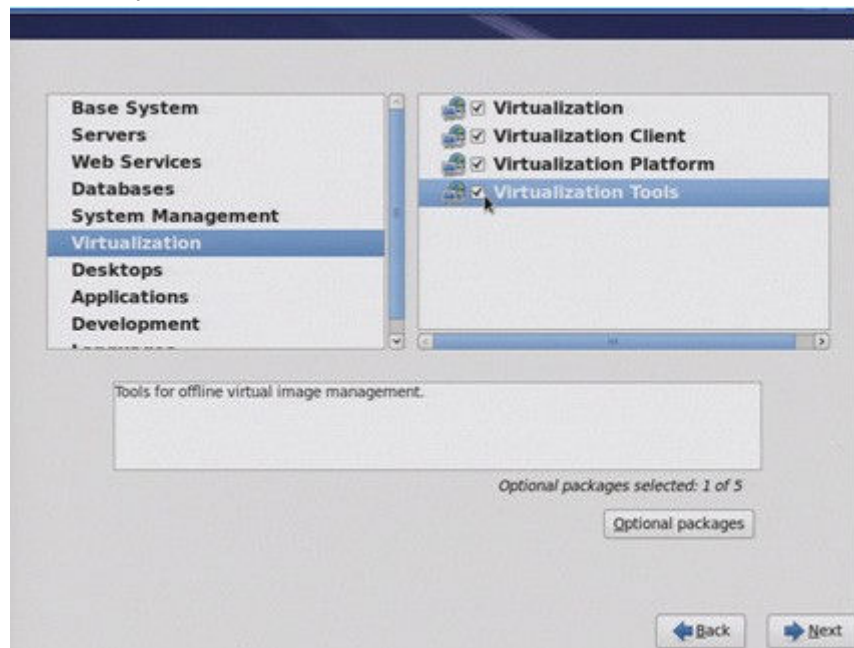
Installing and managing the virtual machine

KVM can be administrated similar to Xen. Both command line and GUI capabilities are available. KVM shares the same management GUI with Xen. The only difference is the virtualization host to be connected.

In both Red Hat Enterprise Linux and SLES environments, the management GUI can be evoked by typing `virt-manager` on the command prompt. In SLES, the GUI can also be accessed from the suite of utilities that are consolidated under YaST.

Before creating virtual machines on SLES hosts running KVM, ensure that the CPU's Virtualization Technology support option is enabled, and that the KVM kernel module kvm-amd, or kvm-intel, is loaded. Typing `lsmod |grep` kvm in the command prompt checks the status of KVM module.

For detailed virtual machine installation procedures, see Virtual machine installation and management.

The following figures show how to install the virtual machine:

Figure 55 Virtual Machine Manager Utility on RHEL
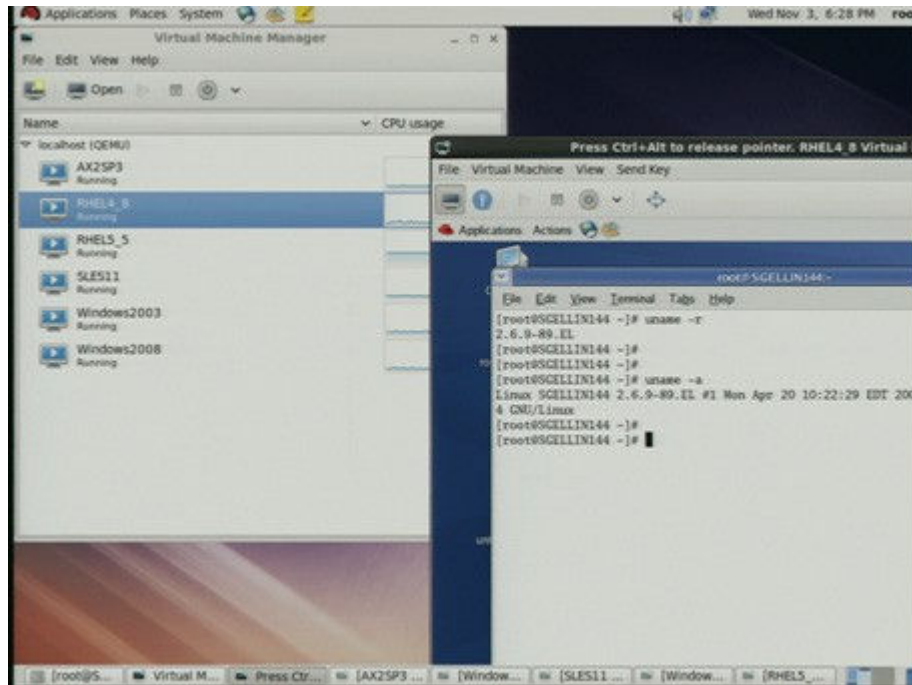
Figure 56 Virtual Machine Manager: virtual machine statuses
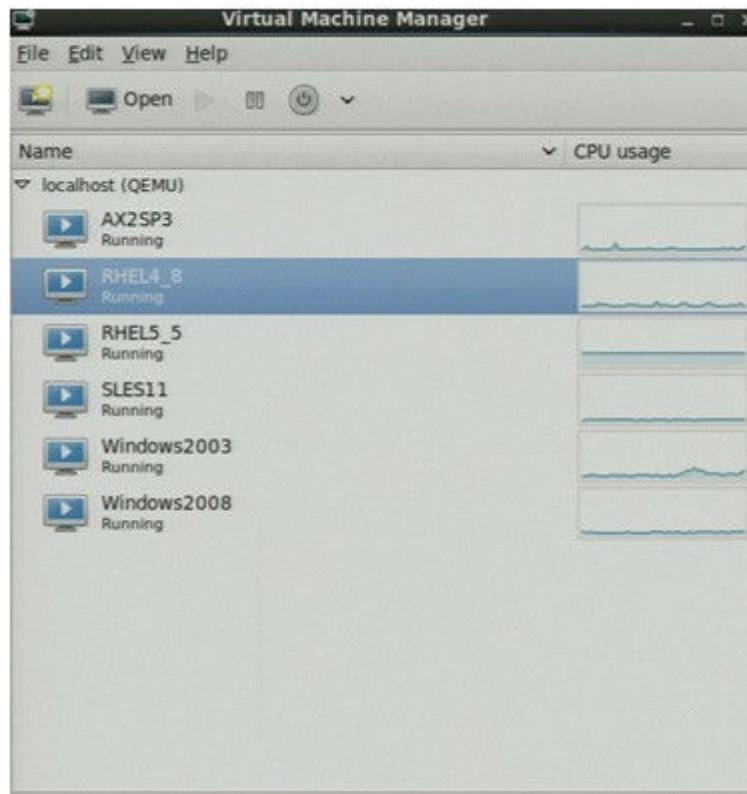
Figure 57 Evoke virtualization management GUI utility in SLES
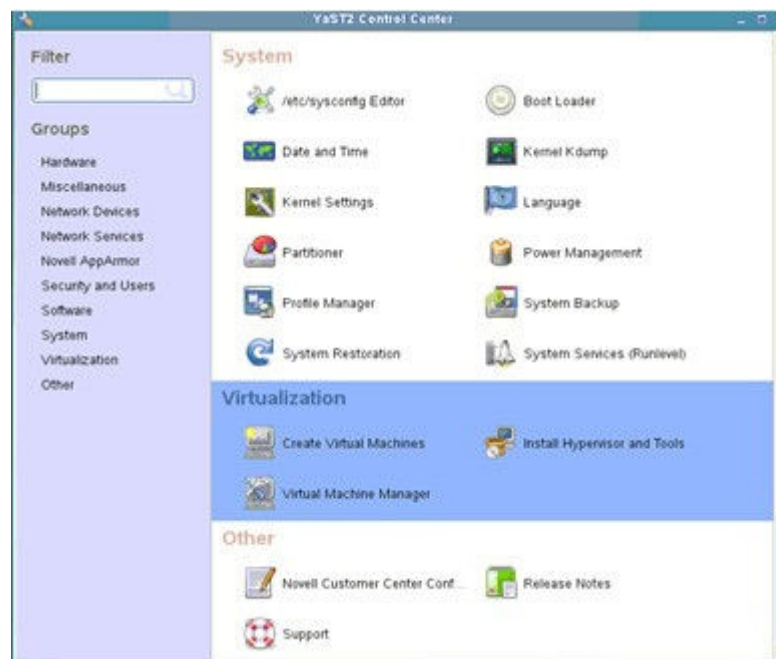
Figure 58 Virtualization Machine Manager

Figure 59 Creating a new Virtual Machine

Figure 60 Virtual Machine Manager Utility on SuSE with VMs running
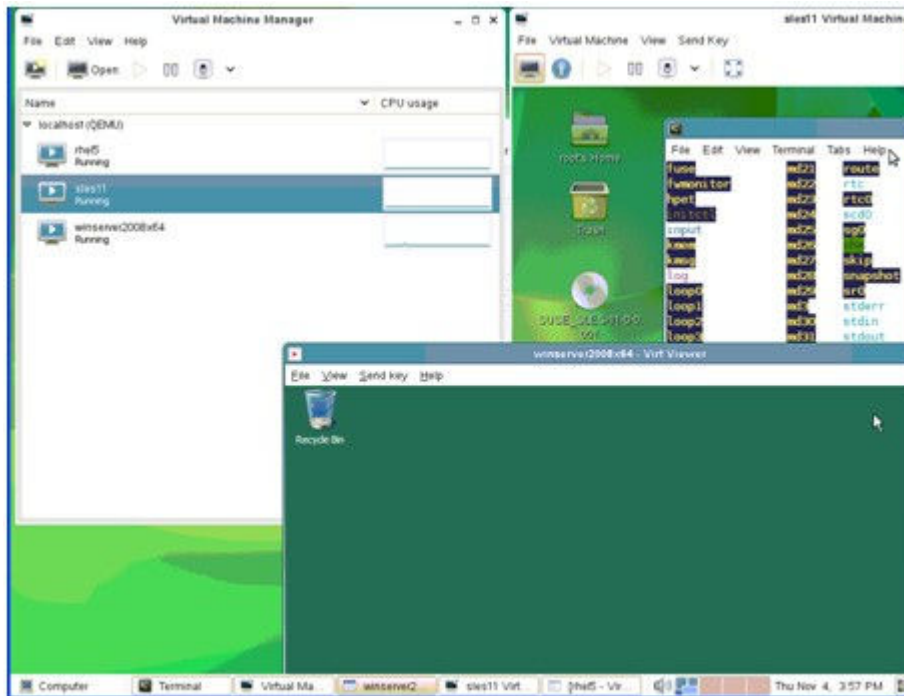
Storage management

As the administration of KVM is similar to Xen, storage management also follows in a similar manner. Most of the concepts Xen adopts apply to KVM, except the dom0 (read "dom zero") concept.

Dell EMC supports any of its supported operating systems which are also, in turn, supported by the operating system vendor on KVM as Guest operating system. By default, the guest operating system is a fully virtualized machine. The development community and operating system vendors have developed different paravirtualized drivers to emulate para-virtualization characteristics. Consult the OS provider for more information.

Connectivity and multipathing functionality

Dell EMC supports storage that is provisioned to hosts running KVM through the following connectivity:

- Fibre Channel

- iSCSI

- Fibre Channel on Ethernet (FCoE)

Dell EMC can support both Linux native multipathing (DM-MPIO) and PowerPath with KVM technology. It is the native Linux kernel in the host server that performs the necessary path failover and recovery for the server. Hence, the virtual machines are spared the burden of implementing load balancing and failover utilities. See the Dell EMC Simple Support Matrix Linux operating system footnotes in the base connectivity section for supported configurations.

Creating a VM on PowerPath

To create a VM on PowerPath devices on Red Hat Enterprise Linux 6.3 and later, turn off cgroups as follows:

About this task

(i) Note: For KVM on Red Hat Enterprise Linux 6.2 and previous releases, Dell EMC does not Support creating a VM directly on PowerPath devices. You must pass PowerPath devices through LVM to KVM as logical volumes.

Procedure

1. Edit `/etc/libvirt/qemu.conf` and put the following line into the cgroup_controllers section, as shown next:

   For Red Hat Enterprise Linux 6, `cgroup_controllers = [ "cpu", "memory" ]`

   For Red Hat Enterprise Linux 7, `cgroup_controllers = ["cpu","memory","cpuacct"]`

2. Reboot or restart libvirtd service after changing /etc/libvirt/qemu.conf for new configuration to take effect.

   For example:
   ```
   # service libvirtd restart
   ```

   For more information about KVM, see the following URL: http://www.linux-kvm.org/page/.

Citrix XenServer

Citrix XenServer is a Linux Xen kernel-based virtualization product that has gained popularity in the recent years. Besides the benefits provided by the Linux Xen kernel, XenServer implements other features for convenient VM management, storage management, and backup management.

XenServer overview

The basic setup of a standard XenServer implementation consists of a server with XenServer operating system installed, and a workstation with XenCenter management suite installed. Only a personal computer is needed to install XenCenter.

XenCenter software connects to the XenServer host through TCP/IP network and facilitates almost all VM-related operations, such as creating VMs, creating and assigning virtual disks, virtual network configuration, and VM backup, can be executed through XenCenter. One XenCenter application can manage multiple XenServer hosts and their VMs.

XenServer hosts can be grouped to form a server pool and facilitate VM live migration across hosts and other advanced operations.

XenServer uses a storage repository concept and converts all storage devices into the specific storage repository format. It is actually an LVM-based logical layer above the Linux block device layer. All virtual disks must be created from existing storage repositories. Virtual disks and storage repositories that are created by one XenServer host are portable to other XenServer hosts, but can only be recognized as LVM volumes by other Linux operating systems.

XenServer provides a StorageLink suite that manages heterogeneous storage environments.

It is recommended that customers interested in implementing XenServer with Dell EMC storage products to see the *Citrix XenServer Support Statement* for supported configurations with XenServer releases prior to Citrix XenServer 6.0.2. The letter of support is available in Dell EMC E-Lab Navigator, under the **Extended Support** tab. For Citrix XenServer 6.0.2 and later, see the Dell EMC Simple Support Matrix.

For more information regarding the virtual machine installation, see the Citrix website.

Connectivity and path management software

Dell EMC supports storage that is provisioned to hosts running XenServer through the following connectivity:

- Fibre Channel
- iSCSI
- Fibre Channel over Ethernet (FCoE)

XenServer support for multipathing is performed by using the Linux native multipathing (DM-MPIO Multipath). It is the native Linux kernel in the host server that performs the necessary path failover and recovery for the server. Hence, the virtual machines are spared the burden of implementing load balancing and failover utilities. See the Dell EMC Simple Support Matrix Linux operating system-specific footnotes in the base connectivity section of Dell EMC E-Lab Navigator for supported configurations.

Live VDI Migration

A Live VDI Migration feature is available in Citrix XenServer Enterprise Edition or later. Live VDI Migration allows system administrators to relocate a virtual machine's Virtual Disk Image (VDI) without shutting down the virtual machine.

This enables administrative operations such as:

- Move a virtual machine from cheap, local storage to fast, resilient, array-backed storage.

- Move a virtual machine from a development to a production environment.
- Move between tiers of storage when a virtual machine is limited by storage capacity.
- Perform storage array upgrades.

Moving virtual disks using XenCenter

To move virtual disks, complete the following steps:

Procedure

1. In the XenCenter pane, select the storage repository where the virtual disk is currently stored and then click **Storage**.
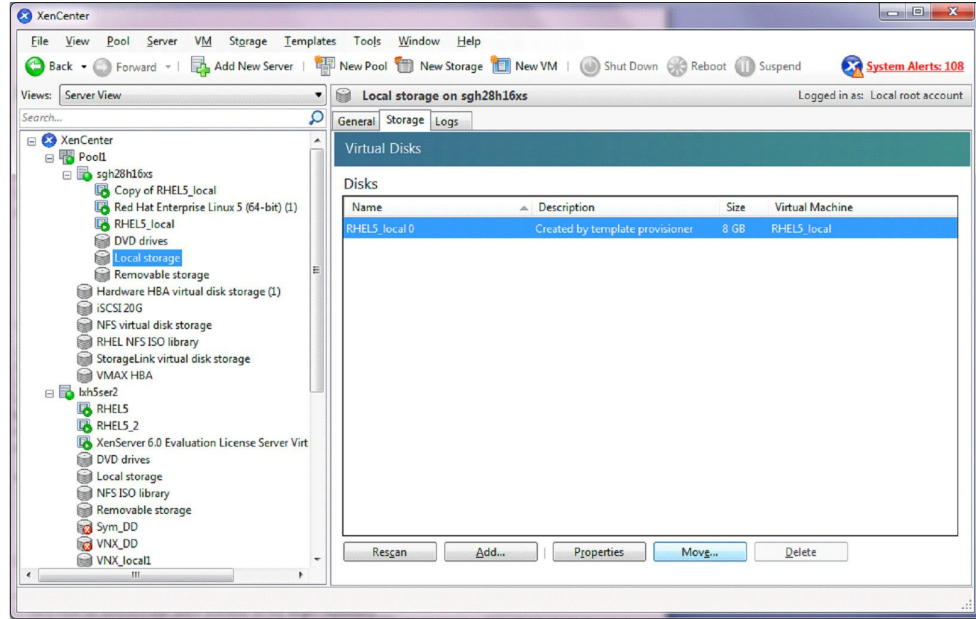
2. In the **Virtual Disks** list, select the virtual disk that you would like to move and then click **Move**. The **Move Virtual Disk** dialog box displays.
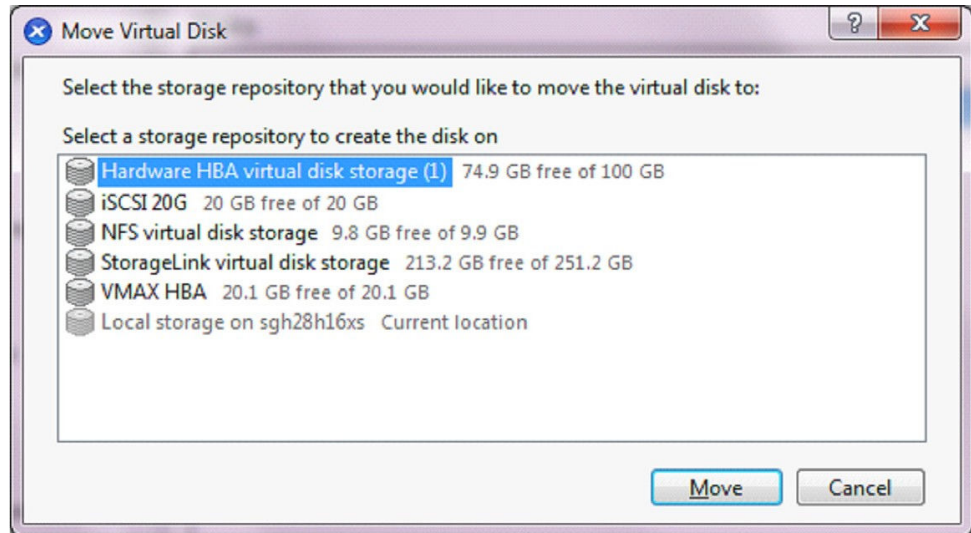
   (i) **Note:** Ensure that the SR has sufficient space for another virtual disk. The available space is shown in the list of available SRs.

3. Click **Move** to move the virtual disk.

Limitations and caveats

Live VDI Migration is subject to the following limitations and caveats:

- There must be sufficient disk space available on the target repository.

- VDIs located on Integrated StorageLink (iSL) SRs cannot be migrated.

- VDIs with more than one snapshot cannot be migrated.

VM migration with XenMotion and Storage XenMotion

XenMotion overview

XenMotion is a feature that allows live migration of virtual machines. With XenMotion, you can move a running virtual machine from one physical host system to another without any disruption or downtime.

Additionally, Storage XenMotion allows a virtual machine to be moved from one host to another, where the virtual machine is not located on storage that is shared between the two hosts. As a result, a virtual machine that is stored on local storage can be migrated without downtime and moved from one pool to another.

Compatibility requirements

When migrating a virtual machine with XenMotion or Storage XenMotion, the new virtual machine host must meet the following compatibility requirements:

- XenServer Tools must be installed on each virtual machine that you want to migrate.

- The target host must have the same or a more recent version of XenServer installed as the source host.

- The target host must have sufficient spare memory capacity or be able to free sufficient capacity using Dynamic Memory Control. If there is not enough memory, the migration fails to complete.

- For Storage XenMotion, note the following:

  - If the CPUs on the source host and target host are different, the target host must provide at least the entire feature set as the source host's CPU. Consequently, it is almost impossible to move a virtual machine between, for example, AMD and Intel processors.

  - Virtual machines with more than one snapshot cannot be migrated.

  - Virtual machines with more than six attached VDIs cannot be migrated.

  - The target storage must have enough free disk space (for the virtual machine and its snapshot) available for the incoming virtual machines. If there is not enough space, the migration fails to complete.

Migrating the virtual machine from one host to another host

To migrate the virtual machine from one host to another host, complete the following steps:

Procedure

1. Right-click the virtual server from the XenCenter pane. In the drop-down menu, select **Migrate to Server** > **Migrate VM wizard**.
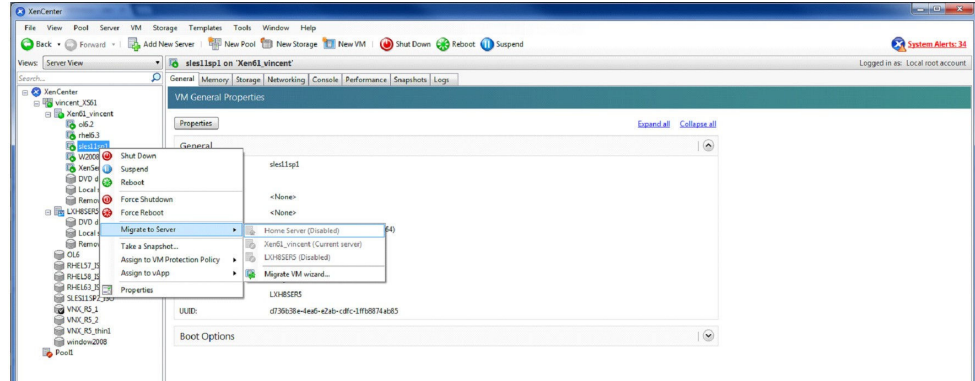
The **Destination** window displays.
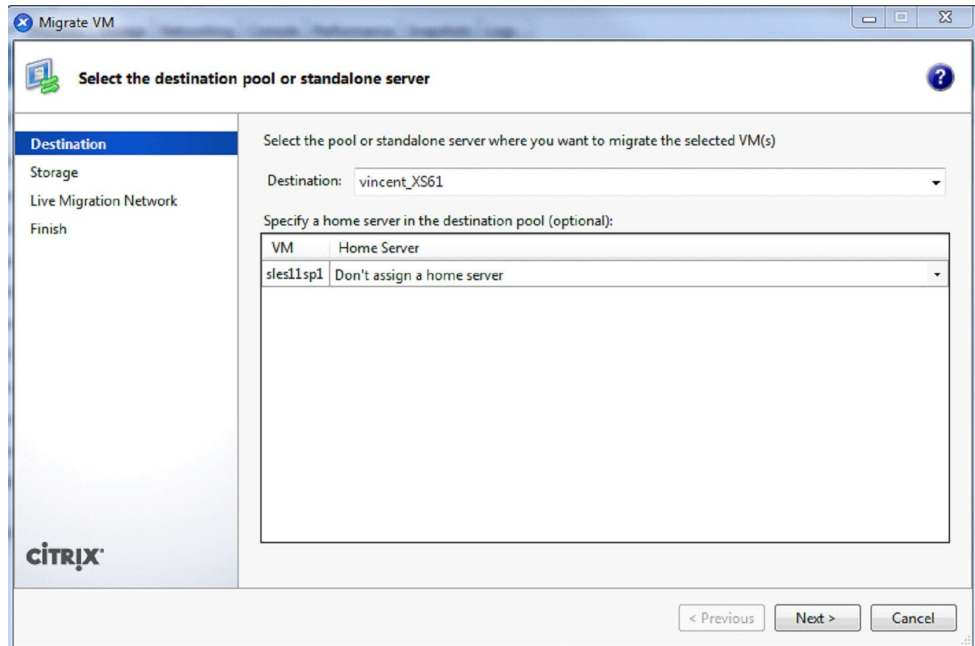
2. In the window, complete the following fields:

   a. In the **Destination** field, choose the destination in the drop-down menu where you want to migrate the virtual machine.

   b. Select a standalone server or a pool from the drop-down menu in the **Specify a home server in the destination pool** .

   c. Click **Next**. The **Storage** window displays.
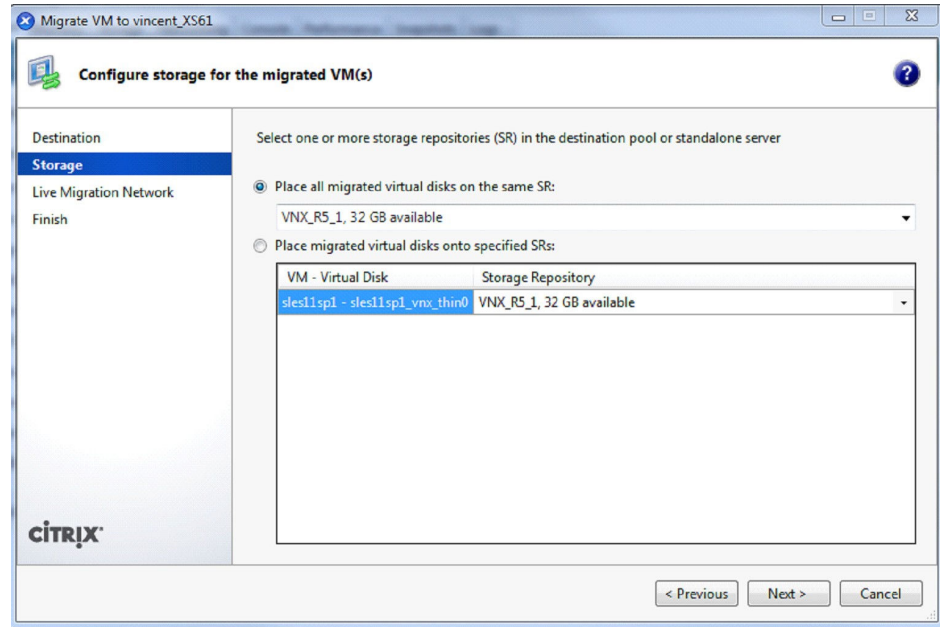
3. In the **Storage** window, complete the following fields:

    a. Specify the storage repository where you would like to place the migrated virtual machine's virtual disks.

    b. Click **Next**.

    The **Live Migration Network** window displays.
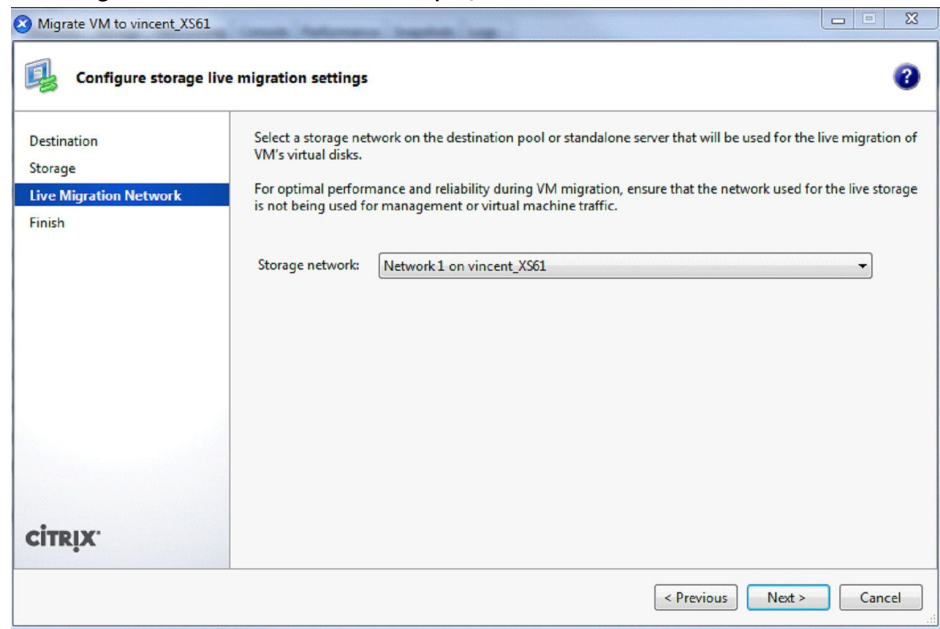
4. From the **Storage network** drop-down list, select a network on the destination pool that will be used for the live migration of the virtual machine's virtual disks and click **Next**.

    The **Finish** window displays.
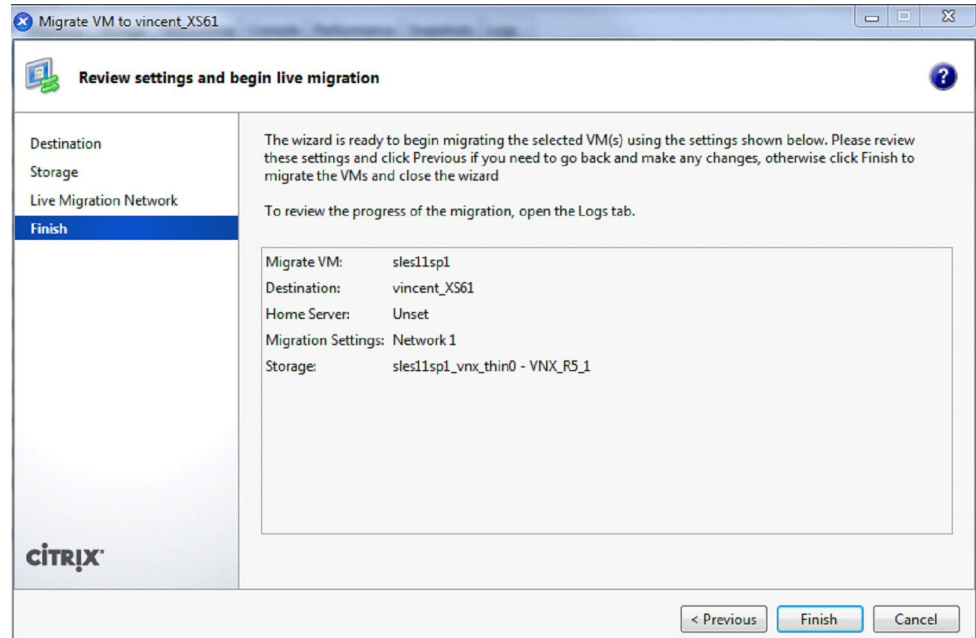
5. Review the configuration settings and click **Finish** to start migrating the virtual machine.

(i) **Note:**

- In Folder View, you can migrate a virtual machine by dragging and dropping it in the Resources panel.

- Due to performance reasons, it is recommended that you do not use your management network for live migration.

- Moving a virtual machine from one host to another preserves the virtual machine state. The state information includes information that defines and identifies the virtual machine as well as the historical performance metrics, such as CPU and network usage.

Lmitations and caveats

XenMotion and Storage XenMotion are subject to the following limitations and caveats:

- Virtual machine with VDIs located on Integrated StorageLink (iSL) SRs cannot be migrated.
- Virtual machine using PCI pass-thru cannot be migrated.
- Virtual machine performance will be reduced during migration.
- For Storage XenMotion, pools protected by High Availability (HA) or Work Load Balancing (WLB) should have HA and WLB disabled before attempting virtual machine migration.
- Time to completion of virtual machine migration will depends on the memory footprint of the virtual machine, and its activity, in addition, virtual machine being migrated with Storage XenMotion will be affected by the size of the VDI and its storage activity.
- IPv6 Linux virtual machine requires a Linux Kernel greater than 3.0.

Oracle Linux Virtualization Manager

Oracle Linux Virtualization Manager is a server virtualization management platform that can be easily deployed to configure, monitor, and manage an Oracle Linux Kernel-based Virtual Machine (KVM) environment with enterprise-grade performance and support from Oracle.

Overview

Based on the open source oVirt project, Oracle Linux Virtualization Manager is a server virtualization management platform that can be used to configure, monitor, and manage an Oracle Linux Kernel-based Virtual Machine (KVM) environment, including hosts, virtual machines, storage, networks, and users. The Manager is accessed through the Administration Portal or VM Portal, web-based portals that are available from a Oracle Linux Virtualization Manager landing page.

Figure 61 Overview of architecture
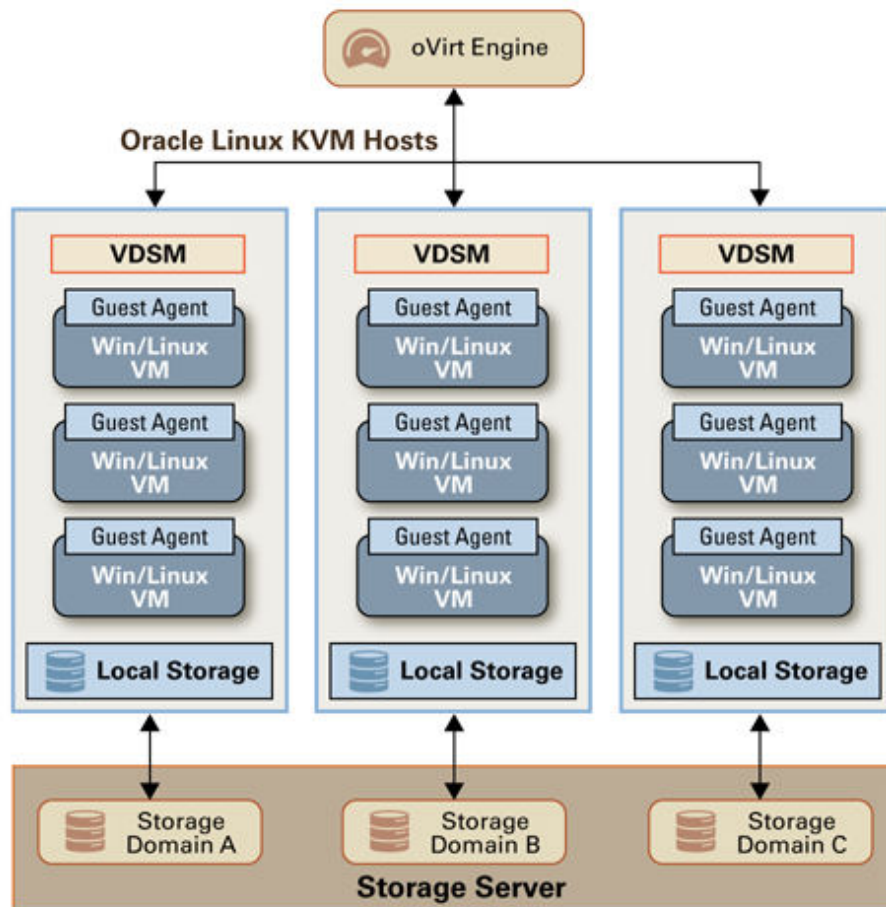
The workhorse of Oracle Linux Virtualization Manager is the engine (ovirt-engine) which is a JBoss-based Java application that runs as a web service and provides centralized management for server and desktop virtualization. The oVirt engine runs on an Oracle Linux server and provides the administration tools for managing the Oracle Linux Virtualization Manager environment. Oracle Linux KVM hosts provide the compute resources for running virtual machines.

Connectivity and multipathing functionality

Dell EMC supports storage that is provisioned to hosts running Oracle Linux Virtualization Manager through the following connectivity:

- Fibre Channel
- iSCSI

Dell EMC only supports Linux native multipathing (DM-MPIO) with Oracle Linux Virtualization Manager technology. See Native Multipath Failover, for Linux native multipath configuration.

Native Clusters

Supported clusters

Dell EMC supports a wide range of cluster and high availability technologies on Linux. The Dell EMC Simple Support Matrix provides a detailed listing of available cluster configurations for operating system versions, path management solutions, and array support.

Red Hat Cluster Suite

Red Hat defines the Red Hat Cluster Suite (RHCS) in RHEL 5 as a collection of technologies working together to provide data integrity and the ability to maintain application availability in the event of a failure. RHCS uses the Cluster Configuration System (CCS) to create the cluster environment, manage the individual nodes, and effect changes in the cluster environment. The Resource Group Manager (rgmanager) provides tools to manage services provided by the cluster.

The Red Hat Cluster Suite provides multiple features, including failover domains, fencing, Ethernet channel bonding, and so on. Discussion of these features is beyond the scope of this document and is instead available in the administration guides published by Red Hat.

Red Hat Cluster Suite is used as part of Dell EMC's qualification effort of the GFS file system. All features of the cluster and possible configuration aspects are not explored in this document; however, they are directly supported by the Red Hat Corporation. It is recommended that you consult Red Hat's documentation of RHCS for detailed configuration instructions and supported configurations on the Red Hat Product Documentation page.

Dell EMC supports the Red Hat release of RHCS and GFS found in RHEL. See Dell EMC Simple Support Matrix for the most up-to-date details of supported configurations.

The following are supported:

- Active/passive or active/active cluster
- Dell EMC PowerPath or Linux native DM-MPIO
- RHEL GFS and GFS2 clustered file system
- RHCS quorum disk
  - With PowerPath, support began with RHEL 5.5 and continued with RHEL 6.0, see Best practices and additional installation information.
- RHCS fence_scsi() and sg3util sg_persist()
  - With PowerPath, support began with RHEL 5.5 and continued with RHEL 6.0, 7.0 and RHEL 8.0, see Best practices and additional installation information.

RHCS ensures high availability and manageability of critical network resources including data, applications, and services. It is a multi node clustering product that supports failover, failback, and migration (load balancing) of individually managed cluster resources.

RHCS includes several important features. These include:

- Support for Fibre Channel or iSCSI storage.
- Multi-node active cluster, up to 128 (consult Red Hat for the latest supported number).
- A single point of administration through either a graphical tool or a command-line tool.
- The ability to tailor a cluster to the specific applications and hardware infrastructure that fit an organization.
- Dynamic assignment and reassignment of server storage on an as-needed basis.
- Time-dependent configuration enables resources to fail back to repaired nodes.
- Support for shared disk systems.
- Support for cluster file systems like GFS and GFS2.
- Although shared disk systems are supported, they are not required.
- Support for cluster aware logical volume managers like CLVM.

Global File System

Global File System (GFS) is an open-source cluster file system that is developed by Red Hat. It operates within the framework provided by Red Hat Cluster Suite and enables shared storage and file systems across multiple nodes in a cluster. A further development, GFS2, was derived from GFS and integrated into Linux kernel version 2.6.19.

GFS is a shared disk file system, whereas, GFS2 can be used on a stand-alone system, like any other file system, or as a shared disk file system in a cluster.

While RHCS provides the infrastructure for configuring cluster nodes, maintaining cluster health, enabling services, and so on, GFS enables shared storage and file system across multiple nodes in a cluster. GFS bundles multiple strategies for Lock Management, such as the Grand Unified Lock Manager (GULM) and the Distributed Lock Manager (DLM).

GFS supports file system features such as quotas, journaling, and dynamic file system changes. In addition, PowerPath is supported on the host for managing multiple redundant paths to the shared storage from a node and for load balancing.

Best practices and additional installation information

- Server cluster heartbeat limitation per Red Hat specification (< 2.0ms RTT)
  - See the Red Hat Knowledge base Article ID: 58412, Support for Red Hat Enterprise Linux High Availability Stretch and Multi-Site Cluster Architectures.
- An RHCS quorum disk has been qualified on Dell EMC storage. If an RHCS quorum disk is configured under a PowerPath pseudo device, it is recommended that the TOTEM TOKEN should be set at 330000ms. This is to protect against a worst case failure scenario regarding paths to the quorum disk. It is further recommended that if the recommended setting is not used; test your desired settings in a non-production environment before implementing them. If not, an unstable cluster may result.
  Find further guidance about RHCS quorum disks in How to Optimally Configure a Quorum Disk in Red Hat Enterprise Linux Clustering and High-Availability Environments in the Red Hat Customer Portal

Follow Red Hat's best practice for use with Linux native DM-MPIO.

- `fence_scsi()` and `sg_persist()` were qualified on a distributed device under both PowerPath and Linux DM-MPIO.

- **Patches for** `fence_scsi()` **and** `fence_scsi_test()` **to support PowerPath pseudo devices may be required for RHEL 5.x configurations in order for them to detect the PowerPath pseudo device.**

```
- Patches to fence_scsi()
# sub get_scsi_devices [This is the original macro]
          #{
 # my ($in, $out, $err);
# my $cmd = "lvs --noheadings --separator : -o vg_attr,devices";
# my $pid = open3($in, $out, $err, $cmd) or die "$!\n";
#
# waitpid($pid, 0);
#
# die "Unable to execute lvs.\n" if ($?>>8);
#
# while (<$out>)
# {
# chomp;
# print "OUT: $_\n" if $opt_v;
#
# my ($vg_attrs, $device) = split(/:/, $_);
#
# if ($vg_attrs =~ /.*c$/)
# {
# $device =~ s/\(.*\)//;
# push(@volumes, $device);
# }
# }
#
# close($in);
# close($out);
# close($err);
#}
sub get_scsi_devices [This is the change]
{
 open (FILE, "/var/run/scsi_reserve") or die "$!\n";
while (<FILE>)
     { chomp; push(@volumes, $_);
 }
    close FILE;
 }LEX
     - Patches to fence_scsi_test() sub get_block_devices
    {
my $block_dir = "/sys/block";
     opendir(DIR, $block_dir) or die "Error: $! $block_dir\n";
 my @block_devices = grep { /^sd*/ | /^emcpower*/ } readdir(DIR);
closedir(DIR);
     for $dev (@block_devices)
    {
      push @devices, "/dev/" . $dev;
```

1. **· Patches for /usr/sbin/fence_scsi to support pseudo devices may be required for RHEL 6.x configurations in order for it to detect the PowerPath pseudo device.**

```
 - Patches to /usr/sbin/fence_scsi sub get_devices_scsi ()
   {
my $self = (caller(0))[3]; my @devices;
   opendir (\*DIR, "/sys/block/") or die "$!\n";
 # @devices = grep { /^sd/ } readdir (DIR); #[This is the original macro]
@devices = grep { /^sd/ | /^emcpower/ } readdir (DIR);#[This is the change]
closedir (DIR);
 return (@devices);
```

Additional installation information

For additional information regarding Red Hat's Cluster Suite implementation or to configure a cluster, go to the Red Hat website.

Red Hat High Availability Add-On

Starting from RHEL 6, Red Hat defines the High Availability Add-On as an integrated set of software components that can be deployed in various configurations to suit your needs for performance, high availability, load balancing, scalability, file sharing, and economy.

Red Hat Enterprise Linux 7 High Availability Add-On introduces a new suite of technologies that underlying high-availability technology based on Pacemaker and Corosync that completely replaces the CMAN and RGManager technologies used in Red Hat Enterprise Linux 6 High Availability Add-On.

The High Availability Add-On consists of the following major components:

- Cluster infrastructure: Provides fundamental functions for nodes to work together as a cluster: configuration file management, membership management, lock management, and fencing.

- High availability Service Management: Provides failover of services from one cluster node to another in case a node becomes inoperative.

- Cluster administration tools: Configuration and management tools for setting up, configuring, and managing a High Availability Add-On. The tools are for use with the Cluster Infrastructure components, the high availability and Service Management components, and storage.

You can supplement the High Availability Add-On with the following components:

- Red Hat GFS2 (Global File System 2): Part of the Resilient Storage Add-On, this provides a cluster file system for use with the High Availability Add-On. GFS2 allows multiple nodes to share storage at a block level as if the storage were connected locally to each cluster node. GFS2 cluster file system requires a cluster infrastructure.

- Cluster Logical Volume Manager (CLVM): Part of the Resilient Storage Add-On, this provides volume management of cluster storage. CLVM support also requires cluster infrastructure.

Red Hat High Availability Add-On is used as part of Dell EMC's qualification effort of the GFS2 file system. All features of the cluster and possible configuration aspects are not explored in this document; however, they are directly supported by the Red Hat Corporation. It is recommended that you consult Red Hat's documentation of High Availability Add On for detailed configuration instructions and supported configurations on the Red Hat Product Documentation page.

Dell EMC supports the Red Hat High Availability Add-On and GFS2 found in RHEL. See the Dell EMC Simple Support Matrix for the most up-to-date details of supported configurations.

The following are supported:

- Active/passive or active/active cluster
- Dell EMC PowerPath or Linux native DM-MPIO
- RHEL GFS2 clustered file system
- CLVM
- Quorum
- fence_scsi() and sg3util sg_persist()

Heartbeat

Dell EMC supports the SuSE release of Heartbeat found in SLES. See the Dell EMC Simple Support Matrix for details of supported configurations.

Heartbeat ensures high availability and manageability of critical network resources including data, applications, and services. It is a multi node clustering product that supports failover, failback, and migration (load balancing) of individually managed cluster resources.

Heartbeat includes several important features. These include:

- Support for Fibre Channel or iSCSI storage.

- Multinode active cluster, up to 16.

- A single point of administration through either a graphical Heartbeat tool or a command-line tool.

- The ability to tailor a cluster to the specific applications and hardware infrastructure that fit an organization.

- Dynamic assignment and reassignment of server storage on an as-needed basis.

- Time-dependent configuration enables resources to fail back to repaired nodes.

- Support for shared disk systems.

- Support for cluster file systems like OCFS 2.

- Shared disk systems are supported, they are not required.

- Support for cluster aware logical volume managers like EVMSS.

Heartbeat cluster component

The following components make up a Heartbeat version 2 cluster found in SuSE's SLES:

- From 2 to 16 Linux servers, each containing at least one local disk device.

- Heartbeat software running on each Linux server in the cluster.

- (Optional) A shared disk subsystem that is connected to all servers in the cluster.

- (Optional) High-speed Fibre Channel cards, cables, and switch used to connect the servers to the shared disk subsystem. Or network adapter cards, cables, and switch used to connect the servers to the shared iSCSI disk subsystem.

- At least two communications mediums over which Heartbeat servers can communicate. These currently include Ethernet (mcast, ucast, or bcast) or serial.

- A STONITH device: A STONITH device is a power switch which the cluster uses to reset nodes that are considered dead. Resetting non-heartbeating nodes is the only reliable way to ensure that no data corruption is performed by nodes that hang and only appear to be dead.

Installation information and additional details

For additional information regarding SuSE's implementation of Heartbeat or to configure a cluster, see the SuSE website.

SuSE High availability extension

SuSE's High Availability Extension (HAE) is an integrated suite of open-source clustering technologies used to implement highly available physical and virtual Linux clusters, and to eliminate single points of failure. It ensures the high availability and manageability of critical resources including data, applications, and services. Therefore, it helps maintain business continuity, protect data integrity, and reduce unplanned downtime of mission-critical Linux workloads.

It ships with essential monitoring, messaging, and cluster resource management functionality, supporting failover, failback, and migration of individually managed cluster resources. HAE is available as add-on product to SLES 11.

Dell EMC supports the SuSE release of High Availability Extension found in SLES. See Dell EMC Online Support for details of supported configurations.

HAE includes several important features. These include:

- Active/active and active/passive cluster configurations (N+1, N+M, N to 1, N to M)
- Multinode active cluster, up to 16
- Hybrid physical and virtual cluster support
- Tools to monitor the health and status of resources, manage dependencies, and automatically stop and start services based on highly configurable rules and policies
- Time-dependent configuration enables resources to fail back to repaired nodes
- Support for cluster aware logical volume managers like cLVM
- Dynamically assign and reassign server storage as needed
- Graphical and CLI-based management tools

HAE components

The following components make up a High Availability Extension cluster found in SuSE's SLES:

- From 1 to 16 Linux servers, each containing at least one local disk device.
- High Availability Extension software running on each Linux server in the cluster.
- (Optional) A shared disk subsystem connected to all servers in the cluster.
- (Optional) High-speed Fibre Channel cards, cables, and switch used to connect the servers to the shared disk subsystem. Or network adapter cards, cables, and switch used to connect the servers to the shared iSCSI disk subsystem.
- At least two TCP/IP communications mediums that support multicasting.
- A STONITH device: A STONITH device is a power switch which the cluster uses to reset nodes that are considered dead. Resetting nonheart beating nodes is the only reliable way to ensure that no data corruption is performed by nodes that hang and only appear to be dead.

Installation information and additional details

For additional information regarding SuSE's implementation of High Availability Extension or to configure a cluster, see the SuSE website.