



# Dell EMC Host Connectivity Guide for Linux

P/N 300-003-865

REV 44

This document is not intended for audiences in China, Hong Kong, Taiwan, and Macao.

Copyright © 2003 – 2017 Dell Inc. or its subsidiaries. All rights reserved.

Published June 2017

Dell believes the information in this publication is accurate as of its publication date. The information is subject to change without notice.

THE INFORMATION IN THIS PUBLICATION IS PROVIDED "AS IS." DELL INC. MAKES NO REPRESENTATIONS OR WARRANTIES OF ANY KIND WITH RESPECT TO THE INFORMATION IN THIS PUBLICATION, AND SPECIFICALLY DISCLAIMS IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Use, copying, and distribution of any Dell EMC software described in this publication requires an applicable software license.

Dell, EMC<sup>2</sup>, EMC, and the EMC logo are registered trademarks or trademarks of Dell Inc. or its subsidiaries. All other trademarks used herein are the property of their respective owners.

For the most up-to-date regulator document for your product line, go to Dell EMC Online Support (<https://support.emc.com>).

# CONTENTS

Preface .....	9
Chapter 1	Introduction
Operating system limits and guidelines.....	14
Host initiators .....	14
Logical unit .....	14
Configuration example.....	16
Storage attach .....	19
Zoning recommendation .....	21
Devices and operations.....	22
SCSI device addressing.....	22
SCSI device operation interfaces .....	23
LUN scanning mechanisms.....	26
System reboot.....	26
HBA driver reload.....	26
SCSI scan function in /proc.....	26
SCSI scan function in /sys.....	27
SCSI scan through HBA vendor scripts .....	27
SCSI scan through Linux distributor provided scripts.....	28
Persistent binding .....	30
HBA persistent binding.....	30
Udev .....	31
Native MPIO .....	31
PowerPath pseudo-names.....	31
Logical volumes .....	31
Mitigating the effects of storage array migration for Linux hosts.....	32
Useful utilities .....	34
Disk partition adjustment for VMAX series, VNX series, VNXe series, Unity series, CLARiiON, or XtremIO.....	36
Track boundaries .....	37
RAID 5 boundaries.....	37
Metastripe boundaries .....	39
VNX series, VNXe series, Unity series, or CLARiiON systems.....	39
Determining the correct offset to partition.....	39
Aligning the partition .....	41
Operating systems .....	44
Host software.....	45
Dell EMC Solutions Enabler for Linux.....	45
Navisphere CLI .....	45
Unisphere CLI.....	47
Dell EMC replication software .....	47
Server vendors .....	48
Host bus adapters.....	49
Converged Network Adapters .....	50
Dell EMC storage.....	51
VMAX series.....	51
Unity series.....	52
VNX series or CLARiiON .....	52
VPLEX .....	52

	XtremIO .....	53
	ScaleIO .....	53
	XtremCache.....	55
<b>Chapter 2</b>	<b>Fibre Channel Connectivity</b>	
	Introduction .....	58
	Configuring the HBAs for a Linux host .....	59
	Prerequisites for first-time installation.....	59
	Emulex Fibre Channel HBA .....	59
	QLogic Fibre Channel HBA .....	63
	Brocade Fibre Channel HBA .....	68
	SNIA API for third-party software (Solution Enabler).....	69
	Hitachi Virtage.....	72
<b>Chapter 3</b>	<b>Fibre Channel over Ethernet Connectivity</b>	
	Introduction .....	76
	Configuring the Linux host .....	78
	Zoning best practices.....	78
	CNAs.....	78
	Cisco Unified Computing System.....	90
<b>Chapter 4</b>	<b>iSCSI Connectivity</b>	
	Introduction .....	94
	iSCSI discovery .....	95
	Digests.....	95
	iSCSI error recovery.....	96
	iSCSI security.....	96
	iSCSI solutions .....	98
	General best practices.....	98
	General supported configurations .....	98
	Dell EMC native iSCSI targets .....	99
	Native Linux iSCSI driver.....	102
	Software and hardware iSCSI initiator.....	102
	Native Linux iSCSI Attach.....	103
	open-iscsi driver .....	103
	Installing the open-iscsi driver .....	106
	Setting initiator name in software iSCSI.....	108
	Selective target(s) login.....	109
	Starting and stopping the iSCSI driver.....	110
	Dynamic LUN discovery .....	111
	Mounting and unmounting iSCSI file systems automatically (RHEL, Asianux, and SLES) .....	111
	Excessive dropped session messages found in /var/log/messages .....	112
	iSCSI Write Optimization in Unity, VNX series, or CLARiiON CX environment.....	113
	Known problems and limitations.....	116
<b>Chapter 5</b>	<b>Booting From SAN</b>	
	Supported environments.....	122
	Notes.....	122
	Limitations and guidelines .....	123
	Preparing host connectivity .....	124



	Guidelines.....	124
	Single and dual path configuration examples .....	125
	Configuring a SAN boot for FC attached host .....	126
	Prepare host connectivity .....	126
	Installing and configuring Fibre Channel HBA .....	126
	Updating HBA BIOS and firmware .....	126
	Enabling HBA port and Selecting boot LUN.....	126
	Configuring SAN boot for iSCSI host.....	132
	Setting up the hardware iSCSI SAN boot .....	132
	Software iSCSI SAN boot.....	136
	Configuring SAN boot for FCoE attached host .....	139
	Installing and configuring Intel card for software FCoE boot .....	139
	Installing an OS on FCoE external devices.....	142
	Multipath booting from SAN .....	146
	Overview .....	146
	Configuring DM-MPIO for SAN boot devices.....	147
	PowerPath booting from SAN.....	154
	Guidelines for booting from Symmetrix, XtremIO, VNX series, VNXe series, Unity series, or CLARiiON.....	155
	Dell EMC Symmetrix-specific guidelines .....	155
	VNX series, VNXe series, Unity series, or CLARiiON-specific guidelines	156
Chapter 6	Path Management	
	Introduction .....	158
	PowerPath.....	159
	Multiple data paths and load balancing feature.....	159
	Automatic path failover feature.....	159
	Veritas Dynamic Multipathing .....	160
	Device-mapper multipath I/O (DM-MPIO) .....	161
Chapter 7	Native Multipath Failover	
	Storage arrays and code revisions .....	164
	VMAX series behavior.....	165
	Unity series, VNX series, and CLARiiON behavior .....	165
	<b>XtremIO behavior</b> .....	167
	Supported host bus adapters .....	168
	Supported operating systems.....	169
	Server platforms .....	170
	DM-MPIO on IBM zSeries .....	170
	Configuration requirements.....	171
	Useful utilities .....	172
	Known issues.....	173
	MPIO configuration for VMAX series.....	178
	RedHat Enterprise Linux (RHEL).....	178
	Oracle Linux and VM server.....	179
	SuSE Linux Enterprise server .....	179
	MPIO configuration for Unity storage, VNX Unified Storage, and CLARiiON	180
	Blacklisting the Unity series, VNX series, or CLARiiON LUNZ.....	180
	Failover mode.....	180
	Red Hat Enterprise Linux (RHEL).....	181
	Red Hat Linux 5.0 (and point releases).....	181
	Red Hat Linux 6.0 (and point releases).....	182
	Red Hat Linux 7.0 (and point releases).....	185

RHEL7.2 and later.....	187
Oracle Linux and VM Server.....	188
SuSE Linux Enterprise Server (SLES).....	188
MPIO configuration for Dell EMC Invista or VPLEX virtualized storage.....	194
Red Hat Enterprise Linux (RHEL).....	194
Oracle Linux and VM Server.....	195
SuSE Linux Enterprise Server (SLES).....	195
OPM.....	196
MPIO configuring for XtremIO storage.....	197
Red Hat Enterprise Linux (RHEL).....	197
Oracle Linux and VM Server.....	197
SuSE Linux Enterprise Server (SLES).....	198
Changing the path selector algorithm.....	199
Configuring LVM2.....	201
Configuring LVM2 for DM-MPIO on RHEL.....	201
Configuring LVM2 for DM-MPIO on SLES.....	202
Disabling Linux Multipath.....	203

Chapter 8

Virtualization

Linux virtualization.....	206
Benefits.....	206
Xen Hypervisor.....	207
Virtualization modes.....	207
Virtual machine installation and management.....	208
Storage management.....	214
Connectivity and path management software.....	214
Kernel-based Virtual Machine (KVM).....	216
Introduction.....	216
Implementing KVM.....	216
Installing and managing the virtual machine.....	220
Storage management.....	223
Connectivity and multipathing functionality.....	223
Citrix XenServer.....	225
XenServer overview.....	225
Connectivity and path management software.....	225
Live VDI Migration.....	226
VM migration with XenMotion and Storage XenMotion.....	228
Oracle VM Server.....	233
OVM overview.....	233
Connectivity and multipathing functionality.....	233

Chapter 9

Virtual Provisioning

Virtual Provisioning on VMAX series.....	236
Terminology.....	236
Virtual Provisioning on VNX, Unity, or CLARiiON.....	238
Virtual Provisioning on XtremIO.....	239
Space reclamation.....	240
Veritas Storage Foundation.....	240
Linux filesystem.....	240
Implementation considerations.....	241
Over-subscribed thin pools.....	241
Thin-hostile environments.....	242
Pre-provisioning with thin devices in a thin hostile environment.....	242

Host /boot, / (root), /swap, and /dump devices positioned on Symmetrix Virtual Provisioning (tdev) devices.....	243
Cluster configuration.....	244
Operating system characteristics .....	245
Thin pool exhaustion.....	245
Filesystem utilities.....	245
Filesystems .....	246
Volume Managers .....	247
Path Management.....	248
Pre-provisioned thin devices .....	250

## Chapter 10

VPLEX	
VPLEX overview .....	252
VPLEX documentation .....	252
Prerequisites .....	253
Veritas DMP settings with VPLEX .....	253
Host Configuration for Linux: Fibre Channel HBA Configuration .....	254
Setting queue depth and execution throttle for QLogic .....	254
Setting Queue Depth and Queue Target for Emulex.....	261
Provisioning and exporting storage.....	263
VPLEX with GeoSynchrony v4.x.....	263
VPLEX with GeoSynchrony v5.x.....	264
VPLEX with GeoSynchrony v6.x.....	264
Storage volumes .....	265
Claiming and naming storage volumes .....	265
Extents .....	265
Devices .....	265
Distributed devices.....	266
Rule sets .....	266
Virtual volumes .....	266
System volumes.....	267
Metadata volumes .....	267
Logging volumes.....	267
Required storage system setup .....	268
Required Symmetrix FA bit settings.....	268
Supported storage arrays .....	269
Initiator settings on back-end arrays .....	269
Host connectivity.....	270
Exporting virtual volumes to hosts.....	271
Front-end paths.....	274
Viewing the World Wide Name for an HBA port .....	274
VPLEX ports.....	274
Initiators .....	274
Configuring Linux hosts to recognize VPLEX volumes.....	276
Linux native cluster support .....	277
Supported Red Hat RHCS configurations and best practices .....	278
..... Supported SUSE HAE configurations and best practices	280
Optimal-Path-Management (OPM) feature.....	282
VPLEX OPM feature overview .....	282
Host connectivity best practices while using OPM.....	283
Host multipathing software configuration while using OPM .....	285
Native MPIO .....	289

Chapter 11	Native Clusters	
	Supported clusters .....	294
	Red Hat Cluster Suite (RHCS) .....	295
	Global File System (GFS) .....	296
	Best practices and additional installation information .....	296
	Heartbeat.....	299
	Heartbeat cluster components .....	299
	Installation information and additional details.....	300
	High Availability Extension (HAE).....	301
	HAE components .....	302
	Installation information and additional details.....	302
Chapter 12	Reference: Supported Linux features and limitations	
	Filesystems and feature limitations .....	304
	Filesystem support.....	304
	Features and limitations .....	305
	Linux volume managers .....	307
	LVM .....	307
	Veritas VxVM and VxFS .....	307
	EVMS.....	307
	LUN limits.....	308
	PATH limits .....	309
Chapter 13	Special Topics	
	Egenera .....	312

# PREFACE

*As part of an effort to improve and enhance the performance and capabilities of its product line, Dell EMC from time to time releases revisions of its hardware and software. Therefore, some functions described in this document may not be supported by all revisions of the software or hardware currently in use. For the most up-to-date information on product features, refer to your product release notes.*

*If a product does not function properly or does not function as described in this document, please contact your Dell EMC representative.*

This guide describes the features and setup procedures for Linux host interfaces to Dell EMC VMAX™ series, EMC VNX™ series, EMC VNXe™ series, Dell EMC Unity™ series, Dell EMC XtremIO™, Dell EMC VPLEX™, 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 VMAX series, and VNX series, VNXe series, Unity series, XtremIO, and host systems.

Readers of this guide are expected to be familiar with the following topics:

- ☒ VMAX series, VNX series, VNXe series, Unity series, XtremIO, and VPLEX system operation
- ☒ Linux operating environment

Any general reference to the VMAX series includes the VMAX3 family, VMAX family, and Symmetrix family.

- The VMAX3 family includes VMAX 400K/200K/100K and VMAX All Flash.
- The VMAX family includes VMAX 40K, 20K/VMAX, VMAX 10K(SN xxx987xxx)/VMAX 10K(SN xxx959xxx), and VMAXe.
- The Symmetrix family includes DMX-4/DMX-3.

**Related documentation** For the most up-to-date information, always consult the [Dell EMC Simple Support Matrix \(ESM\)](#), available through E-Lab Interoperability Navigator (ELN).

For documentation, refer to [Dell EMC Online Support](#).

**Conventions used in this guide** Dell EMC uses the following conventions for notes and cautions.

---

**Note:** A note presents information that is important, but not hazard-related.

---

## **IMPORTANT**

---

An important notice contains information essential to software or hardware operation.

---

## Typographical Conventions

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

Normal font	In running text: <ul style="list-style-type: none"><li>• Interface elements (for example, button names, dialog box names) outside of procedures</li><li>• Items that user selects outside of procedures</li><li>• Java classes and interface names</li><li>• Names of resources, attributes, pools, Boolean expressions, buttons, DQL statements, keywords, clauses, environment variables, filenames, functions, menu names, utilities</li><li>• Pathnames, URLs, filenames, directory names, computer names, links, groups, service keys, file systems, environment variables (for example, command line and text), notifications</li></ul>
<b>Bold</b>	In procedures: <ul style="list-style-type: none"><li>• Names of dialog boxes, buttons, icons, menus, fields</li><li>• Selections from the user interface, including menu items and field entries</li><li>• Key names</li><li>• Window names</li></ul> In running text: <ul style="list-style-type: none"><li>• Command names, daemons, options, programs, processes, notifications, system calls, man pages, services, applications, utilities, kernels</li></ul>
<i>Italic</i>	Used for: <ul style="list-style-type: none"><li>• Full publications titles referenced in text</li><li>• Unique word usage in text</li></ul>
<b><i>Bold Italic</i></b>	Anything requiring extra emphasis
Courier	Used for: <ul style="list-style-type: none"><li>• System output</li><li>• Filenames</li><li>• Complete paths</li><li>• Command-line entries</li><li>• URLs</li></ul>
<b>Courier, bold</b>	Used for: <ul style="list-style-type: none"><li>• User entry</li><li>• Options in command-line syntax</li></ul>
<i>Courier, italic</i>	Used for: <ul style="list-style-type: none"><li>• Arguments used in examples of command-line syntax</li><li>• Variables in examples of screen or file output</li><li>• Variables in path names</li></ul>
<b><i>Courier, bold, italic</i></b>	Variables used in a command-line sample
<>	Angle brackets enclose parameter or variable values supplied by the user
[ ]	Square brackets enclose optional values
	Vertical bar indicates alternate selections - the bar means “or”
{ }	Braces indicate content that you must specify (that is, 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.

Dell EMC support, product, and licensing information can be obtained on the Dell EMC Online Support site as described next.

---

**Note:** To open a service request through the Dell EMC Online Support site, you must have a valid support agreement. Contact your Dell EMC sales representative for details about obtaining a valid support agreement or to answer any questions about your account.

---

### Product information

For documentation, release notes, software updates, or for information about Dell EMC products, licensing, and service, go to [Dell EMC Online Support](#) (registration required).

### Technical support

Dell EMC offers a variety of support options.

**Support by Product** — Dell EMC offers consolidated, product-specific information on the Web at [Dell EMC Online Support](#).

The Support by Product web pages offer quick links to Documentation, White Papers, Advisories (such as frequently used Knowledgebase articles), and Downloads, as well as more dynamic content, such as presentations, discussion, relevant Customer Support Forum entries, and a link to Dell EMC Live Chat.

**Dell EMC Live Chat** — Open a Chat or instant message session with an Dell EMC Support Engineer.

### eLicensing support

To activate your entitlements and obtain your Symmetrix license files, visit the Service Center on [Dell EMC Online Support](#), as directed on your License Authorization Code (LAC) letter e-mailed to you.

For help with missing or incorrect entitlements after activation (that is, expected functionality remains unavailable because it is not licensed), contact your Dell EMC Account Representative or Authorized Reseller.

For help with any errors applying license files through Solutions Enabler, contact the Dell EMC Customer Support Center.

If you are missing a LAC letter, or require further instructions on activating your licenses through the Online Support site, contact Dell EMC's worldwide Licensing team at [licensing@emc.com](mailto:licensing@emc.com) or call:

- ☒ North America, Latin America, APJK, Australia, New Zealand: SVC4EMC (800-782-4362) and follow the voice prompts.
- ☒ EMEA: +353 (0) 21 4879862 and follow the voice prompts.

**We'd like to hear from you!**

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@emc.com](mailto:techpubcomments@emc.com)



# CHAPTER 1

## Introduction

This chapter provides an overview of the following:

☒ Operating system limits and guidelines .....	14
☒ Devices and operations .....	22
☒ LUN scanning mechanisms .....	26
☒ Persistent binding .....	30
☒ Mitigating the effects of storage array migration for Linux hosts.....	32
☒ Useful utilities .....	34
☒ Disk partition adjustment for VMAX series, VNX series, VNXe series, Unity series, CLARiiON, or XtremIO 36	
☒ Operating systems.....	44
☒ Host software .....	45
☒ Server vendors .....	48
☒ Host bus adapters .....	49
☒ Converged Network Adapters.....	50
☒ Dell EMC storage .....	51

## Operating system limits and guidelines

This section provides operating system limits and restrictions imposed in a SAN environment. Factors such as number of supported host bus adapters, logical units (LUNs), scalability of targets, file system, and volume management limits are detailed. The following areas are discussed:

- ☒ “Host initiators”
- ☒ “Logical unit”
- ☒ “Configuration example” on page 16
- ☒ “Storage attach” on page 19
- ☒ “Zoning recommendation” on page 21

### Host initiators

On all Linux environments, Dell EMC supports up to 16 Fibre Channel initiator ports on a single host. The host initiators may be single or dual channel 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.

---

#### Notes:

- ☒ Dell EMC does not support the mixing of HBAs from different vendors.
  - ☒ 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 is dependent on the SCSI scan algorithm employed by the operating system and the LUN scan limits imposed by the host bus adapter.

The HBA initiator and host system limits are theoretical maximums. [Table 1](#) illustrates these limits.

**Table 1** Maximum SCSI devices

Operating system	Per initiator devices	Host system devices	Dell EMC supported
Asianux 3.0	Emulex: 65536 QLogic: 65536	65536	1024
Asianux 4.0	Emulex: 65536 QLogic: 65536	65536	8192
OEL 5.0 <sup>1</sup>	Emulex: 65536 QLogic: 65536	65536	1024
OEL 6.0	Emulex: 65536 QLogic: 65536	65536	8192
OEL 7.0	Emulex: 65536 QLogic: 65536	65536	8192
RHEL 5 <sup>1</sup>	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	1024

**Table 1** Maximum SCSI devices

Operating system	Per initiator devices	Host system devices	Dell EMC supported
RHEL 6	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	8192
RHEL 7	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	16384
SLES 10 <sup>2</sup>	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	1024
SLES 11	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	8192
SLES 12	Emulex: 65536 QLogic: 65536 Brocade: 256	65536	16384

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

**Notes** A related limitation is the highest LUN instance or device number that a host system can address. This number is dependent on the ability of the HBA to address high LUN numbers as well as the total number of SCSI devices that have been exposed to the host system.

The HBA addressable LUN ID is the theoretical maximum. [Table 2](#) illustrates these limits and the supported limits for Dell EMC storage attach.

**Table 2** Highest addressable LUN ID (page 1 of 2)

Operating system	HBA addressable LUN ID	Dell EMC supported
Asianux 3.0	256 - Default; 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
Asianux 4.0	256 - Default; 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
OEL 5.0	256 – Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
OEL 6.0	256 – Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)
OEL 7.0	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)

**Table 2** Highest addressable LUN ID (page 2 of 2)

Operating system	HBA addressable LUN ID	Dell EMC supported
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	256 – Default, 32768 - Max (Emulex) 65536 (QLogic)	16384 (Emulex) 16384 (QLogic)

## Configuration example

The following hardware requirements are needed. The server:

- ☒ 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.

---

**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.

This process is different for RHEL 6 and above, which is explained in [“Multiple LUNs for RHEL 6.0 and later”](#) on page 18.

---

For PowerPath, skip [Step 1](#) and proceed to [Step 2](#). There are no changes needed when using PowerPath.

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.

```

172.23.224.69 - PuTTY
###
### This is a template multipath-tools configuration file
### Uncomment the lines relevant to your environment
###
defaults {
#   udev_dir                /dev
#   polling_interval        10
#   selector                "round-robin 0"
#   path_grouping_policy    multibus
#   getuid_callout          "/sbin/scsi_id -g -u -s /block/%n"
#   prio_callout            /bin/true
#   path_checker            readsector0
#   rr_min_io               100
#   max_fds                 8192
#   rr_weight               priorities
#   failback                immediate
#   no_path_retry           fail
#   user_friendly_names     yes
#   flush_on_last_del       no
#   queue_without_daemon    no
#   mode                    0666
#   uid                     0
#   gid                     0
}
blacklist {
#   wwid 35000cca0005c2250
#   wwid 35000cca00074a158
#   devnode "^ (ram|raw|loop|fd|md|dm-|sr|scd|st) [0-9] *"
#   devnode "^ hd[a-z] [[0-9] *]"
#   devnode "^ hd[a-z]"
#   devnode "^ cciss!c[0-9]d[0-9] *"
}

```

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.

```

root@localhost:~/usr/sbin/hbanyware
[root@localhost hbanyware]# more /etc/modprobe.conf
alias eth0 bnx2
alias eth1 bnx2
alias eth2 bnx2
alias eth3 bnx2
options scsi_mod max_luns=8191 max_report_luns=8192
alias scsi_hostadapter mptbase
alias scsi_hostadapter1 mptsas
alias scsi_hostadapter2 lpfc
alias scsi_hostadapter3 usb-storage
# Emulex lpfc options
options lpfc lpfc_max_luns=8192
###BEGINPP
include /etc/modprobe.conf.pp
###ENDPP
[root@localhost hbanyware]#

```

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 affect.

## 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 prior versions. Therefore, module options cannot be changed in RHEL 6 by adding a `.conf` file entry within the `/etc/modprobe.d` directory. Settings should go on the kernel command line.

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:

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. If so, 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

1. Modify /etc/default/grub; Add the highlighted text.:

```
[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. 4. 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
```

(No modification in modprobe for QLogic 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
```

5. The configuration changes above must be compiled into the initrd by executing the following:

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

## Storage attach

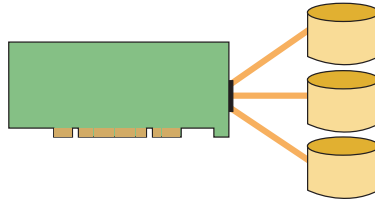
This section discusses fan-in and fan-out.

**Fan-in** With Dell EMC VNX™ series and Dell EMC CLARiiON™ systems, Dell EMC supports the scanning of a maximum of 4 VNX series and CLARiiON systems or 32 VNX series and CLARiiON SP ports (whichever is lesser) per host initiator port.

With the Unity series and VNXe series, Dell EMC supports the 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 VMAX series does 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.

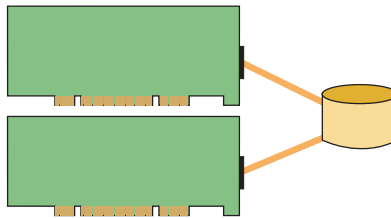
Figure 1 shows an example of fan-in.



**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 will attempt to scan. On Emulex, this limit is set to 256 and on QLogic this limit is set to 512. Again, these are theoretical limits as exported by the host bus adapter.

Figure 2 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 is dependent on the number of SCSI devices and targets exposed. With a large number of SCSI devices, the boot process will be 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, refer to the *Dell EMC XtremIO Storage Array Host Configuration Guide* on Dell EMC Online Support.

### **IMPORTANT**

---

A single zone should *not* contain multiple initiator ports.

Multiple target ports from multiple arrays are supported in a single zone.

## Devices and operations

This section provides an overview of mechanisms provided by a Linux operating system for addressing and utilizing SCSI devices, included in the following sections:

- ☒ “SCSI device addressing”, next
- ☒ “SCSI device operation interfaces” on page 23

### 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` filesystem 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 or CLARiiON:

```
# 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 a 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.

---

**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, refer to [Linux 2.4 SCSI subsystem How To](#).

## SCSI device operation interfaces

Linux provides various device operation interfaces. This includes block and character devices as well as raw device interfaces. In addition, with the 2.6 kernel a new framework for device management, the *device-mapper*, was introduced. This section expands on these concepts.

### 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 physical drive

*n* is a number denoting the partition on that physical 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.

## Red Hat RHEL implementation

The raw interface is available on RHEL 5.

---

**Note:** Although RHEL includes support for rawio, it is now a deprecated interface. Dell EMC recommends 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 populated 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 mechanisms

Linux provides multiple mechanisms to rescan the SCSI bus and recognize SCSI devices exposed to the system. With the 2.6 kernel and later, significant improvements have been made and dynamic LUN scanning mechanisms are available. Linux currently lacks a kernel command that allows for a dynamic SCSI channel reconfiguration like `drvconfig` or `ioscan`.

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.

### **IMPORTANT**

Dell EMC recommends 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`

In the 2.4 kernel, the `/proc` file system provides a listing of available SCSI devices. If SCSI devices 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 from the Dell EMC-approved site on the QLogic website.

### Usage examples

- ☒ To re-scan 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 re-scan 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 from the Dell EMC-approved site 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
```

---

**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

Novell's 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.

```
182bi094:~ # 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:

```
182bi094:~ # /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.)
182bi094:~ #

```

---

**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 explores the persistent binding features available on Linux:

- ☒ [“HBA persistent binding” on page 30](#)
- ☒ [“Udev” on page 31](#)
- ☒ [“Native MPIO” on page 31](#)
- ☒ [“PowerPath pseudo-names” on page 31](#)
- ☒ [“Logical volumes” on page 31](#)

## HBA persistent binding

In Emulex and QLogic drivers available for the 2.4 kernel, 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. Refer to the current Dell EMC HBA documentation provided on the Dell EMC-approved web page of Emulex (now Broadcom) or QLogic for information on how to configure the appropriate drivers.

- ☒ *EMC Installation and Configuration Guide for Emulex HBAs and the Linux 2.4 Kernel*
- ☒ *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 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 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 meta-data 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

This section provides information on 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 will load 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 filesystems 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 filesystems using mount by label should not be adversely affected by having migrated to a new storage array. The filesystem label will be copied to the new target LUN and mount will be able to 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. Therefore, 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 filesystems 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.

We recommend 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 on 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. Refer to PowerPath Migration Enabler (PPME) and Open Replicator (OR) documentation available on [Dell EMC Online Support](#), for additional information regarding implementation and usage.

## Useful utilities

Table 3 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 3** Useful system utilities on Linux (page 1 of 2)

Command name	Purpose [From the respective 'man' pages]
<b>Create partitions, file systems, mount file system, and monitor IO status</b>	
fdisk	Command used to create and manipulate partition tables.
parted	a partition manipulation program.
mkfs	Command used to create a Linux filesystem on a device partition.
fsck	Command used to check and repair a Linux filesystem.
mount	Command used to attach the filesystem on a device to the file tree.
umount	Command used to detach a filesystem.
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.
<b>LVM command</b>	
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.
<b>Multipath command</b>	
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.
<b>Driver module utility</b>	

**Table 3** Useful system utilities on Linux (page 2 of 2)

<b>Command name</b>	<b>Purpose [From the respective ‘man’ pages]</b>
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.
<b>udev utility</b>	
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.
<b>iSCSI utility command</b>	
iscsiadm	Open-iscsi administration utility.
iscsi-ls	List iscsi device utility.
<b>Other utility</b>	
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 VMAX series, VNX series, VNXe series, Unity series, CLARiiON, or XtremIO

This section discusses the basic theory of alignment providing the reader with an understanding of how aligning the data with the physical layout of Dell EMC storage may benefit overall system performance. For specific information on methods of how alignment may be performed on different filesystems and volume managers, the reader is encouraged to contact the operating system vendor.

To maximize disk performance, any I/O to a VMAX series, VNX series, VNXe series, Unity series, CLARiiON, 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 series that is discussed briefly in this section.

- ☒ Cache Slot Boundaries (one Track [64 Blocks] - 64 KB)
- ☒ RAID 5 Boundaries (four Tracks [256 Blocks] - 256 KB)
- ☒ Metastripe Boundaries (two Cylinders [3840 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.

Owing to the legacy of the IBM PC BIOS, 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 far more 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, use the default value (2048), but 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 an VMAX array, a track is considered to be 64 blocks and a VMAX cache memory slot is based on this track size and offset. On VNX series or CLARiiON storage, the unit of allocation is an element, which is (by default) 128 KB.

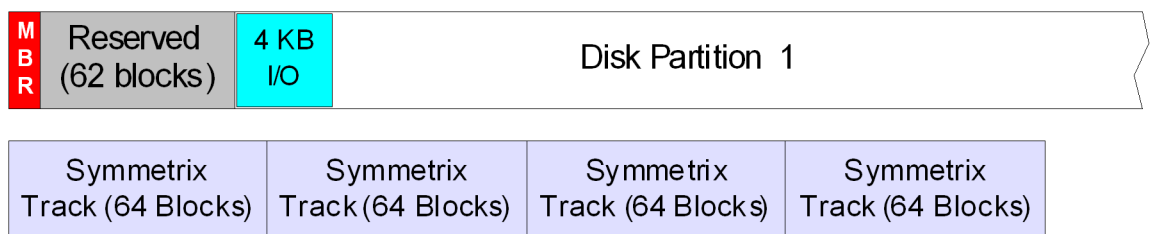


## Track boundaries

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:

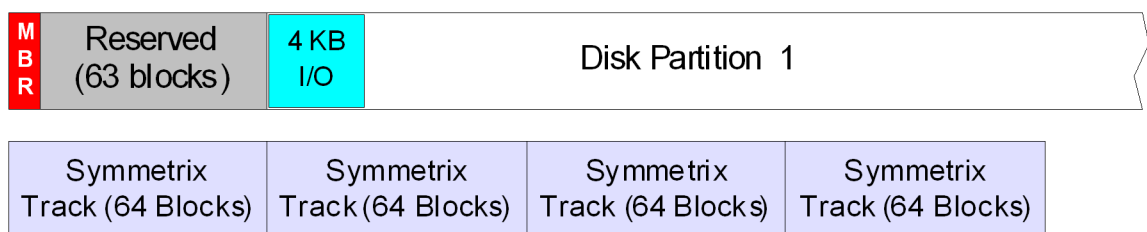
- ☒ Any I/O of 32 KB or larger will always cause a boundary crossing.
- ☒ Any random I/O of 16 KB will cause a boundary crossing 50 percent of the time.
- ☒ Any random I/O of 8 KB will cause a boundary crossing 25 percent of the time.
- ☒ Any random I/O of 4 KB will cause a boundary crossing 12.5 percent of the time.



**Figure 3** Misaligned partition (not to scale)

As [Figure 3](#) 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)

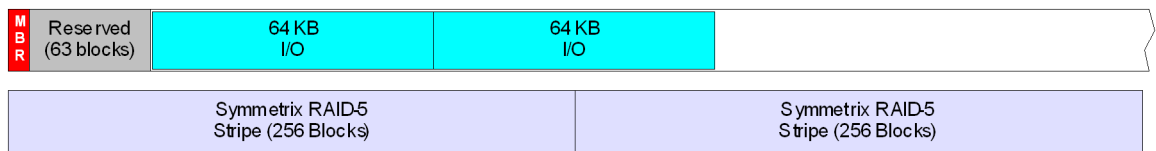
As [Figure 4](#) shows, starting the first partition on block address 64 will align the I/Os and will not cause boundary crossings.

## RAID 5 boundaries

If you now consider RAID 5 volumes, things change. There are two differences to consider:

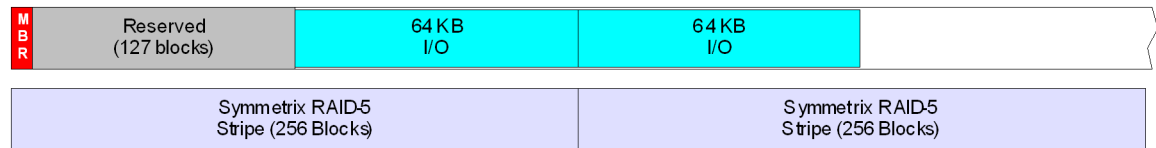
- ☒ In Linux kernels prior to 2.6.18, the maximum I/O size that Linux will issue is 64 KB. Larger I/Os are broken up into 64 KB chunks. In Linux kernels 2.6.18 and higher, this is handled by the `/sys/block/<device>/queue/max_sectors_kb` parameter (which is less than the `max_hw_sectors_kb` setting for 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)

As [Figure 5 on page 38](#) 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 [Figure 6](#) shows, starting the first partition on block address 64 will not cause stripe boundary crossings for 64 KB I/Os.

## Metastripe boundaries

Metastripes 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 acknowledgement 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, Unity series, or CLARiiON systems

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

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

## Determining the correct offset to partition

You must first determine where to start the partition:

- ☒ For a VMAX array, the start address is at block 128.
- ☒ For a VNX series or CLARiiON 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 8KB).

To determine the element size, follow these steps:

1. Start **Unisphere/Navisphere**.
2. Navigate to the appropriate storage group and LUN to be used.

For example, suppose you want to create an aligned partition on LUN 2 on the lab-w2k host. Note that 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**.

[Figure 7 on page 40](#) shows the **LUN Properties** dialog box.

4. Ensure that the Alignment Offset value is 0. If not, the VNX series or CLARiiON LUN has been deliberately misaligned, possibly to correct the partition misalignment that you intend to correct. If this is the case, 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.

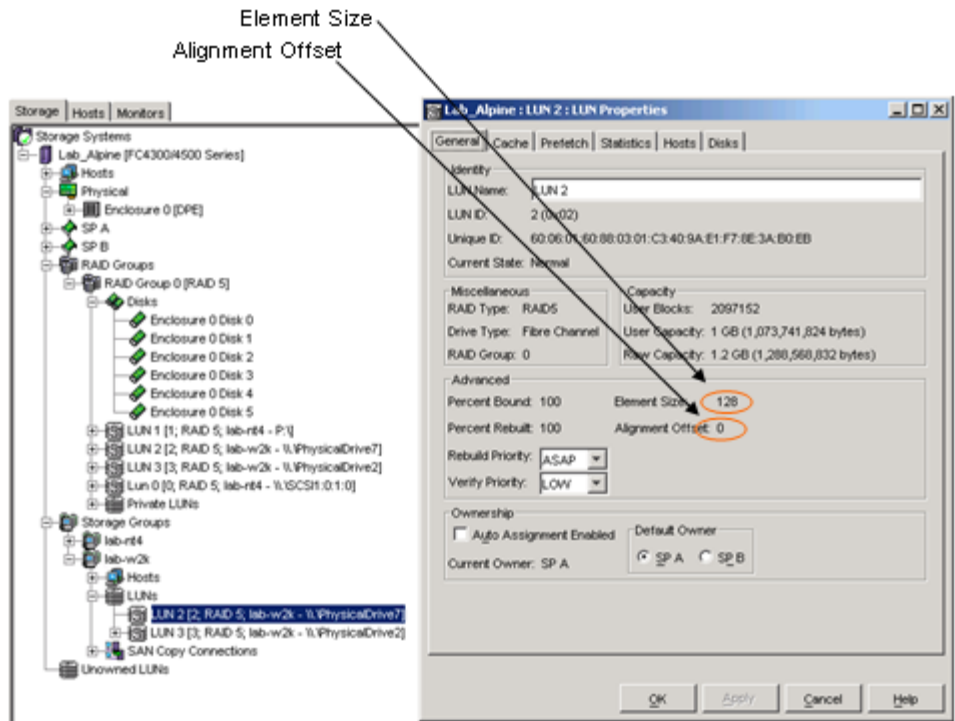


Figure 7 LUN properties

## 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.

In order 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 proposed by Intel as a replacement for the outdated PC BIOS, one of the few remaining relics of the original IBM PC.

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

### IMPORTANT

You must include GPT support in kernel in order 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.

#### 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) unit TB
(parted) mkpart primary 0 0
(parted) print
(parted) quit
```

```
[root@lin105100 Hellya]# parted /dev/emcpowerj
GNU Parted 3.1
Using /dev/emcpowerj
Welcome to GNU Parted! Type 'help' to view a list of commands.
(parted) mklabel gpt
(parted) unit TB
(parted) mkpart primary 0.00TB 3TB
(parted) p
Model: Unknown (unknown)
Disk /dev/emcpowerj: 3.30TB
Sector size (logical/physical): 4096B/4096B
Partition Table: gpt
Disk Flags:

Number  Start   End     Size    File system  Name      Flags
  1      0.00TB  3.30TB  3.30TB                primary

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

[root@lin105100 Hellya]# partprobe /dev/emcpowerj
[root@lin105100 Hellya]# parted /dev/emcpowerj print
Model: Unknown (unknown)
Disk /dev/emcpowerj: 3299GB
Sector size (logical/physical): 4096B/4096B
Partition Table: gpt
Disk Flags:

Number  Start   End     Size    File system  Name      Flags
  1      1049kB  3299GB  3299GB                primary
```

## 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)
- ☒ Novell 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 Novell's SuSE Linux and RedHat Linux. Dell EMC does *not* support Xen implementations from XenSource directly.

Refer to the [Dell EMC Simple Support Matrix](#) Linux OS footnotes in the base connectivity section for supported configurations.

For more information on Xen, refer to the [Micro Focus \(Novell\)](#) 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” on page 45
- ☒ “Navisphere CLI” on page 45
- ☒ “Dell EMC replication software” on page 47

### 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 series or CLARiON 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 an Dell EMC CX-Series 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 navisecli command is used as follows:

```
navisecli -help  
or  
navisecli
```

```
[-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, refer to 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). Specifically, 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. Refer the [Dell EMC Simple Support Matrix](#) for the most current support information for the OS 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 vendors

Dell EMC supports Intel, AMD, and PowerPC architecture-based servers from a range of server vendors.

On the Intel platform 32-bit, EM64T and IA-64 IPF processor-based servers are supported. On the AMD platform, the 64-bit Opteron processor based servers are supported. Both stand-alone rack/tower and blade servers are supported. For more detailed information, please refer to O/S vendor.

---

**Note:** Dell EMC provides support for the IBM PowerPC architecture. The IBM P6 and P7 series standard LPAR and VIO clients are supported for VMAX3, VMAX, Symmetrix, XtremIO, VNX series, VNXe series, Unity series, and CLARiiON CX family attach. Linux native DM-MPIO and PowerPath are supported on Linux LPARs.

Support for booting Linux LPARS from the SAN is available. Consult 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, refer to the vendor's website.

## Host bus adapters

QLogic, Emulex, and Brocade Fibre Channel host bus adapters are supported on Dell EMC storage. Refer to the Linux "Base Connectivity" section of the [Dell EMC Simple Support Matrix](#) for supported HBA models and refer to the appropriate installation guide, available at [Dell EMC Online Support](#), for configuring the host adapter and driver for the system.

Both the QLogic iSCSI hardware initiator and the generic NIC iSCSI software initiator are supported with Dell EMC iSCSI storage arrays. Refer to the Linux "iSCSI Connectivity" section of the [Dell EMC Simple Support Matrix](#) for supported configurations and required driver revision levels. An Dell EMC-published QLogic iSCSI guide is available on the QLogic website. [Chapter 4, "iSCSI Connectivity,"](#) provides more information.

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 will function 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](#).

## Dell EMC storage

Basic Dell EMC storage information is provided in this section for the following:

- ☒ “VMAX series” on page 51
- ☒ “Unity series” on page 52
- ☒ “VNX series or CLARiiON” on page 52
- ☒ “VPLEX” on page 52
- ☒ “XtremIO” on page 53
- ☒ “ScaleIO” on page 53
- ☒ “XtremCache” on page 55

### VMAX series

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

**Applications** Information about VMAX applications can be found at the [Dell EMC VMAX All Flash Storage page](#) on EMC.com.

**Offerings** [Table 4](#) lists the 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](#).

Refer to the Base Connectivity and the iSCSI Connectivity tables in the [Dell EMC Simple Support Matrix](#) for supported HBAs, operating system revisions, and servers.

**Table 4** Supported VMAX arrays (page 1 of 2)

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

**Table 4** Supported VMAX arrays (page 2 of 2)

Storage arrays	Array code requirements
VMAXe [Fibre Channel and iSCSI]	Engenuity 5875 /5876
Symmetrix DMX-4 [Fibre Channel and iSCSI]	Engenuity 5772 /5773
Symmetrix DMX-3 [Fibre Channel and iSCSI]	Engenuity 5771/5772

**VMAX series array-specific settings**

When attaching a Linux host to a 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**

VMAX documentation is available at [Dell EMC Online Support](#).

## Unity series

This section contains information on Unity series applications, offerings, and documentation:

- ☒ Refer to information about Unity applications on the [Unity All Flash Storage page](#) on EMC.com and on Dell EMC Online Support.
- ☒ The Unity series includes All-Flash storage arrays (Unity 300F, 400F, 500F, 600F) or Hybrid Flash storage arrays (Unity 300, 400, 500, 600).

## VNX series or CLARiiON

This section contains information on VNX series or CLARiiON applications, offerings, array-specific settings, and documentation.

**Applications**

Information about VNX application can be found on the [EMC VNX Hybrid Flash Storage Family page](#) on EMC.com.

**VNX series or CLARiiON system-specific settings**

When attaching a Linux host to a VNX series or CLARiiON 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, consult [Table 20 on page 166](#).
- ☒ 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. For more information refer to [Chapter 10, "VPLEX."](#)

**Documentation**

Refer to ["VPLEX documentation"](#) on [page 252](#) for a list of VPLEX documentation.



## 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 very 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 throughput 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-Bricks. Performance scales linearly, ensuring that two X-Bricks 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. Please consult the versions of Dell EMC PowerPath and Veritas DMP supported in the [Dell EMC Simple Support Matrix](#).

## ScaleIO

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

### Applications

Find information about ScaleIO applications on the [Dell EMC ScaleIO page](#) on EMC.com.

## Implementation

Implementing a ScaleIO 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 ScaleIO software installed on them.

To implement the physical layer, perform the following steps:

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 ScaleIO 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 will contribute some or all of their physical storage:
  - Starting with ScaleIO 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 will access the data exposed by the ScaleIO 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:

- ☒ ScaleIO Installation Manager
- ☒ ScaleIO 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 a 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, ScaleIO 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** Refer to the *Dell EMC Simple Support Matrix - ScaleIO Node* and *EMC Simple Support Matrix - ScaleIO 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 EMC XtremCache™ /SF applications, offerings, installation/management /configuration processes, supports for Linux, and documentation.

**Applications** Find information about XtremCache applications on the [Dell EMC ScaleIO page](#) on EMC.com.

- 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:  
**vfcm add -cache\_dev <device>**
  - ☒ Enable and start a source device:  
**vfcm add -source\_dev <device>**
  - ☒ You can review the configuration by typing the following command:  
**vfcm 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** Refer to the *Dell EMC Simple Support Matrix - EMC XtremCache*, located at [Dell EMC E-Lab Interoperability Navigator](#).

**Documentation** XtremCache documentation can be found at [Dell EMC Online Support](#).

# CHAPTER 2

## Fibre Channel Connectivity

This chapter provides information on connectivity, including the following:

- ☒ Introduction..... 58
- ☒ Configuring the HBAs for a Linux host ..... 59
- ☒ Hitachi Virtage..... 72

## Introduction

Fibre Channel captures some of the benefits of both channels and networks. A Fibre Channel fabric is a switched network, providing a set of 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 an Dell EMC-approved Emulex, QLogic, and Brocade adapter into a Linux host environment and configuring the host for connection to an Dell EMC storage array over Fibre Channel (FC).

This section contains the following information:

- ☒ “Prerequisites for first-time installation” on page 59
- ☒ “Emulex Fibre Channel HBA” on page 59
- ☒ “QLogic Fibre Channel HBA” on page 63
- ☒ “Brocade Fibre Channel HBA” on page 68

### Prerequisites for first-time installation

---

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

---

- ☒ Review the [Dell EMC Simple Support Matrix](#) or contact your Dell EMC representative for the latest information on qualified adapters, drivers, and Linux distributions.
- ☒ Refer to the vendor's Fibre Channel Host Adapter (HBA) product documentation to properly install an HBA in your server.
- ☒ Refer to the vendor's product documentation to verify and update HBA firmware and boot BIOS to Dell EMC-qualified versions.

### 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. Refer to 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.

---

**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 OS and version.

---

**Note:** The support stated in the [Dell EMC Simple Support Matrix](#) supersedes versions listed in this document.

---

**Dell EMC-supported Emulex driver versions**

Table 5 lists the Emulex driver versions supported with the corresponding OS updates. These driver versions are included by default in the kernel and do not require any installation.

**Table 5** Dell EMC-supported Emulex Linux in-kernel drivers (page 1 of 4)

OS	In-kernel Driver Version	Supported adapter			
		1/2 Gb	4 Gb	8 Gb	16 Gb
SuSe SLES 10 GA	8.1.6	√	√		
Red Hat RHEL 5.0 Oracle OEL 5.0 SuSE SLES 10 SP1 (errata kernels 2.6.16.53-0.8, 2.6.16.53-0.16)	8.1.10.3	√	√		
Red Hat RHEL 5.1 Oracle OEL 5.1 SuSE SLES 10 SP1 (errata kernels 2.6.16.53-0.8, 2.6.16.53-0.16)	8.1.10.9	√	√		
SuSE SLES 10 SP1 (errata kernels equal to or greater than 2.6.16.54-0.2.3)	8.1.10.12-update	√	√		
Red Hat RHEL 5.2 SuSE SLES 10 SP2 Oracle OEL 5.2 Asianux 3.0 SP1	8.2.0.22	√	√	√	
RedHat RHEL 5.4 SuSE SLES 10 SP3	8.2.0.48.2p	√	√	√	
RHEL 5.3 OEL 5.3	8.2.0.33.3p	√	√	√	
RHEL 5.4 OEL 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)	√	√	√	
RHEL 5.5 OEL 5.5	8.2.0.63.3p	√	√	√	
RHEL 5.6	8.2.0.87.1p	√	√	√	
RHEL 5.7	8.2.0.96.2p	√	√	√	
RHEL 5.8	8.2.0.108.4p	√	√	√	
RHEL 5.9 <b>RHEL 5.10</b> <b>RHEL 5.11</b>	8.2.0.128.3p	√	√	√	√
	8.2.0.128.3p	√	√	√	√
RHEL 6.0	8.3.5.17	√	√	√	
RHEL 6.1	8.3.5.30.1p	√	√	√	



**Table 5** Dell EMC-supported Emulex Linux in-kernel drivers (page 2 of 4)

OS	In-kernel Driver Version	Supported adapter			
		1/2 Gb	4 Gb	8 Gb	16 Gb
RHEL 6.2	8.3.5.45.4p	√	√	√	
RHEL 6.3	8.3.5.68.5p	√	√	√	√
RHEL 6.4	8.3.5.86.1p	√	√	√	√
RHEL 6.5	8.3.7.21.4p	√	√	√	√
RHEL 6.6	10.2.8020.1	√	√	√	√
RHEL 6.7	10.6.0.20	√	√	√	√
RHEL 7.0	8.3.7.34.3p	√	√	√	√
RHEL 7.1	10.2.8021.1	√	√	√	√
RHEL 7.2	10.7.0.1	√	√	√	√
SLES 10 SP4	8.2.0.92.1p	√	√	√	√
SuSE SLES 11 GA	8.2.8.14	√	√	√	
SuSE SLES 11 SP1	8.3.5.8.1p 8.3.5.8.2p	√	√	√	√
SuSE SLES 11 SP2	8.3.5.48.2p 8.3.5.48.3p	√	√	√	√
SLES 11 SP3	8.3.7.10.6p 8.3.7.10.7p	√	√	√	√
SLES 11 SP4	10.4.8000.0	√	√	√	√
SLES 12	10.2.8040.1	√	√	√	√
Oracle Linux 5. x UEK R1 [2.6.32-100]	8.3.18	√	√	√	
Oracle Linux 6. x UEK R1 [2.6.32-100]	8.3.5.30.1p	√	√	√	
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	√	√	√	

**Table 5** Dell EMC-supported Emulex Linux in-kernel drivers (page 3 of 4)

OS	In-kernel Driver Version	Supported adapter			
		1/2 Gb	4 Gb	8 Gb	16 Gb
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]	8.3.5.45.4p	√	√	√	√
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	√	√	√	√
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	√	√	√	√
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	√	√	√	√
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	√	√	√	√
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	√	√	√	√
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	√	√	√	√
Oracle Linux 6.x UEK R3 [3.8.13-16]	8.3.7.26.2p	√	√	√	√

**Table 5** Dell EMC-supported Emulex Linux in-kernel drivers (page 4 of 4)

OS	In-kernel Driver Version	Supported adapter			
		1/2 Gb	4 Gb	8 Gb	16 Gb
Oracle Linux 6.x UEK R3 U1 [3.8.13-26]	8.3.7.34.4p	√	√	√	√
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]					
Oracle Linux 6.x UEK R3 U4 [3.8.13-55]	10.2.8061.0	√	√	√	√
Oracle Linux 7.x UEK R3 U4 [3.8.13-55]					
Oracle Linux 6.x UEK R3 U5 [3.8.13-68]	10.6.61.0	√	√	√	√
Oracle Linux 7.x UEK R3 U5 [3.8.13-68]					
Oracle Linux 6.x UEK R3 U6 [3.8.13-98]	10.6.61.1	√	√	√	√
Oracle Linux 7.x UEK R3 U6 [3.8.13-98]					

## 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. Refer to 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.

---

**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 OS and version.

**Note:** The support stated in the [Dell EMC Simple Support Matrix](#) supersedes versions listed in this document.

**Dell EMC-supported QLogic driver versions**

Table 6 lists the QLogic driver versions supported with the corresponding OS updates. These driver versions are included by default in the kernel and do not require any installation.

**Table 6** Dell EMC supported QLogic Linux in-kernel drivers (page 1 of 4)

OS	In-kernel Driver Version	Supported adapter			
		½ Gb	4 Gb	8 Gb	16 Gb
SLES 10 GA	8.01.04-k	√	√		
RHEL 5.0 OEL 5.0	8.01.07-k1	√	√		
SLES 10 SP1	8.01.07-k3	√	√		
RHEL 5.1 OEL 5.1	8.01.07-k7	√	√	√	
RHEL 5.2 OEL 5.2	8.02.00-k5-rhel5.2-03	√	√	√	
RHEL 5.2 (errata kernels equal to or greater than 2.6.18-92.1.6.el5) OEL 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	√	√	√	
SLES10 SP2	8.02.00-k6-SLES10-05	√	√	√	
RHEL 5.3 OEL 5.3	8.02.00.06.05.03-k	√	√	√	
SuSE SLES 11 GA	8.02.01.03.11.0-k9	√	√	√	
RHEL 5.4 OEL 5.4	8.03.00.10.05.04-k	√	√	√	
RHEL 5.4 (errata kernels equal to or greater than 2.6.18-164.2.1.el5) OEL 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	√	√	√	
SuSE SLES 10 SP3	8.03.00.06.10.3-k4	√	√	√	

Table 6 Dell EMC supported QLogic Linux in-kernel drivers (page 2 of 4)

OS	In-kernel Driver Version	Supported adapter			
		½ Gb	4 Gb	8 Gb	16 Gb
RHEL 5.5 OEL 5.5	8.03.01.04.05.05-k	√	√	√	
RHEL 5.6 OEL 5.6	8.03.01.05.05.06-k	√	√	√	
SLES 10 SP4	8.03.07.03.06.1-k	√	√	√	
SLES 11 SP1 (kernel < 2.6.32.13-0.4.1)	8.03.01.06.11.1-k8	√	√	√	
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	√	√	√	
SLES 11 SP1 (kernel > 2.6.32.27-0.2.2)	8.03.01.08.11.1-k8	√	√	√	
SLES 11 SP2	8.03.07.07-k	√	√	√	
SLES 11 SP3	8.04.00.13.11.3-k	√	√	√	√
SLES 11 SP4	8.07.00.18-k	√	√	√	√
SLES 12	8.07.00.08.12.0-k	√	√	√	√
RHEL 5.7	8.03.07.00.05.07-k	√	√	√	
RHEL 5.8	8.03.07.09.05.08-k	√	√	√	
RHEL 5.9 RHEL 5.10 RHEL 5.11	8.03.07.15.05.09-k	√	√	√	
RHEL 6.0	8.03.05.01.06.1-k0	√	√	√	
RHEL 6.1	8.03.07.03.06.1-k	√	√	√	
RHEL 6.2	8.03.07.05.06.2-k	√	√	√	
RHEL 6.3	8.04.00.04.06.3-k	√	√	√	√
RHEL 6.4	8.04.00.08.06.4-k	√	√	√	√
RHEL 6.5	8.05.00.03.06.5-k2	√	√	√	√
RHEL 6.6	8.07.00.08.06.6-k1	√	√	√	√
RHEL 6.7	8.07.00.16.06.7-k	√	√	√	√

**Table 6** Dell EMC supported QLogic Linux in-kernel drivers (page 3 of 4)

OS	In-kernel Driver Version	Supported adapter			
		½ Gb	4 Gb	8 Gb	16 Gb
RHEL 7.0	8.06.00.08.07.0-k2	√	√	√	√
RHEL 7.1	8.07.00.08.07.1-k2	√	√	√	√
RHEL 7.2	8.07.00.18.07.2-k	√	√	√	√
Oracle Linux 5. x UEK R1 [2.6.32-100]	8.03.01.02.32.1-k9	√	√	√	
Oracle Linux 6. x UEK R1 [2.6.32-100]	8.03.07.03.32.1-k	√	√	√	
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	√	√	√	
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	√	√	√	
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	√	√	√	
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	√	√	√	
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	√	√	√	
Oracle Linux 5.x UEK R2 U3 [2.6.39-400] Oracle Linux 6.x UEK R2 U3 [2.6.39-400]	8.04.00.11.39.0-k, 8.05.00.03.39.0-k	√	√	√	

Table 6 Dell EMC supported QLogic Linux in-kernel drivers (page 4 of 4)

OS	In-kernel Driver Version	Supported adapter			
		½ Gb	4 Gb	8 Gb	16 Gb
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	√	√	√	
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	√	√	√	√
Oracle Linux 6.x UEK R3 [3.8.13-16]	8.05.00.03.39.0-k	√	√	√	√
Oracle Linux 6.x UEK R3 U1 [3.8.13-26]	8.06.00.14.39.0-k	√	√	√	√
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	√	√	√	√
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	√	√	√	√
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	√	√	√	√

## Brocade Fibre Channel HBA

Brocade and QLogic have announced and signed an agreement to sell the Brocade Adapter business to QLogic effective January 17, 2014. The product portfolio includes Fibre Channel Host Bus Adapters (HBAs), Converged Network Adapters (CNAs), and mezzanine adapters for OEM blade server platforms. 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 listed in [Table 7](#) are included by default in the kernel and do **not** require installation.

### Dell EMC-supported Brocade driver versions

[Table 7](#) lists the Brocade driver versions supported with the corresponding OS updates.

**Table 7** Dell EMC-supported Brocade Linux in-kernel drivers

OS	In-kernel Driver Version	Supported adapter		
		4 Gb	8 Gb	16 Gb
RHEL 5.5	2.1.2.0	√	√	
RHEL 5.6	2.1.2.2	√	√	
RHEL 5.7	2.3.2.3	√	√	
RHEL 5.8	3.0.2.2	√	√	
RHEL 5.9	3.0.23.0			
RHEL 5.10		√	√	√
RHEL 5.11				
RHEL 6.0	2.1.2.1	√	√	
RHEL 6.1	2.3.2.3	√	√	
RHEL 6.2	3.0.2.2 (BFA) 3.0.2.2r (BNA)	√	√	
RHEL 6.3	3.0.2.2	√	√	√
RHEL 6.4	3.0.23.0	√	√	√
RHEL 6.5	3.2.21.1	√	√	√



**Table 7** Dell EMC-supported Brocade Linux in-kernel drivers

OS	In-kernel Driver Version	Supported adapter		
		4 Gb	8 Gb	16 Gb
RHEL 6.6	3.2.23.0	√	√	√
RHEL 6.7				
RHEL 7.0				
RHEL 7.1				
SLES 10 SP3	2.0.0.0	√	√	
SLES 10 SP4	2.3.2.1	√	√	
SLES 11	1.1.0.2	√	√	
SLES 11 SP1	2.1.2.1	√	√	
SLES 11 SP2	3.0.2.2	√	√	
SLES 11 SP3	3.1.2.1	√	√	√
SLES 11 SP4	3.2.23.0	√	√	√
SLES 12				

## 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 inter-operable 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 won't 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 above 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:

1. Use 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](http://driverdownloads.qlogic.com/QLogicDriverDownloads_UI/DefaultNewSearch.aspx)

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

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

5. Unzip 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 doesn't exist) and will have the following entries:

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

**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 SNIA API libraries for Emulex HBA:

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. Use 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 OS (**elxocmcore-rhel5-rhel6-rhel7-11.0.243.13-1.tgz**).

4. Unzip 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 doesn't 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 built-in (firmware) feature on Hitachi server products, including the BladeSymphony 2000, the BladeSymphony 500, and BladeSymphony 320. This feature requires no separate OS 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 (LPAR), Hitachi also offers Fibre Channel I/O virtualization for Hitachi Virtage. This allows multiple logical partitions (LPARs) 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, refer to the [Hitachi website](#).

The Fibre Channel drivers used for logical partitions (LPARs) 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 (LPAR), 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 above step.
3. Reboot the OS 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. Refer to <http://support.emc.com/kb/209658> for details.

---



# CHAPTER 3

## Fibre Channel over Ethernet Connectivity

This chapter provides information on Fibre Channel over Ethernet (FCoE) connectivity, including the following:

☒ Introduction.....	76
☒ Configuring the Linux host.....	78
☒ Cisco Unified Computing System .....	90

# Introduction

I/O consolidation has been long sought by the IT industry to unify the multiple transport protocols in the data center. 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.

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

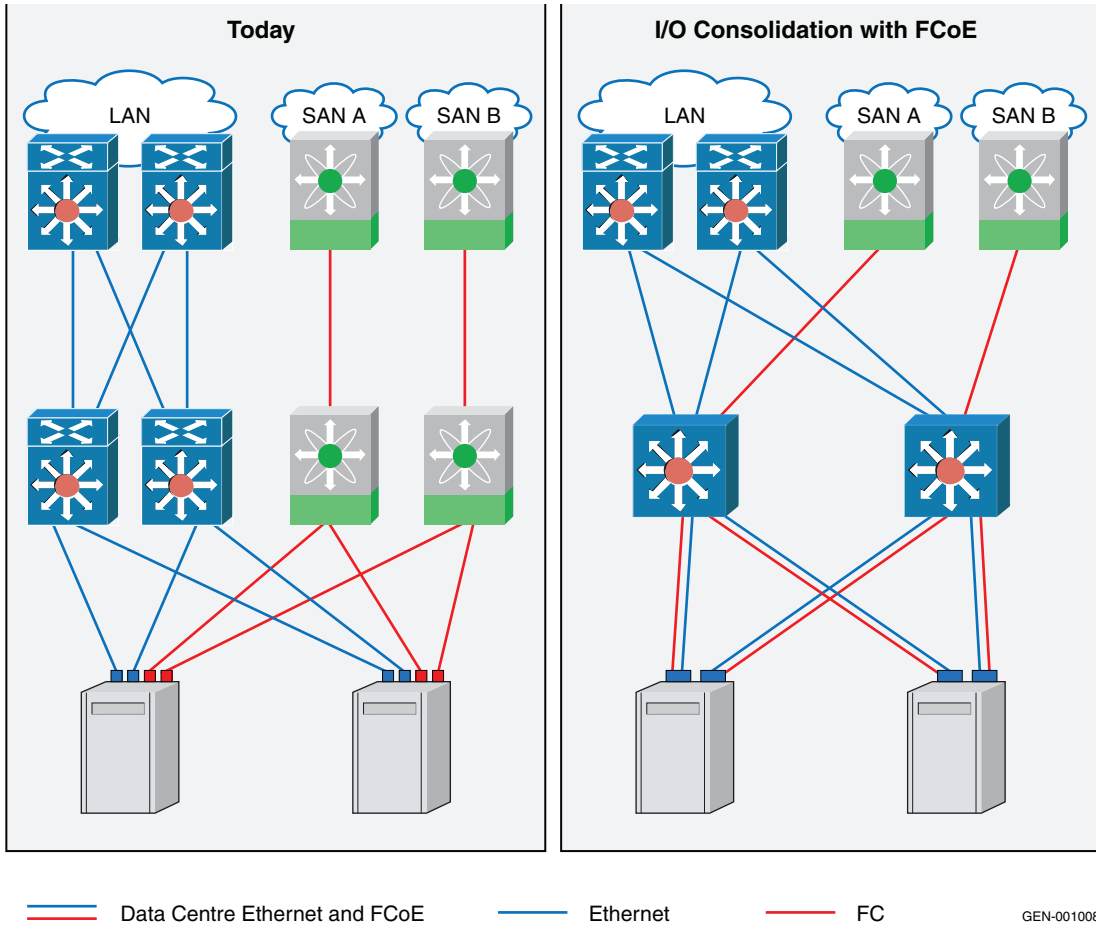
- ▣ *Fibre Channel over Ethernet - A review of FCoE Today*
- ▣ Silvano, Gai, *Data Center Network and Fibre Channel over Ethernet*, Nuova Systems Inc., 2008

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 [Figure 8 on page 77](#).

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).





**Figure 8** Typical topology versus FCoE example using NEX-5020

For more information on Fibre Channel over Ethernet, refer to 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.

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

### Emulex

Using the Emulex Covered 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 8](#) lists the Emulex driver versions supported with the corresponding OS updates. [Table 9 on page 81](#) lists the Emulex out-of-kernel driver versions supported with the corresponding OS updates.

Refer to the latest [Dell EMC Simple Support Matrix](#) for your specific Linux distribution, kernel version, and driver to determine whether or not 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 8** Supported FCOE in-kernel drivers (page 1 of 3)

OS	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 higher)	√ <sup>1</sup>
RHEL 5.5	8.2.0.63.3p	√ <sup>1</sup>
RHEL 5.6	8.2.0.87.1p	√ <sup>1</sup>
RHEL 5.7	8.2.0.96.2p	√ <sup>1</sup>
RHEL 5.8	8.2.0.108.4p	√ <sup>2</sup>
RHEL 5.9 RHEL 5.10 RHEL 5.11	8.2.0.128.3p	√ <sup>2</sup>
RHEL 6.0	8.3.5.17	√ <sup>2</sup>
RHEL 6.1	8.3.5.30.1p	√ <sup>2</sup>
RHEL 6.2	8.3.5.45.4p	√ <sup>2</sup>
RHEL 6.3	8.3.5.68.5p	√ <sup>2</sup>
RHEL 6.4	8.3.5.86.1p	√ <sup>2</sup>
RHEL 6.5	8.3.7.21.4p	√ <sup>2</sup>
RHEL 6.6	10.2.8020.1	√ <sup>2</sup>
RHEL 6.7	10.6.0.20	√ <sup>2</sup>
RHEL 7.0	8.3.7.34.3p	√ <sup>2</sup>
RHEL 7.1	10.2.8021.1	√ <sup>2</sup>
Linux OL 6.x [x86_64] UEK R1 [2.6.23-100]	8.3.5.30.1p	√ <sup>1</sup>
Linux OL 5.x [x86_64] UEK R1 [2.6.32-100]	8.3.18	√ <sup>1</sup>

**Table 8** Supported FCOE in-kernel drivers (page 2 of 3)

OS	Driver Version	Supported Adapters
		CNA
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	√ <sup>1</sup>
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] 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]	8.3.5.45.4p	√ <sup>2</sup>
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	√ <sup>1</sup>
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	√ <sup>1</sup>
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	√ <sup>2</sup>
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)	√ <sup>2</sup>
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	√ <sup>2</sup>
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	√ <sup>2</sup>
Linux OL 6.x [x86_64] UEK R3 [3.8.13-16]	8.3.7.26.2p	√ <sup>2</sup>
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	√ <sup>2</sup>
Linux OL 6.x [x86_64] UEK R3 U4 [3.8.13-55]	10.2.8061.0	√ <sup>2</sup>

**Table 8** Supported FCOE in-kernel drivers (page 3 of 3)

OS	Driver Version	Supported Adapters
		CNA
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	√ <sup>2</sup>
	8.3.7.34.4p	√ <sup>2</sup>
Linux OL 7.x [x86_64] UEK R3 U4 [3.8.13-55]	10.2.8061.0	√ <sup>2</sup>
Linux OL 7.x [x86_64] UEK R3 U5 [3.8.13-68]	10.6.61.0	√ <sup>2</sup>
SLES 10 SP4	8.2.0.92.1p	√ <sup>1</sup>
SuSE SLES 11 GA	8.2.8.14	√ <sup>1</sup>
SuSE SLES 11 SP1	8.3.5.8.1p 8.3.5.8.2p	√ <sup>1</sup>
SuSE SLES 11 SP2	8.3.5.48.2p 8.3.5.48.3p	√ <sup>2</sup>
SuSE SLES 11 SP3	8.3.7.10.6p	√ <sup>2</sup>
SuSE SLES 11 SP4	10.4.8000.0.	√ <sup>2</sup>
SuSE SLES 12	10.2.8040.1	√ <sup>2</sup>
XenServer [x86_64] 6.0.2	8.3.5.39.1p	√ <sup>1</sup>
XenServer [x86_64] 6.1	8.3.5.77.1p	√ <sup>1</sup>
XenServer [x86_64] 6.2	8.3.7.14	√ <sup>2</sup>
XenServer [x86_64] 6.5	10.2.340.0	√ <sup>2</sup>

1. For models LP21xxx, OCe10xxx and OCe11xxx CNAs.
2. For models LP21xxx, OCe10xxx, OCe11xxx and OCe14xxx CNAs.

**Table 9** Supported FCOE out-of-kernel drivers

OS	Driver Version	Supported Adapters
		CNA
RHEL 5.5 - 5.8	10.4.255.16	√ <sup>1</sup>
RHEL 6.4 - 6.6	10.4.255.16	√ <sup>1</sup>
RHEL 7.0	10.4.255.16	√ <sup>1</sup>

**Table 9** Supported FCOE out-of-kernel drivers

OS	Driver Version	Supported Adapters
		CNA
SLES 10 SP3-SP4	8.2.2.10.1p-1	√ 1
SLES 11 SP2-SP3	10.4.255.16	√ 1
SLES 12	10.4.255.16	√ 1
XenServer 6.1-6.2	10.0.794.0	√ 1

1. For models OCe11xxx and OCe14xxx CNAs 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 the [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 adapters in 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. Refer to 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. Refer to [Table 10 on page 83](#) for supported in-kernel driver versions. [Table 11 on page 86](#) lists the out-of-kernel driver versions supported with the corresponding OS updates.

Refer to the latest [Dell EMC Simple Support Matrix](#) for your specific Linux distribution, kernel version, and driver to determine whether or not 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 page](#).

**Table 10 Supported FCOE in-kernel drivers (page 1 of 3)**

OS	Driver Version	Supported Adapters
		CNA
RHEL 5.4	8.03.00.10.05.04-k 8.03.00.1.05.05-k	√ <sup>1</sup>
RHEL 5.5	8.03.01.04.05.05-k	√ <sup>1</sup>
RHEL 5.6	8.03.01.05.05.06-k	√ <sup>2</sup>
RHEL 5.7	8.03.07.00.05.07-k	√ <sup>2</sup>
RHEL 5.8	8.03.07.09.05.08-k	√ <sup>2</sup>
RHEL 5.9 RHEL 5.10 RHEL 5.11	8.03.07.15.05.09-k	√ <sup>2</sup>
RHEL 6.0	8.03.05.01.06.1-k0	√ <sup>2</sup>
RHEL 6.1	8.03.07.03.06.1-k	√ <sup>2</sup>
RHEL 6.2	8.03.07.05.06.2-k	√ <sup>2</sup>
RHEL 6.3	8.04.00.04.06.3-k	√ <sup>2</sup>
RHEL 6.4	8.04.00.08.06.4-k	√ <sup>2</sup>
RHEL 6.5	8.05.00.03.06.5-k2	√ <sup>2</sup>
RHEL 6.6	8.07.00.08.06.6-k1	√ <sup>2</sup>

**Table 10 Supported FCOE in-kernel drivers (page 2 of 3)**

OS	Driver Version	Supported Adapters
		CNA
RHEL 6.7	8.07.00.16.06.7-k	√ 2
RHEL 7.0	8.06.00.08.07.0-k2	√ 2
RHEL 7.1	8.06.00.08.07.1-k2	√ 2
Linux OL 6.x [x86_64] UEK R1 [2.6.23-100]	8.03.07.03.32.1-k	√ 1
Linux OL 5.x [x86_64] UEK R1 [2.6.32-100]	8.03.01.02.32.1-k9	√ 1
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	√ 1
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	√ 2
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	√ 1
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	√ 1
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	√ 2
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)	√ 2
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	√ 2
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	√ 2



**Table 10 Supported FCOE in-kernel drivers (page 3 of 3)**

OS	Driver Version	Supported Adapters
		CNA
Linux OL 6.x [x86_64] UEK R3 [3.8.13-16]	8.05.00.03.39.0-k	√ <sup>2</sup>
Linux OL 6.x [x86_64] UEK R3 U1 [3.8.13-26]	8.06.00.14.39.0-k	√ <sup>2</sup>
Linux OL 6.x [x86_64] UEK R3 U2 [3.8.13-35]	8.07.00.08.39.0-k1	√ <sup>2</sup>
Linux OL 6.x [x86_64] UEK R3 U3 [3.8.13-44]	8.07.00.08.39.0-k1	√ <sup>2</sup>
Linux OL 6.x [x86_64] UEK R3 U4 [3.8.13-55]	8.07.00.16.39.0-k	√ <sup>2</sup>
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	√ <sup>2</sup>
Linux OL 7.x [x86_64] UEK R3 U4 [3.8.13-55]	8.07.00.16.39.0-k	√ <sup>2</sup>
Linux OL 7.x [x86_64] UEK R3 U5 [3.8.13-68]	8.07.00.16.39.0-k	√ <sup>2</sup>
SLES 10 SP4	8.03.01.12.10.3-k4	√ <sup>2</sup>
SuSE SLES 11 GA	8.02.01.03.11.0-k9	√ <sup>1</sup>
SuSE SLES 11 SP1	8.03.01.06.11.1-k8 8.03.01.07.11.1-k8 8.03.01.08.11.1-k8	√ <sup>2</sup>
SuSE SLES 11 SP2	8.03.07.07-k	√ <sup>2</sup>
SuSE SLES 11 SP3	8.04.00.13.11.3-k	√ <sup>2</sup>
SuSE SLES 11 SP4	8.07.00.18-k	√ <sup>2</sup>
SuSE SLES 12	8.07.00.08.12.0-k	√ <sup>2</sup>
XenServer [x86_64] 6.0.2	8.03.07.03.55.6-k2	√ <sup>2</sup>
XenServer [x86_64] 6.1	8.04.00.02.55.6-k	√ <sup>2</sup>
XenServer [x86_64] 6.2	8.05.00.03.55.6-k	√ <sup>2</sup>
XenServer [x86_64] 6.5	8.07.00.09.66.5-k	√ <sup>2</sup>

1. For models QLE80xx, and QLE81xx CNAs.
2. For models QLE80xx, QLE81xx, and QLE82xx CNAs.

**Table 11** Supported FCOE out-of-kernel drivers

OS	Driver Version	Supported Adapters
		CNA
SLES 10 SP4	8.04.00.15.10.3-k 8.06.00.10.10.3-k	√ <sup>1</sup>
SLES 11 SP2	8.05.00.03.11.1-k 8.06.00.10.11.1-k	√ <sup>1</sup>
SLES 11 SP3	8.06.00.10.11.1-k	√ <sup>1</sup>
RHEL 5.8-5.9	8.04.00.15.5.6-k 8.06.00.11.5.6-k	√ <sup>1</sup>
RHEL 5.10	8.06.00.11.5.6-k	√ <sup>1</sup>
RHEL 6.2	8.06.00.10.06.0-k	√ <sup>1</sup>
RHEL 6.3-6.4	8.05.00.03.06.0-k 8.06.00.10.06.0-k	√ <sup>1</sup>

1. 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 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. Refer to the package read1st.txt for package details.

The firmware and bootcode can be updated by using QConvergeConsole GUI/CLI Management tools. Refer to the QConvergeConsole User Guide available from the [QLogic website](#).

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

### QLogic 1000/1800 series

**Note:** : On January 17, 2014 Brocade's adapter business was acquired by QLogic Corp. All Fibre Channel Host Bus Adapters (HBAs), Converged Network Adapters (CNAs), and mezzanine adapters for OEM blade server platforms are now QLogic products.

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. Refer to [Table 12 on page 87](#) for supported in-kernel driver versions. [Table 13 on page 89](#) lists the out-of-kernel driver versions supported with the corresponding OS updates.

Refer to the latest [Dell EMC Simple Support Matrix](#) for your specific Linux distribution, kernel version, and driver to determine whether or not 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 12** Supported FCOE in-kernel drivers (page 1 of 2)

OS	Driver Version	Supported Adapters
		CNA
RHEL 5.5	2.1.2.0	√ 1
RHEL 5.6	2.1.2.0	√ 2
RHEL 5.7	2.3.2.5	√ 2
RHEL 5.8	3.0.2.2	√ 2
RHEL 5.9 RHEL 5.10 RHEL 5.11	3.0.23.0	√ 2
RHEL 6.0	2.1.2.1	√ 2
RHEL 6.1	2.3.2.3	√ 2
RHEL 6.2	3.0.2.2	√ 2
RHEL 6.3	3.0.2.2	√ 2
RHEL 6.4	3.0.23.0	√ 2
RHEL 6.5	3.2.21.1	√ 2
RHEL 6.6	3.2.23.0	√ 2
RHEL 6.7	3.2.23.0	√ 2
RHEL 7.0	3.2.23.0	√ 2
RHEL 7.1	3.2.23.0	√ 2
SLES 10 SP4	2.3.2.1	√ 1

**Table 12** Supported FCOE in-kernel drivers (continued) (page 2 of 2)

OS	Driver Version	Supported Adapters
		CNA
SLES 11	1.1.0.2	√ <sup>1</sup>
SLES 11 SP1	2.1.2.1	√ <sup>1</sup>
SLES 11 SP2	3.0.2.2	√ <sup>1</sup>
SLES 11 SP3	3.1.2.1	√ <sup>2</sup>
SLES 11 SP4	3.2.23.0	√ <sup>2</sup>
SLES 12	3.2.23.0	√ <sup>2</sup>
Linux OL 6.x [x86_64] UEK R1 [2.6.23-100]	2.3.2.3	√ <sup>2</sup>
Linux 6.x [x86_64] UEK R1 U2 [2.6.32-300]	3.0.2.2	√ <sup>2</sup>
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	√ <sup>2</sup>
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	√ <sup>2</sup>
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	√ <sup>2</sup>
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	√ <sup>2</sup>
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	√ <sup>2</sup>
XenServer [x86_64] 6.0.2	2.3.2.7	√ <sup>2</sup>
XenServer [x86_64] 6.1	3.1.0.0	√ <sup>2</sup>
XenServer [x86_64] 6.2	3.2.1.1	√ <sup>2</sup>
XenServer [x86_64] 6.5	3.2.1.1	√ <sup>2</sup>

1. For models EM1010/1020 only.
2. For models EM1010/1020, EM-BR1860.

**Table 13** Supported FCOE out-of-kernel drivers

OS	Driver Version	Supported Adapters
		CNA
RHEL 5.8 -5.9	3.2.3.0	√ <sup>1</sup>
RHEL 6.3-6.4	3.2.3.0	√ <sup>1</sup>
SLES10 SP3-SP4	3.2.3.0	√ <sup>1</sup>
SLES 11 SP2-SP3	3.2.3.0	√ <sup>1</sup>

1. 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

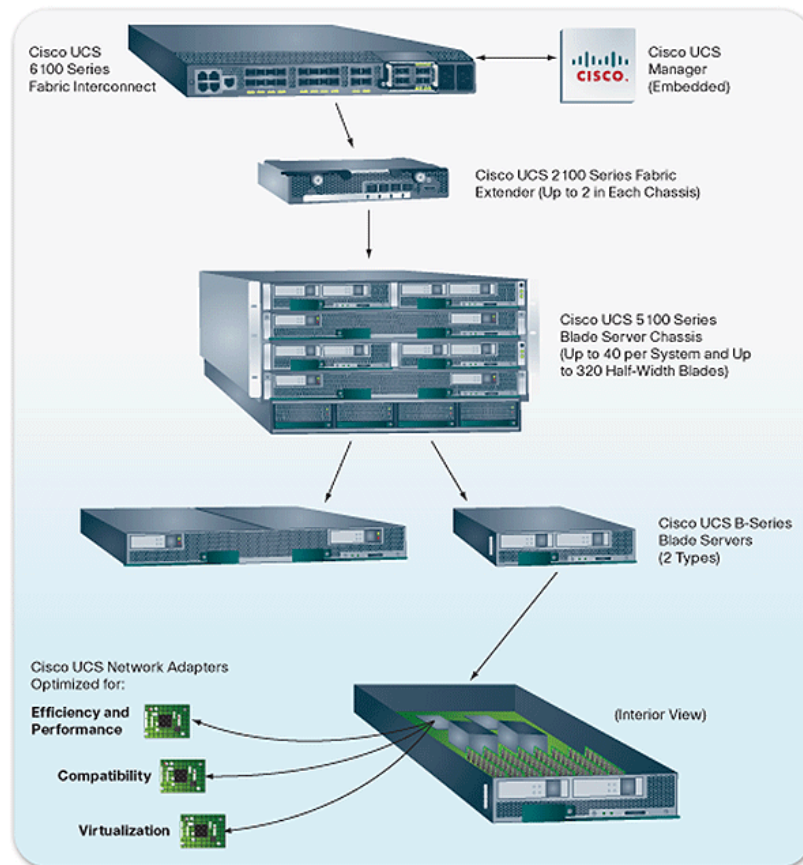
The Brocade CNA firmware and bootcode are provided in a multi-boot firmware image, which is available from the [QLogic website](#).

The image can be flashed using either the BCU CLI or Host Connectivity Manager on systems. Refer to the *BR-Series Adapters Administrator's Guide* available from the [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 [Figure 9](#), 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 designed for virtualization, FCoE connectivity, and centralized management.



**Figure 9** 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. Refer to 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 board 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. Refer to the *Fibre Channel over Ethernet (FCoE) and Data Center Bridging (DCB) Case Studies TechBook* on [Dell EMC E-Lab Interoperability Navigator \(ELN\)](#) at the Resource Center tab, for information on supported features and topologies.





# CHAPTER 4

## iSCSI Connectivity

This chapter provides information on Fibre Channel connectivity, including the following:

☒ Introduction.....	94
☒ iSCSI discovery .....	95
☒ iSCSI solutions .....	98
☒ Native Linux iSCSI Attach .....	103
☒ Known problems and limitations.....	116

## Introduction

iSCSI (Internet Small Computer System Interface) is an IP-based storage networking standard for linking data storage facilities 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 in conjunction 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

In order 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 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.

---

**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 as well as digest failure recovery
- ☒ Error recovery level 2 implies level 1 capabilities as well as connection recovery



### **IMPORTANT**

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 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:

#### **CHAP (Challenge Handshake Authentication Protocol)**

The Challenge-Handshake Authentication Protocol (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 through the use of 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” on page 98
- ☒ “General supported configurations” on page 98
- ☒ “Dell EMC native iSCSI targets” on page 99
- ☒ “Native Linux iSCSI driver” on page 102
- ☒ “Software and hardware iSCSI initiator” on page 102

## 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 will lead 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.

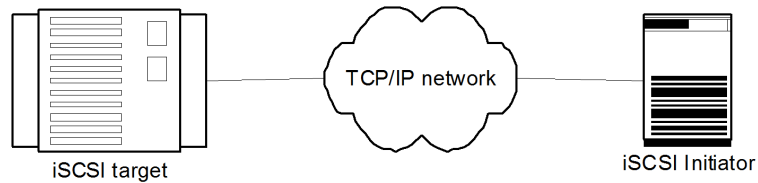
### Direct connect

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 10** Direct connect example

**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, we recommend using enterprise class equipment because they provide more memory to address network congestion.



**Figure 11** Network connect example

## Dell EMC native iSCSI targets

This section discusses the following Dell EMC native iSCSI targets:

- ☒ “VMAX series” on page 99
- ☒ “VNX series iSCSI/FC system” on page 99
- ☒ “VNXe/Unity Series” on page 100
- ☒ “XtremIO” on page 100
- ☒ “CLARiiON” on page 101
- ☒ “Celerra” on page 101

**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 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 CHAP (Challenge Handshake Authentication Protocol).

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 VMAX series iSCSI target, refer to the VMAX configuration guide.

For up-to-date iSCSI host support refer to [Dell EMC E-Lab Interoperability Navigator](#).

For configuration of an iSCSI server, refer to 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 ports.
- ☒ iSNS protocol is supported. The authentication mechanism is CHAP (Challenge Handshake Authentication Protocol).
- ☒ 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 (Challenge Handshake Authentication Protocol).

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, refer to the appropriate configuration guide.

For up-to-date iSCSI host support, refer to Dell EMC Simple Support Matrix, available through [Dell EMC E-Lab Interoperability 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.

Refer to the Server/OS vendor support matrix for 10 Gb iSCSI compatibility.

**References**

For configuration of XtremIO targets, refer to the appropriate configuration guide.

For up-to-date iSCSI host support refer to [Dell EMC E-Lab Interoperability Navigator](#).

For configuration of the iSCSI server, refer to 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
<code>replacement_timeout</code>	120	Specifies the number of seconds the iSCSI layer holds for a timed-out path/session to re-establish before failing any commands on that path/session. The default value is 120.
<code>FirtsBurstLength</code>	<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 default setting for this parameter (256KB).

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`**



**CLARiiON** This can be configured to support iSCSI and FC connectivity simultaneously; however, no Linux host may be connected to the same VNX series or CLARiiON system 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 (Challenge Handshake Authentication Protocol).

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

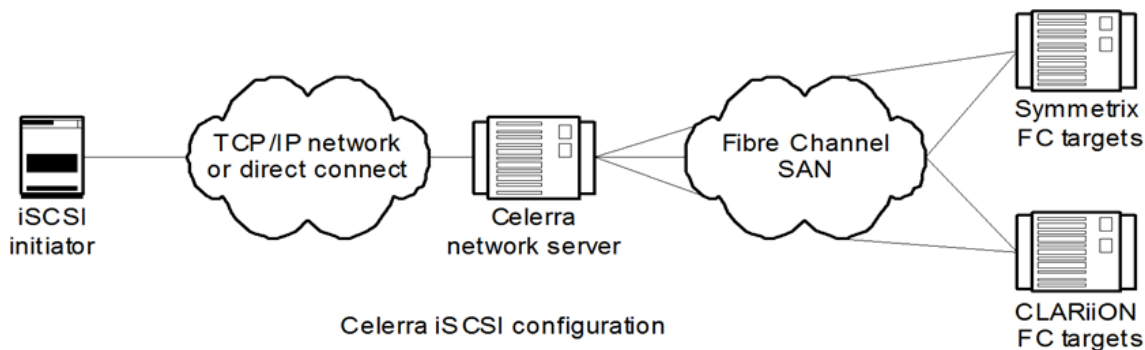
## References

For configuration of CLARiiON iSCSI targets, refer to the appropriate configuration guide.

For up-to-date iSCSI host support, refer to [Dell EMC E-Lab Interoperability Navigator](#).

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

**Celerra** Celerra Network Server provides iSCSI target capabilities combined with NAS capabilities. The Celerra iSCSI system is defined by creating a file system. The file system is built on FC LUNs that are accessible on Symmetrix or VNX series or CLARiiON systems. The file system is then mounted on the Celerra server data movers. Out of the file system iSCSI LUNs are defined and allocated to iSCSI targets. The targets are then associated with one of the Celerra TCP/IP interfaces.



**Figure 12** Celerra Network server

All Celerra network servers can be configured to provide iSCSI services.

The following are some of the characteristics of the Celerra Network Server:

- ☒ iSCSI error recovery level 0 (session-level recovery)
- ☒ Supports CHAP with unlimited entries for one-way authentication and one entry for reverse authentication
- ☒ Uses iSNS protocol for discovery
- ☒ Provides 10 Gb/s and 1 Gb/s interfaces
- ☒ Supports Dell EMC storage Symmetrix, VNX series, and CLARiiON on the back end

## 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, refer to the [Internet Engineering Task Forces \(IETF\) website](#).

For more information about the iSCSI protocol, refer to the [RFC 3720 page](#).

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 via 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 via the IP network between it and the storage devices attached to it.

For detailed information regarding the native Linux iSCSI driver refer to “[Native Linux iSCSI Attach](#)” on page 103.

## 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 a number of 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 an HBA and CNA support matrix, refer to [Dell EMC E-Lab Interoperability Navigator](#).

## Native Linux iSCSI Attach

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

- ☒ “open-iscsi driver” on page 103
- ☒ “Setting initiator name in software iSCSI” on page 108
- ☒ “Selective target(s) login” on page 109
- ☒ “Starting and stopping the iSCSI driver” on page 110
- ☒ “Dynamic LUN discovery” on page 111
- ☒ “Mounting and unmounting iSCSI file systems automatically (RHEL, Asianux, and SLES)” on page 111
- ☒ “Excessive dropped session messages found in /var/log/messages” on page 112

### open-iscsi driver

This section discusses the following open-iscsi driver information:

- ☒ “open-iscsi driver introduction” on page 103
- ☒ “Features” on page 104
- ☒ “README file and main pages” on page 104
- ☒ “Environment and system requirements” on page 106
- ☒ “Installing the open-iscsi driver” on page 106
- ☒ “Known problems and limitations” on page 116

#### open-iscsi driver introduction

This driver is found in RHEL 5, 6, and 7; SLES 10, 11, and 12; 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 `iscsiadm`. Refer to 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 refer to the [Open-iSCSI website](#).

**README file and main pages**

- ☒ For RHEL 5, 6, and 7, refer to the following:
  - `/usr/share/doc/iscsi-initiator-utils-x.x.x/README`
  - main pages for:
    - `iscsid`
    - `iscsiadm`
- ☒ For SLES 10, 11, and 12, refer to the following:
  - `/usr/share/doc/packages/open-iscsi/README`
  - main pages for:
    - `iscsid`
    - `iscsiadm`
- ☒ For Asianux 3 and 4, refer to the following:
  - `/usr/share/doc/iscsi-initiator-utils-x.x.x/README`
  - main pages for:
    - `iscsid`
    - `iscsiadm`



**IMPORTANT**

Follow the configuration guidelines that Dell EMC outlines. Using improper settings can 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 13 on page 105 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.

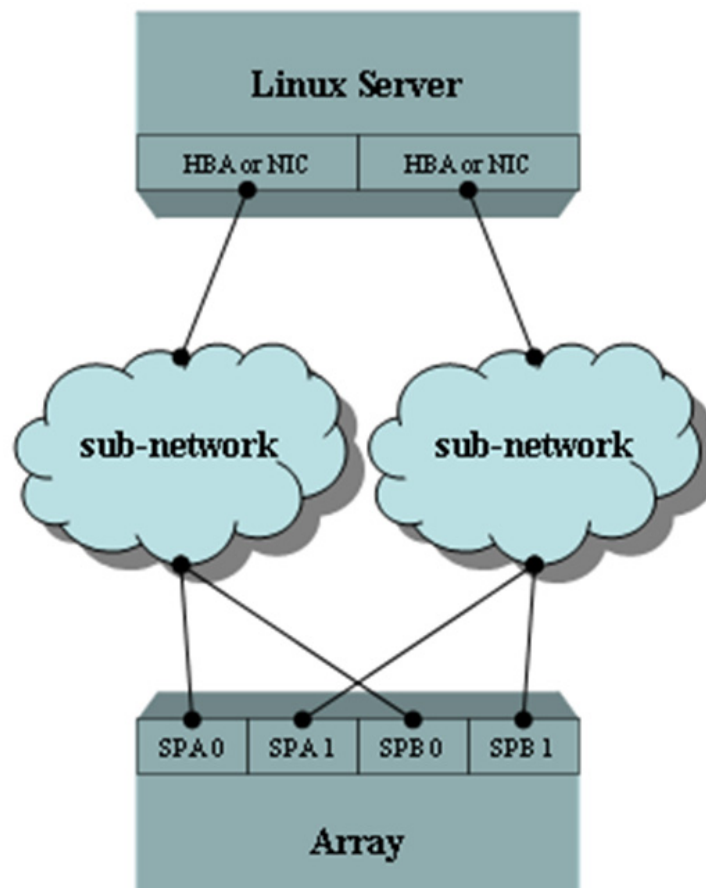
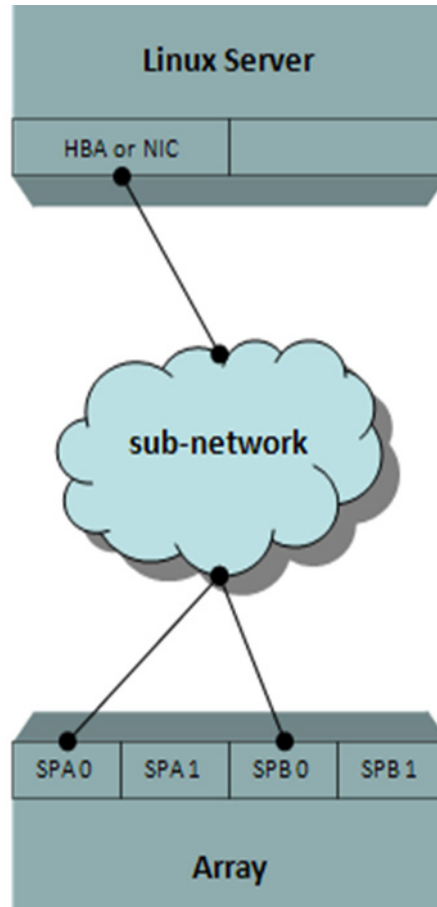


Figure 13 Linux iSCSI with multiple NICs/HBAs to multiple subnets example

Figure 14 on page 106 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 14** Linux iSCSI with single NIC/HBA to single subnet example

### Environment and system requirements

Consult the [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

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.

**Note:** Complete the following steps before continuing on to the Asianux 3.0, or SLES 10 installation subsections.

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 Asianux 3.0, and SLES 10 GA are located in `/var/lib/open-iscsi/`. For SLES 10 SP1 they are located in `/etc/iscsi/`.

Dell EMC recommends the following steps to complete the installation. These recommendations are generic to all distributions unless otherwise noted.

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 utilizes a pound (#) symbol to comment out a line in the configuration file. You can enable a variable listed in [Table 14](#) by deleting the pound (#) 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; as well as, the configuration file itself.

**Table 14** RHEL 5/6/ 7, Asianux 3/ 4, SLES 10, SLES 11/12 (Linux 2.6 kernel and up)

Variable name	Default settings	Dell EMC recommended settings
<code>node.startup</code>	Asianux and SLES: Manual Redhat: automatic	Auto
<code>node.session.iscsi.InitialR2T</code>	No	Yes
<code>node.session.iscsi.ImmediateData</code>	Yes	No
<code>node.session.timeo.replacement_timeout</code>	120	60
<code>node.conn[0].timeo.noop_out_interval</code>	5	Higher in congested networks <sup>a</sup>
<code>node.conn[0].timeo.noop_out_timeout</code>	5	Higher in congested networks <sup>b</sup>

- 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].timeo.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-iscsi on`
- RHEL 7 and SLES 12:  
`# systemctl enable iscsid`  
`# systemctl enable iscsi`

3. Refer to “[SLES 10, SLES 11, and SLES 12](#)” on page 108 to complete your installation on a SLES server or “[RHEL 5, 6 and 7 and Asianux 3, 4](#)” on page 145 to complete your installation on a RHEL 5 and Asianux 3.0.

**SLES 10, SLES 11, and SLES 12**

Dell EMC recommends using 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.

1. Open YaST and select **Network Services > iSCSI Initiator**. From here you can open the **Discovered Targets** tab by entering the IP address of the target.  
  
With an XtremIO, Unity, VNX series, or CLARiiON system 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 VMAX arrays you must enter each individual target you wish to discover.
2. After discovering your targets, select the **Connected Targets** tab to log in to the targets you wish to be connected to and select those you wish to log in to automatically at boot time.

**RHEL 5, 6, and 7 and Asianux 3 and 4**

For RHEL 5, 6 and 7, or Asianux 3 and 4, you must perform a series of `iscsiadm(8)` commands to configure the targets you want to connect to with `open-iscsi`. Consult the main pages for `iscsiadm` for a detailed explanation of the command and its syntax.

1. First, discover the targets you want to connect your server to via iSCSI.  
  
For XtremIO, Unity, VNX series or CLARiiON systems 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 VMAX arrays you must perform the discovery process on each individual target.
2. Before you perform the discovery of each port on the VMAX series, configure the targeted VMAX series iSCSI ports to accept the IQN of your initiator.  
  
Refer to the appropriate VMAX series installation and administration documentation posted at [Dell EMC Online Support](#).

**IMPORTANT**

If you do not configure the initiator on a VMAX series array first, `Sendtarget` 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`

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

For further details, consult the [iSCSI protocol specification \(RFC 3720\) page](#).

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, refer to operating system vendor specific documentation.

## Selective target(s) 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
```

Refer to the following sections for more information:

- ☒ “Starting and stopping the iSCSI driver” on page 110
- ☒ “Dynamic LUN discovery” on page 111
- ☒ “iSCSI Write Optimization in Unity, VNX series, or CLARiiON CX environment” on page 113
- ☒ “Mounting and unmounting iSCSI file systems automatically (RHEL, Asianux, and SLES)” on page 111
- ☒ “Excessive dropped session messages found in /var/log/messages” on page 112

## 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 will be placed in the `/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.

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

- ☒ To manually stop the iSCSI driver for RHEL 5, RHEL 6, and Asianux 3.0, 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 will 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, we recommend that you type `/etc/init.d/iscsi start`.

- ☒ To manually start the iSCSI driver on RHEL 7 and SLES 12, 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 IO quiesced. To reload the iSCSI driver on RHEL, Asianux and SLES, use the following command as *root*:

- ☒ RHEL 5 and RHEL 6, and Asianux 3 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 and SLES 12:

```
# systemctl restart iscsi
```

```
# systemctl restart iscsid
```

---

**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 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 will be unmounted automatically by the `netfs` `initscript` (before iSCSI is stopped) during normal shutdown, and you can more 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/emcpower1 /mnt/emcpower1 ext2 _netdev 0 0
```

```
/dev/sde1 /mnt/sde1 ext2 _netdev 0 0
```

For SLES, you can use YaST to choose which device(s) you want to mount upon system boot. For details, refer to 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/emcpower1 /mnt/emcpower1 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
```

---

**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/emcpower1 /mnt/emcpower1 ext3 acl,user,nofail 0 0
```

For information, see TID 7004427: */etc/fstab entry does not mount iSCSI device on boot up* at <http://www.novell.com/support/php/search.do?cmd=displayKC&docType=kc&externalId=7004427>.

## 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 will recover the session, it can rob the system of 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, VNX series, or CLARiiON CX environment

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 CLARiiON CX family with FLARE code 04.29.000.5.013 and later, as well as all 04.30.000.5 releases, and VNX OE for Block version 31 and later, support 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 this task in privileged mode, as shown in [Table 15](#).

**Table 15** Enabling flow control on a Brocade switch

Step	Task	Command
1	Set the flow control parameters on a gigabit Ethernet port.	<ul style="list-style-type: none"> <li>• no switchport</li> <li>• no cee</li> <li>• switchport</li> <li>• switchport mode access</li> <li>• no shutdown</li> <li>• qos flowcontrol tx on rx</li> </ul>
2	Verify the flow control configuration.	<ul style="list-style-type: none"> <li>• On sh run int te 0/7</li> </ul>

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 CSMH 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 this task in privileged mode, as shown in [Table 16](#).

**Table 16** 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 Receive  FlowControl RxPause  TxPause  Unsupported adminoper
admin      oper                opcodes
-----  -----
2/1  on      on      on      on      0      0      0

```

Perform the task in [Table 17](#) to enable flow control on a gigabit Ethernet NIC and disable window scaling on a Linux host.

**Table 17** 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)`, refer 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

Table 18 lists the known problems and limitations.

**Table 18** Known problems and limitations

Issue number	Description	Workaround/resolution
Novell BZ: #4 74455 (SLES 10 SP3) #498369 (SLES 11)	There may be data corruption over iSCSI when working with large files.	Novell 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 <code>/var/log/messages</code> displays a message similar to this one: <pre>Jan 25 09:31:44 lin048126 kernel: scsi scan: 192 byte inquiry failed. Consider BLIST_INQUIRY_36 for this device.</pre>	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)	<code>Open-iscsi</code> on SLES 10 SP3 does not reestablish the connection after a logout request from a VNXe.	This is seen in the GA release with <code>open-iscsi-2.0.868-0.6.11</code> . The issue is resolved by updating to <code>open-iscsi-2.0.868-0.11.1</code> and later.
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 <code>powermt display dev=all</code> 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 <code>powermt display dev=all</code> command shows paths as alive while the target cable pull is done against the owner SP.	No solution exists.



Table 18 Known problems and limitations

Issue number	Description	Workaround/resolution
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 located in the <code>/var/log/messages</code> file. To prevent the call trace from logging into the kernel log messages, add the <code>nousb</code> kernel parameter when the system is booting.	Redhat 7.0 To prevent the call trace from logging into the kernel log messages, add the <code>nousb</code> 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 same SAN MAC address is exposed to both the <code>correspondingifconfig</code> 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 will attempt to acquire DHCP using this iSCSI function, which will fail, and Anaconda will then try to acquire DHCP indefinitely.	To work around this problem, if DHCP is desired, you must use the <code>asknetwork</code> installation parameter and provide a "dummy" static IP address to the corresponding network interface of the iSCSI function. This prevents Anaconda from entering an infinite loop and allows it to instead request the iSCSI offload function to perform DHCP acquisition.
<code>iscsi-initiator-utils</code> component, BZ # 825185 Redhat 6.5	If the corresponding network interface has not been brought up by <code>dracut</code> or the tools from the <code>iscsi-initiator-utils</code> 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: <code>iscsiadm: Error while adding record: no available memory.</code> The host is non-operational when connecting to storage.	An upstream patch was applied and the <code>/var/lib/iscsi</code> file is now successfully attached.
BZ # 983553 Redhat 6.5	Prior to this update, a single unreachable target could block rescans of others. The <code>iscsiadm</code> utility could halt in the D state and the rest of the targets could remain unscanned.	To fix this bug, <code>iscsiadm</code> was made terminable and all the targets were updated. Functioning sessions will now be rescanned properly without long delays.

**Table 18** Known problems and limitations

Issue number	Description	Workaround/resolution
BZ # 917600 Redhat 6.5	Support for managing flash nodes from the <code>open-iscsi</code> utility was added to this package.	If you use <code>iscsi-initiator-utils</code> , 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.
BZ # 865739 Redhat 6.5	Previously, the <code>tgt</code> daemon did not report its exported targets properly if configured to report them to an Internet Storage Name Service (iSNS) server. Consequently, running the <code>iscsiadm -m discoverydb -t isns</code> command failed.	This bug was fixed and <code>tgt</code> now reports its exported targets correctly in the described scenario
OPT # 212991	Multiple Linux hosts have the same <code>iqn</code> identifiers.	The administrator should check the <code>/etc/initiatorname.iscsi</code> file and ensure that each host has a unique <code>iqn</code> name. To generate a new <code>iqn</code> identifier, use the iSCSI <code>iscsi-iname</code> utility. For RHEL, use <code>/sbin/iscsi-iname</code> and for SLES, use <code>/usr/sbin/iscsi-iname</code> .
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. Refer to “Excessive dropped session messages found in <code>/var/log/messages</code> ” on page 112.
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.

**Table 18** Known problems and limitations

Issue number	Description	Workaround/resolution
Novell BZ # 212760 Novell BZ # 251675 Novell 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 3.0, the same sleep is installed in /etc/init.d/iscsid after modprobe -q iscsi_tcp.
OPT #287890 Asianux BZ #5259	The Asianux 3.0 GA host may reboot when I/O is generated to an PowerPath devices.  This is caused by bnx2 (Broadcom NetXtreme II BCM5706/5708 Driver) version 1.4.44-1.  For further information, refer to the <a href="#">Red Hat Bugzilla – Bug 212055 page</a> .	Update to Asianux 3 SP1. Asianux 3 SP1 has the updated version of the bnx2 driver included in the kernel.
DIMS # 145916 Novell BZ # 172447	A server may not automatically register with a a VNX series or CLARiiON iSCSI system via Unisphere/Navisphere or the Unisphere/Navisphere Server Utility when installed on SLES 10.	This was fixed in SLES 10 SP1.



# CHAPTER 5

## Booting From SAN

Installing and booting Linux from a SAN (storage area network) environment is supported on all Dell EMC storage arrays. The Linux host operating system can reside on an external device managed by Linux native DM-MPIO utility or PowerPath software. This chapter is a summary of the major steps involved in the configuration process and considerations to be taken to prevent possible issues and includes the following information:

☒ Supported environments .....	122
☒ Limitations and guidelines.....	123
☒ Preparing host connectivity .....	124
☒ Configuring a SAN boot for FC attached host .....	126
☒ Configuring SAN boot for iSCSI host.....	132
☒ Configuring SAN boot for FCoE attached host.....	139
☒ Multipath booting from SAN.....	146
☒ PowerPath booting from SAN.....	154
☒ Guidelines for booting from Symmetrix, XtremIO, VNX series, VNXe series, Unity series, or CLARiiON 155	

## Supported environments

Dell EMC storage environments such as VMAX series, VNX, Unity, VPLEX, and XtremIO are supported.

Refer to [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 attached to the SAN environment, the host bus adapter's Boot BIOS must be installed and enabled on the adapter. Refer to the driver manuals and configuration guides found on the Dell EMC section of the Emulex (now Broadcom), QLogic, and Brocade websites as well as operating system, HBA and server vendor documentation.

## Notes

- ⊗ The AX100/100i, AX150/150i are supported only with the low-cost HBAs. Refer to the [Dell EMC Simple Support Matrix](#) for supported HBAs with these arrays.
- ⊗ iSCSI booting from SAN is supported in limited configurations. Refer to the [Dell EMC Simple Support Matrix](#) for supported environments.

## Limitations and guidelines

Boot configurations must not deviate from the following limitations established by Dell EMC:

- ☒ The Dell EMC Storage device must have enough disk space to hold the Linux operating system.
- ☒ The VMAX series, VNX series, VNXe series, Unity series, or CLARiiON device that is to contain the Master Boot Record (MBR) for the host must have a lower logical unit number (LUN) than any other device visible to the host.
- ☒ Space reclamation, available with Enginuity 5874 and later, is prohibited for use on VMAX Virtual Provisioning (thin) devices which are utilized for host /boot, / (root), /swap, and /dump volumes.
- ☒ Boot from SAN using Linux DM-MPO when configured to VNX series or CLARiiON storage is supported with PNR devices and ALUA devices. Boot from a SAN using Linux DM-MPO when configured to 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. Refer to the restrictions and notes in the Overview section of [“Multipath booting from SAN” on page 146](#).
- ☒ 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 OS. The configuration can be altered after installation.
- ☒ If multiple HBAs are attached to the host, make sure 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.
- ☒ Dell EMC recommends 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 or CLARiiON 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 lunoffset) 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 VMAX series, when you add devices to SG, you can use the `switch -lun 0`. (`symaccess -sid xxx -type stor -name host_sg_name add devs lun_id -lun 0`) command.
- ☒ In XtremIO version 4.0.0 or above, volumes are numbered by default starting from LUN ID 1. We do not recommend manually adjusting the LUN ID to 0, as it may lead to issues with some operating systems. In XtremIO 3.x and previous versions, LUN ID starts from 0, and still remains accessible when XtremIO cluster is updated from 3.0.x to 4.x.
- ☒ Additional LUNs can be added after OS 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. Refer to your Red Hat RHEL installation documentation for details of this type of installation.
- ☒ 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. Refer to your SuSE SLES installation documentation for details of this type of installation.



- When configuring a Linux boot LUN under PowerPath control, the boot LUN may be configured using only a single path during the OS installation; however, you will want to add the additional paths for PowerPath control. Refer to the PowerPath installation document for details of PowerPath installation.

## Single and dual path configuration examples

Figure 15 shows an example of single path configuration.

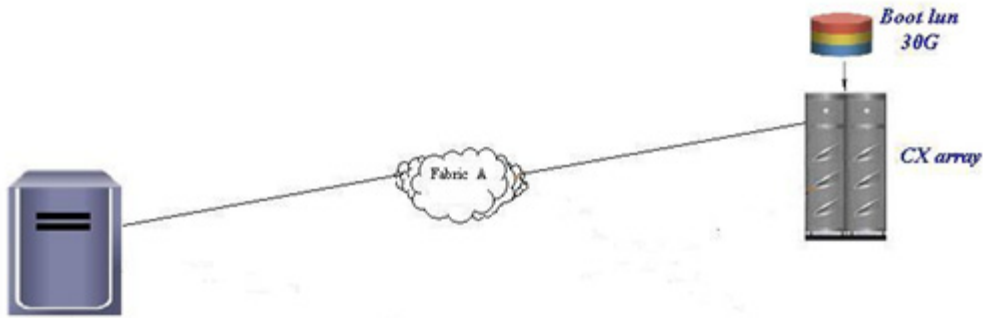


Figure 15 Single path configuration

Figure 16 shows an example of dual path configuration.

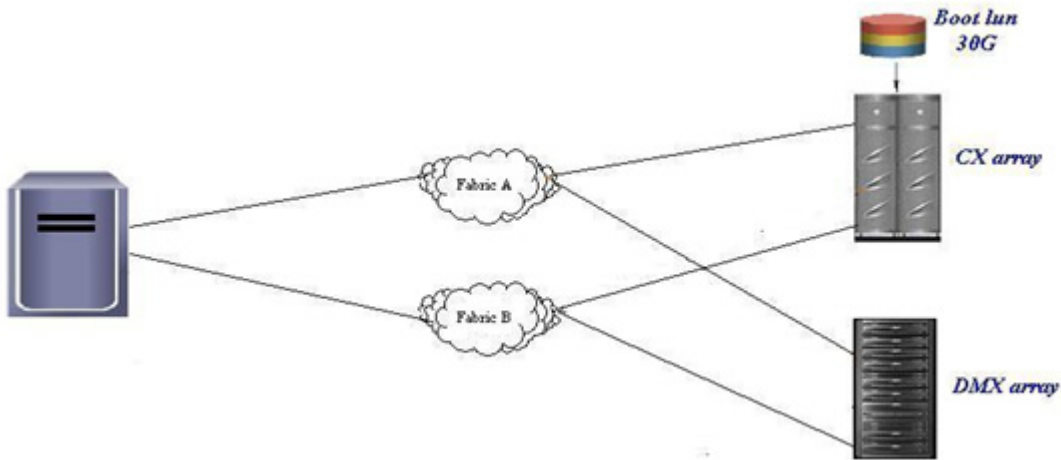


Figure 16 Dual path configuration

# Configuring a SAN boot for FC attached host

## Prepare host connectivity

Refer to Chapter 2, “Fibre Channel Connectivity” on page 57 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 OS kernel resides on an external device, it will not be loaded during a boot by the host system's hard disk controller. The OS 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. Refer to the respective vendor site for details

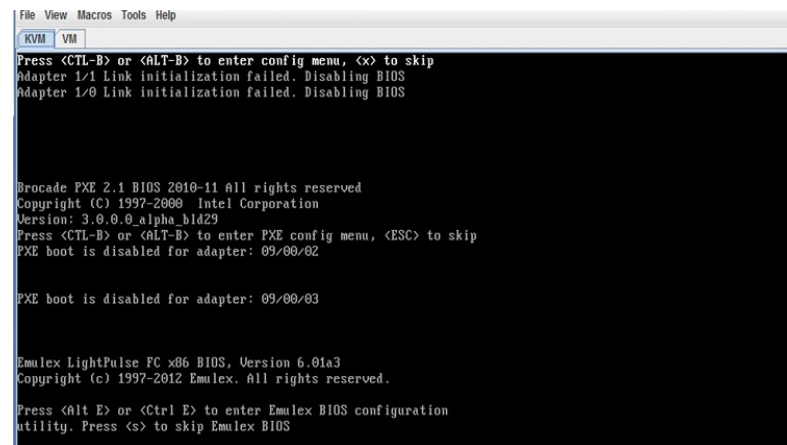
## Updating HBA BIOS and firmware

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

[Figure 17](#) is an example for setup RHEL 5.3 with Emulex HBA, and is done by single path connection during installation. Different vendor HBA and OS versions might be similar but require slight modifications. Refer to HBA and OS vendor websites for more details

Boot the server, and press ALT-E to enter Emulex BIOS when you see the message in [Figure 17](#).



```

File View Macros Tools Help
KVM VM
Press <CTL-B> or <ALT-B> to enter config menu, <X> to skip
Adapter 1/1 Link initialization failed. Disabling BIOS
Adapter 1/0 Link initialization failed. Disabling BIOS

Brocade PXE 2.1 BIOS 2010-11 All rights reserved
Copyright (C) 1997-2009 Intel Corporation
Version: 3.0.0.0_alpha_bld29
Press <CTL-B> or <ALT-B> to enter PXE config menu, <ESC> to skip
PXE boot is disabled for adapter: 09/00/02

PXE boot is disabled for adapter: 09/00/03

Emulex LightPulse FC x86 BIOS, Version 6.01a3
Copyright (c) 1997-2012 Emulex. All rights reserved.
Press <Alt E> or <Ctrl E> to enter Emulex BIOS configuration
utility. Press <s> to skip Emulex BIOS
  
```

**Figure 17** Entering the Emulex BIOS configuration

1. Select the adapter port you want to configure, as shown in Figure 18. (In a single path configuration, this should be the HBA port zoned to storage).

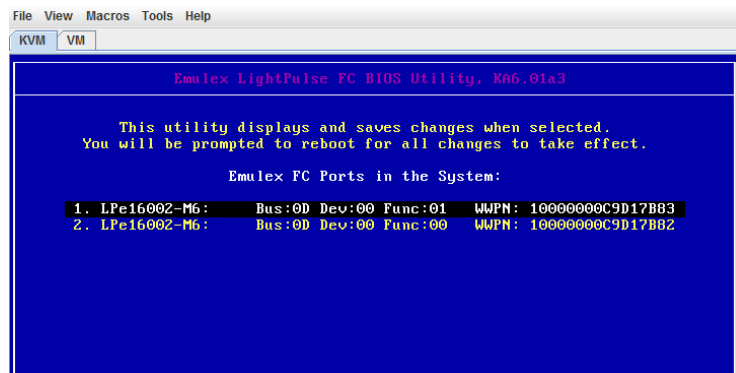


Figure 18 Selecting the HBA adapter port

2. If the link status is shown as **Link UP**, select **Enable/Disable Boot from SAN** to enable the HBA BIOS, as shown in Figure 18.

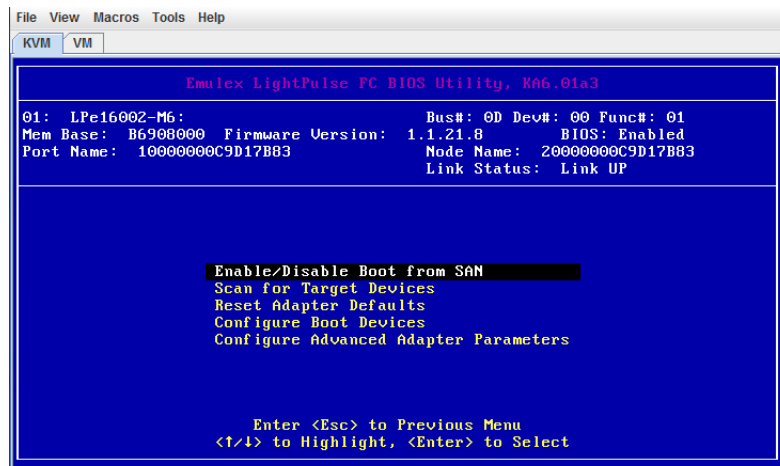


Figure 19 Selecting the Enable/Disable Boot from SAN

3. To enable the boot BIOS on the adapter port, select **Enable**, as shown in Figure 20.

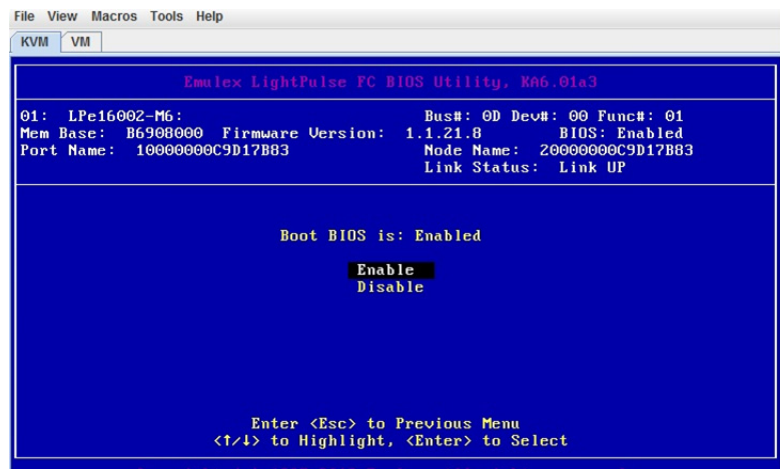
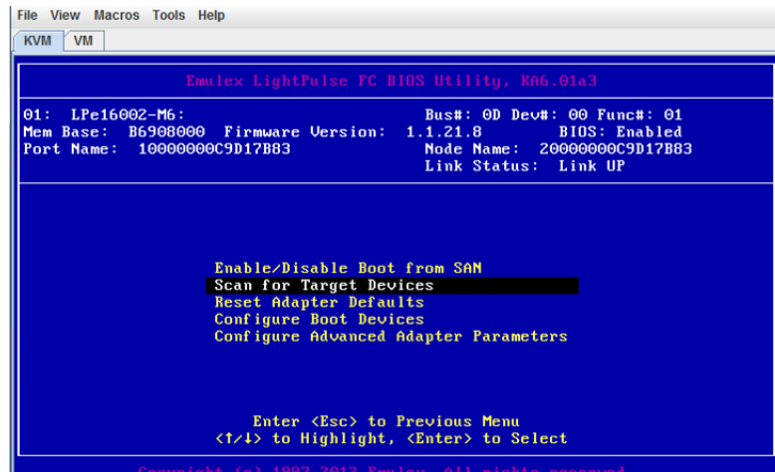


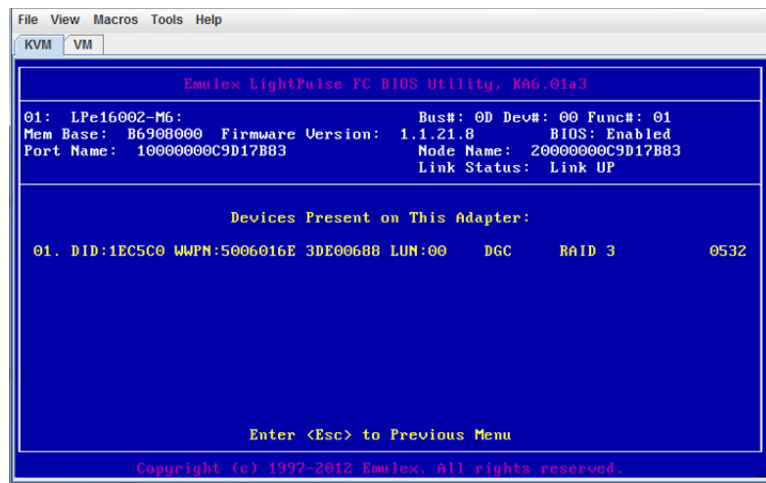
Figure 20 Enabling boot BIOS on the adapter port

4. After BIOS is enabled, scan for available target devices, as shown in Figure 21.



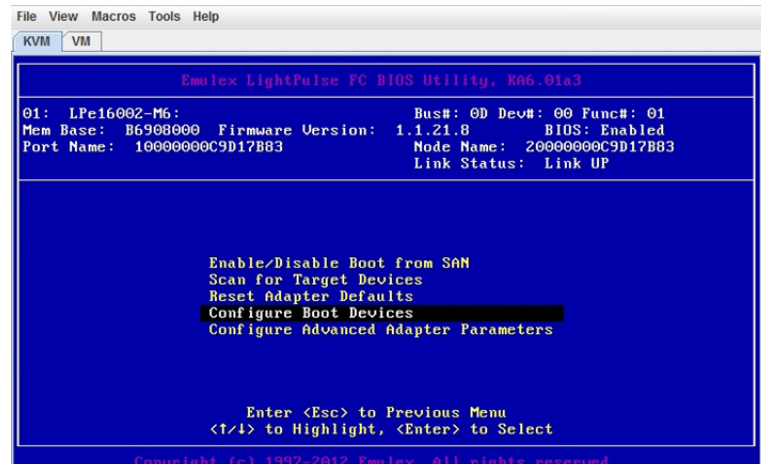
**Figure 21** Scanning for Target Devices

All of the attached LUNs should be listed after scanning. The example in Figure 22 shows the attached boot LUN, which is a VNX RAID3 LUN.



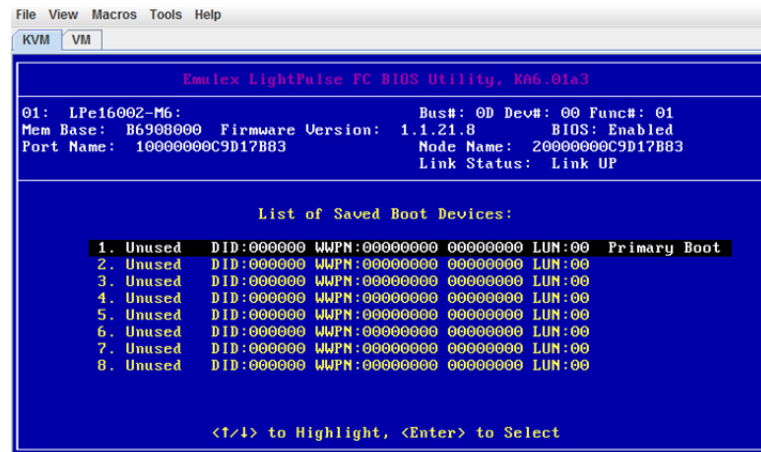
**Figure 22** Listed LUNs after scanning

- When the boot LUN is visible to the HBA, select **Configure Boot Devices**, as shown in Figure 23.



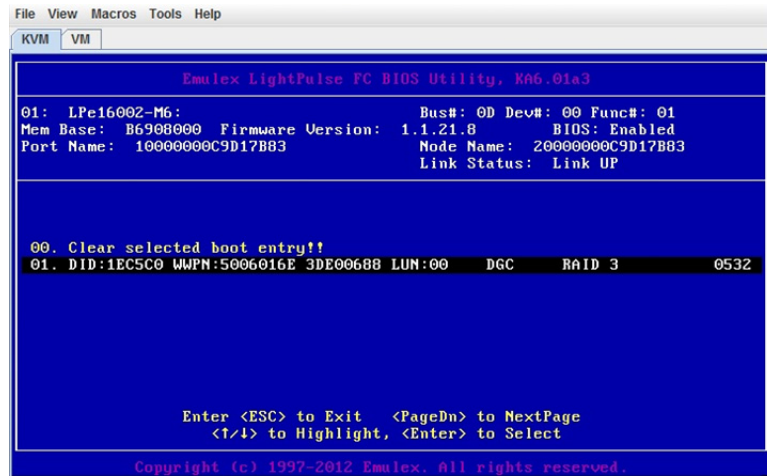
**Figure 23** Configuring the boot devices

- Review the list of boot devices and select the required LUN as the primary boot, as shown in Figure 24



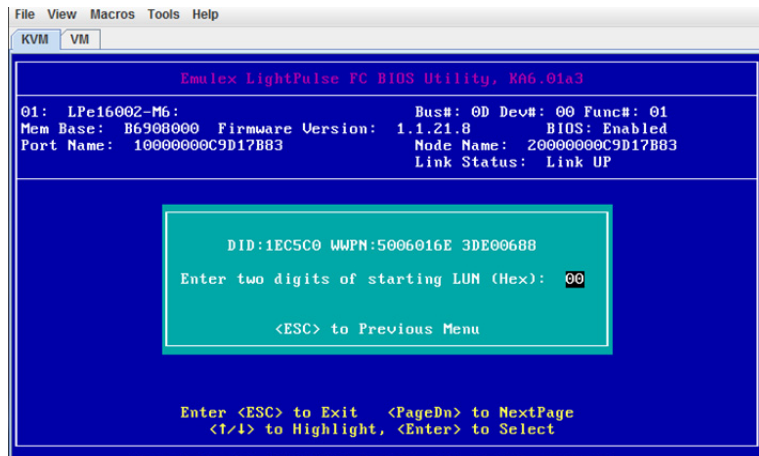
**Figure 24** Selecting a LUN as the primary boot

You can view the device for boot details as shown in Figure 25.



**Figure 25** Viewing the device for boot details

7. Set the boot LUN ID as **00**, as shown in Figure 26.



**Figure 26** Setting the LUN ID for the boot device

8. 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 up, with a BIOS is successfully installed message, as shown in [Figure 27](#). The server is then ready for OS installation.

```
Emulex LightPulse FC x86 BIOS, Version 6.01a3
copyright (c) 1997-2012 Emulex. All rights reserved.

Press <Alt E> or <Ctrl E> to enter Emulex BIOS configuration
utility. Press <S> to skip Emulex BIOS

Installing Emulex BIOS .....
Bringing the Link up, Please wait...
Link Up : Physical Link Established.
Bringing the Link up, Please wait...
Link Up : Physical Link Established.
--Adapter 1 LPe16002-M6: S_ID:240C00 PCI Bus, Device, Function (00,00,01)
DID:1EC5C0 WWPN:5006016E3DE09608 LUN: 00
--Adapter 2 LPe16002-M6: S_ID:240B00 PCI Bus, Device, Function (00,00,00)

Emulex BIOS is installed successfully!!!
```

**Figure 27** Confirming the BIOS installation after a server restart

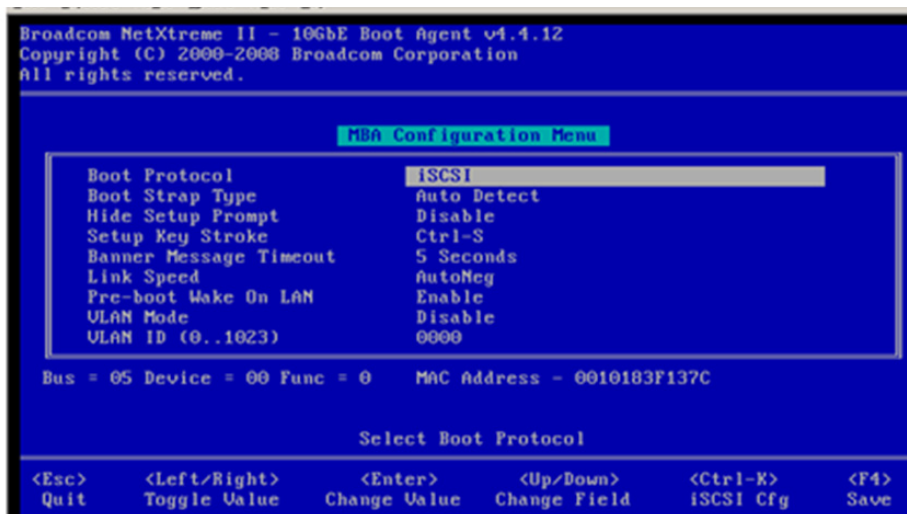
## Configuring SAN boot for iSCSI host

Systems can boot to iSCSI LUNs in primarily 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 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 HDD, using an ISO image, or using PXE server.

### Setting up the hardware iSCSI SAN boot

1. Press **Ctrl + S** when prompted to enter the Broadcom MBA setup.
2. Press **Ctrl - K** to enter the iSCSI configuration, as shown in [Figure 28](#).



**Figure 28** Entering the iSCSI configuration

---

**Note:** The figures in this section show the setup of RHEL 5.4 on a Dell server with Broadcom NICs.

---



3. Select the primary device, as shown in Figure 29.

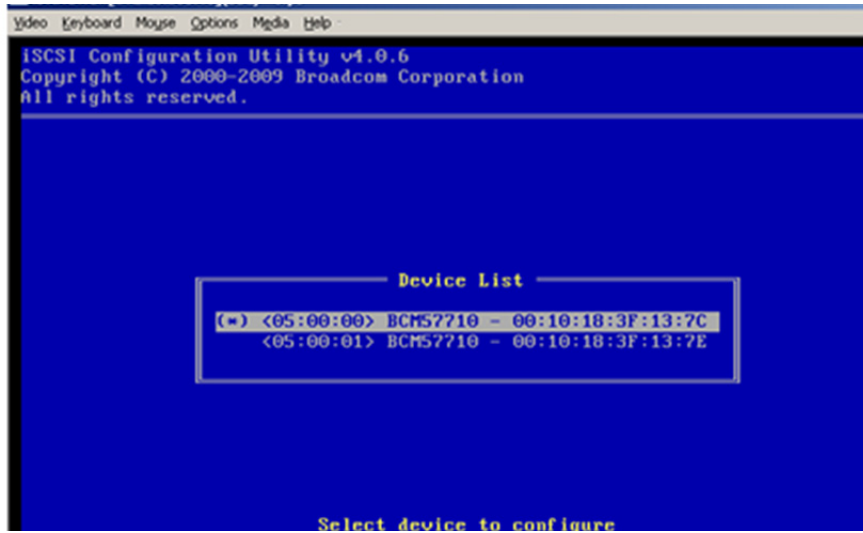


Figure 29 Selecting the primary device

4. Select **General Parameters**, as shown in Figure 30.

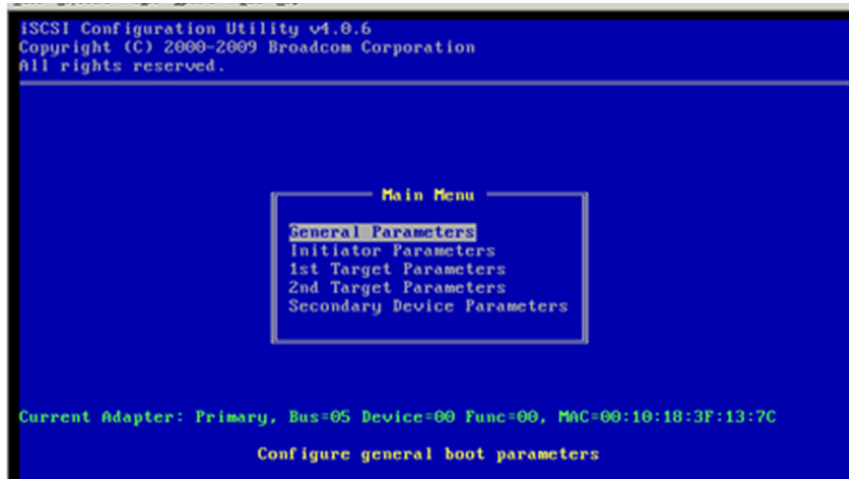


Figure 30 Selecting General Parameters

- In **General Parameters**, change the **Boot to iSCSI target** parameter from **Enabled**, as shown in Figure 31, to **Disabled**.



Figure 31 Changing the **Boot to iSCSI target** parameter

- Set up the **Initiator Parameters**, as shown in Figure 32.

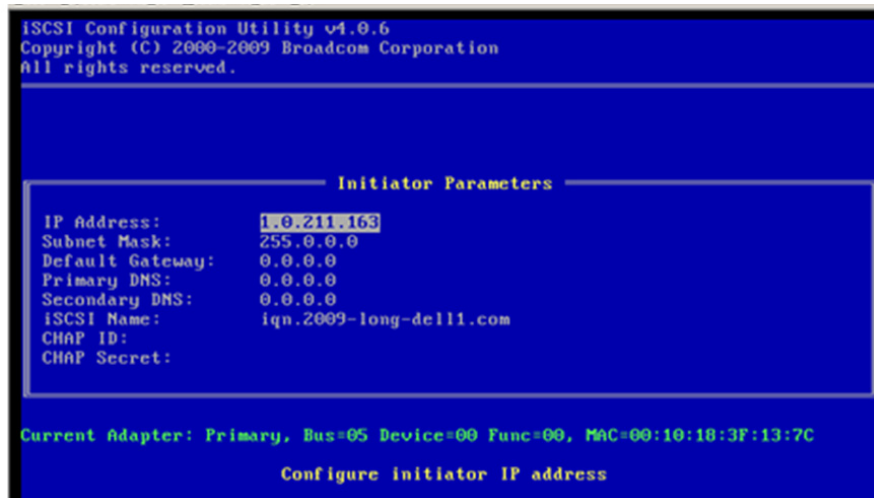


Figure 32 Setting up the Initiator Parameters

7. Set up the **1st Target Parameters**, as shown in the example in Figure 33.

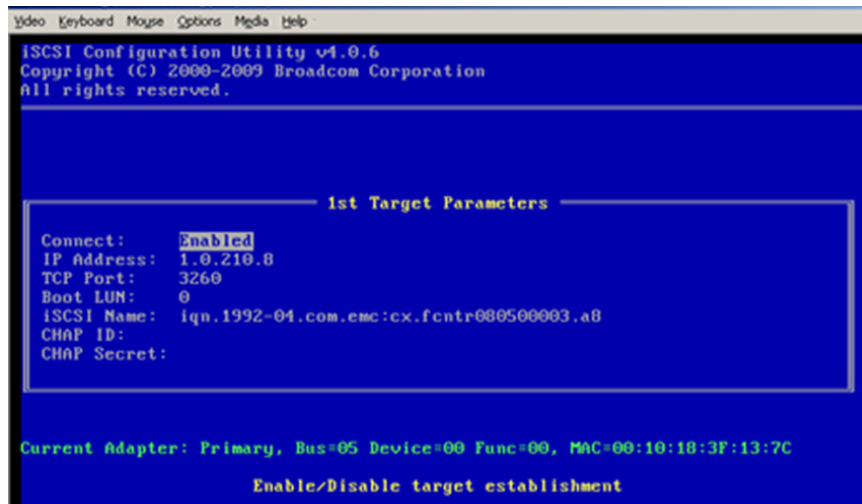


Figure 33 Setting up 1st Target Parameters

8. Set up the **Secondary Device Parameters**, as shown in the example in Figure 34.

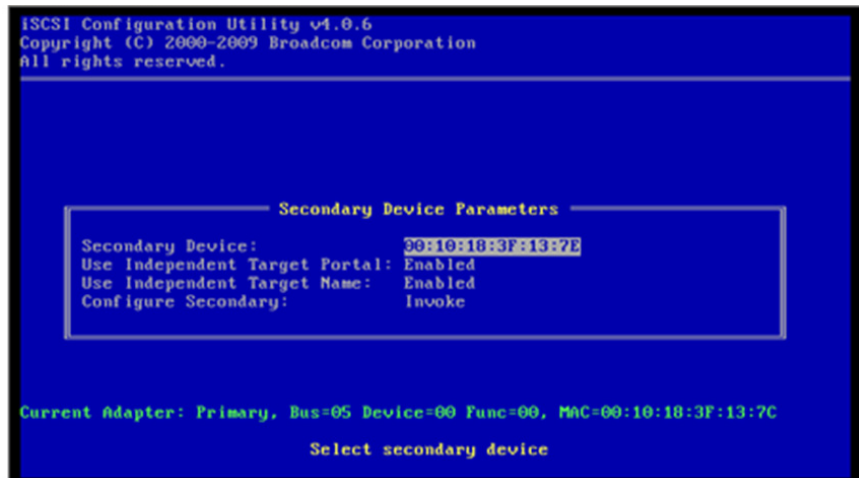


Figure 34 Setting up the Secondary Device Parameters

The following steps are the same as for the 1st target:

9. For the secondary device, select **General Parameters**, as shown in [Figure 30](#).
10. Change the **Boot to iSCSI target parameter** from Enabled to Disabled, as shown in [Figure 31](#).
11. Set up the **Initiator Parameters**, as shown in [Figure 32](#).
12. Set up the **2nd Target Parameters**, as shown in the example in [Figure 33](#).
13. Set up the **Secondary Device Parameters**, as shown in the example in [Figure 34](#).
14. Save and exit the MBA configuration.
15. Reboot the system and enter BIOS to set the boot option to the iSCSI NIC.
16. Install the OS 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 greater, NOVEL SUSE start to support it since SUSE10. Oracle Linux and Debian operating systems also support it.

---

**Note:** The figures 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:

1. Update the latest iSCSI FLB firmware. (Refer to Intel documents for reference.)
2. (Optional) You can install an OS 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 OS.

Anaconda can discover (and then log in to) iSCSI disks in the following ways:

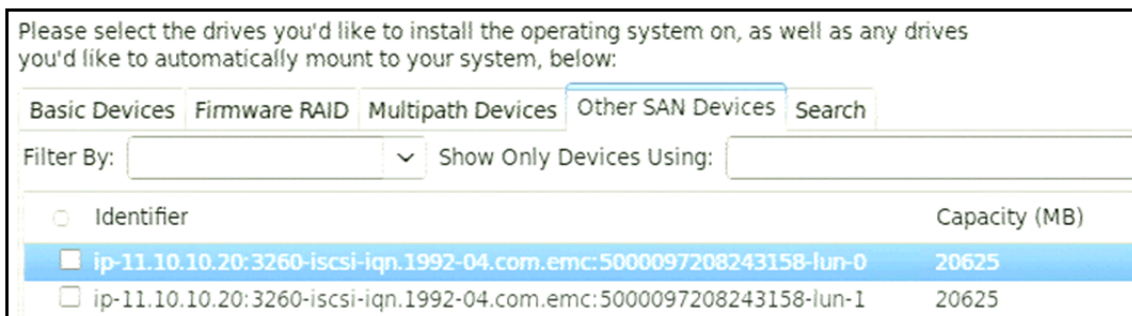
- When Anaconda starts, it checks whether the BIOS or add-on boot ROMs of the system support iSCSI Boot Firmware Table (iBFT), which is a BIOS extension for systems which can boot from iSCSI. If the BIOS supports iBFT, Anaconda will read the iSCSI target information for the configured boot disk from 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, as shown in [Figure 35](#).



**Figure 35** Selecting drives to install the OS

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`
  - Confirm that the Intel NIC IBFT firmware menu appears during the OS boot.
9. Boot the OS from the iSCSI LUN, as shown in [Figure 36](#).

---

**Note:** During the boot process, if you can see `iscsistart` messages about initiator logging into a target, as shown in [Figure 36](#), this confirms that your OS 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:
```

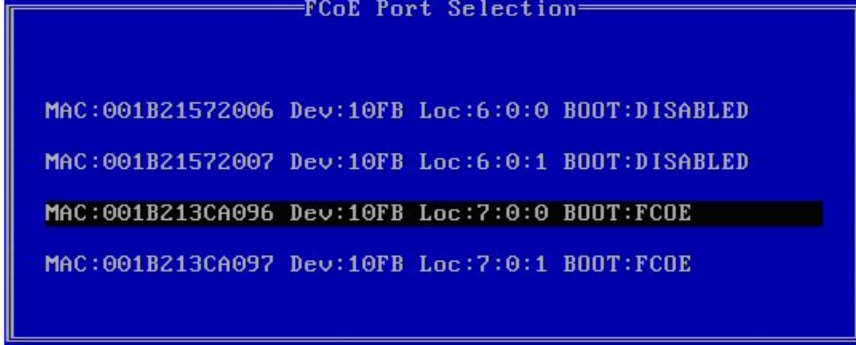
**Figure 36** Booting the OS from an iSCSI LUN

# Configuring SAN boot for FCoE attached host

## Installing and configuring Intel card for software FCoE boot

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 [Figure 37](#).



```

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
  
```

**Figure 37** Selecting the adapter for configuration

2. Select the desired port and press **Enter**.

**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, as shown in [Figure 38](#). If the **Discover VLAN** value displayed is not what you want, enter the correct value.

- 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, as shown in Figure 38.



Figure 38 FCoE Boot Targets Configuration menu

- When you are finished, press **Save** as shown in Figure 39.

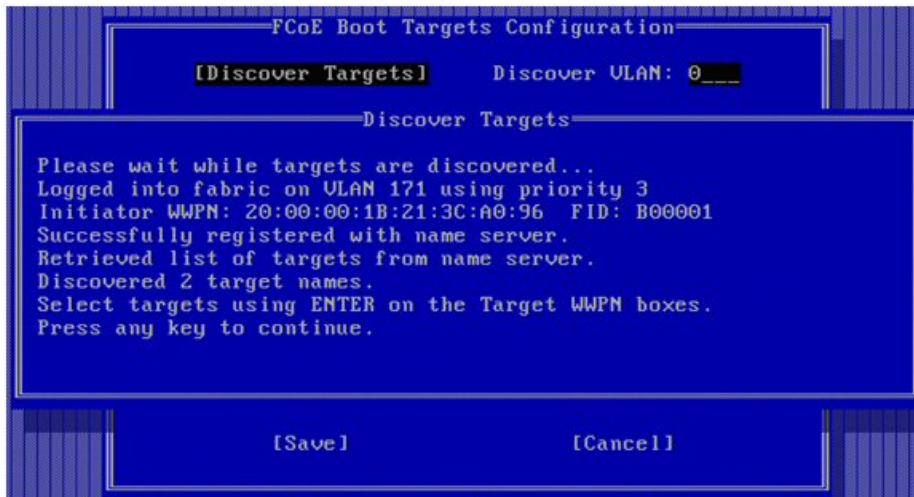


Figure 39 Discovering the remote device



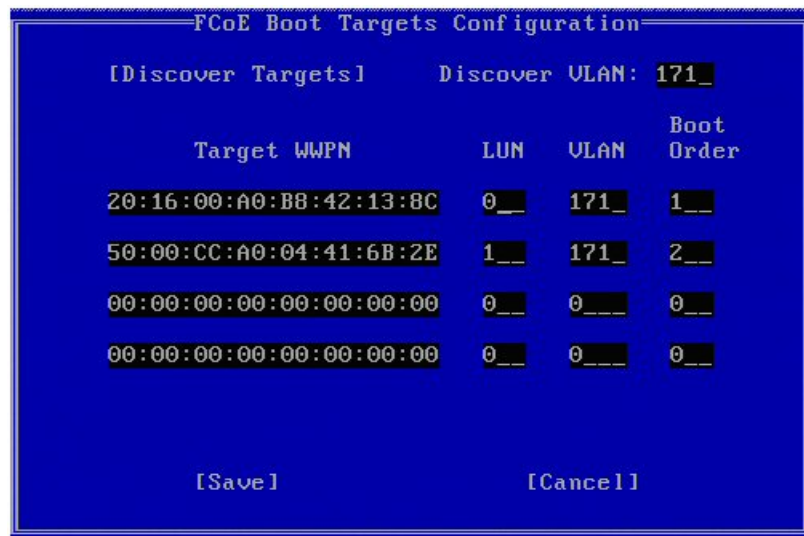
6. Select the proper device for the OS deployment by highlighting a target from the list, and then press **Enter**, as shown in Figure 40.



**Figure 40** Selecting a target from the list

7. Manually fill in the LUN and Boot Order values, as shown in Figure 41.

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.



**Figure 41** Configuring the LUN and boot order

After rebooting the server, the remote device is displayed in the BIOS, as shown in Figure 42.

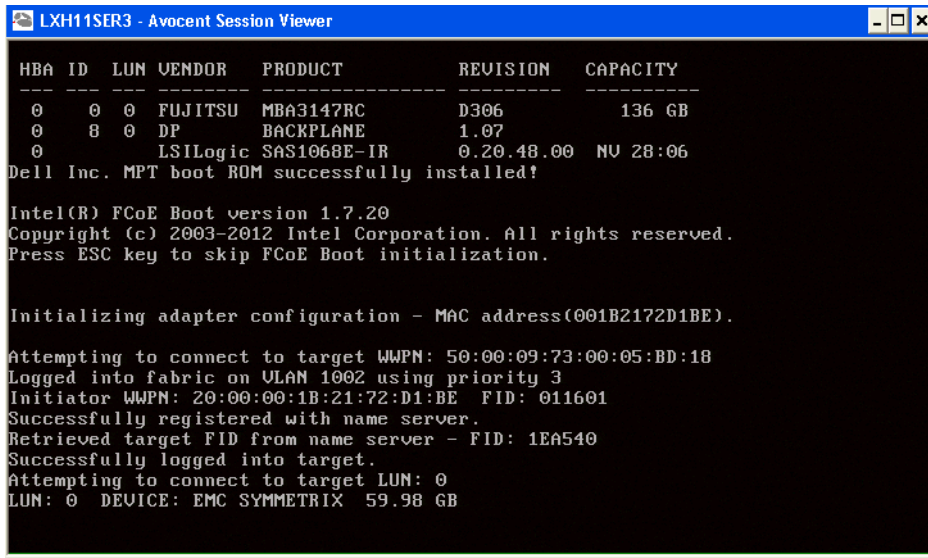


Figure 42 Remote device displayed in the BIOS after restarting the server

## Installing an OS on FCoE external devices

Installation of Linux OS on an external boot device and normal installation on internal hard disks attached to the server has one main difference: The partitions where the boot image and OS packages are installed are hosted on an external device.

To install an OS on FCoE external devices:

1. Select **Specialized Storage Devices**, as shown in [Figure 43](#)



Figure 43 Selecting a storage devices installation type

- Then select the remote device in **Other SAN Devices** or in **Multipath Devices**, as shown in Figure 44.

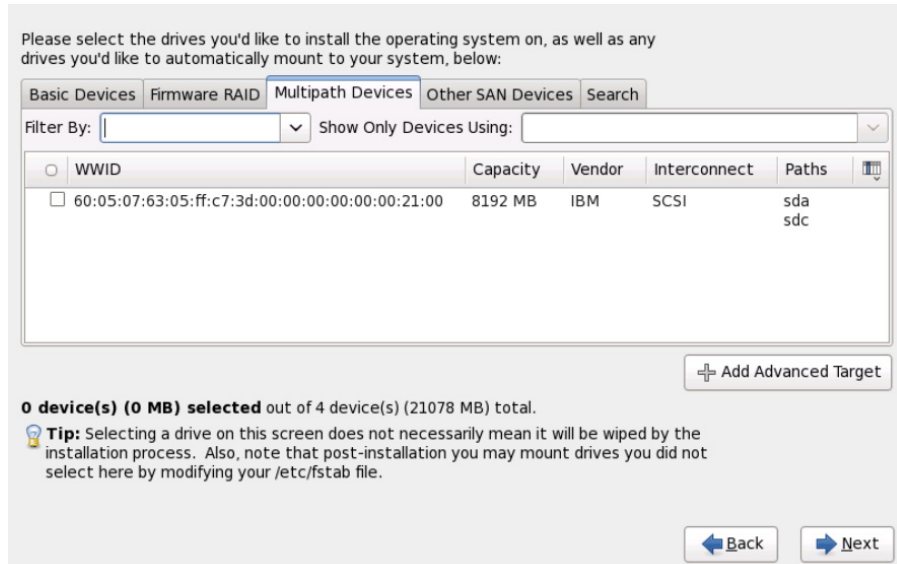


Figure 44 Selecting Multipath Devices

- Find the remote device on the **Other SAN Devices** tab and click **Add Advanced Target**, as shown in Figure 45.

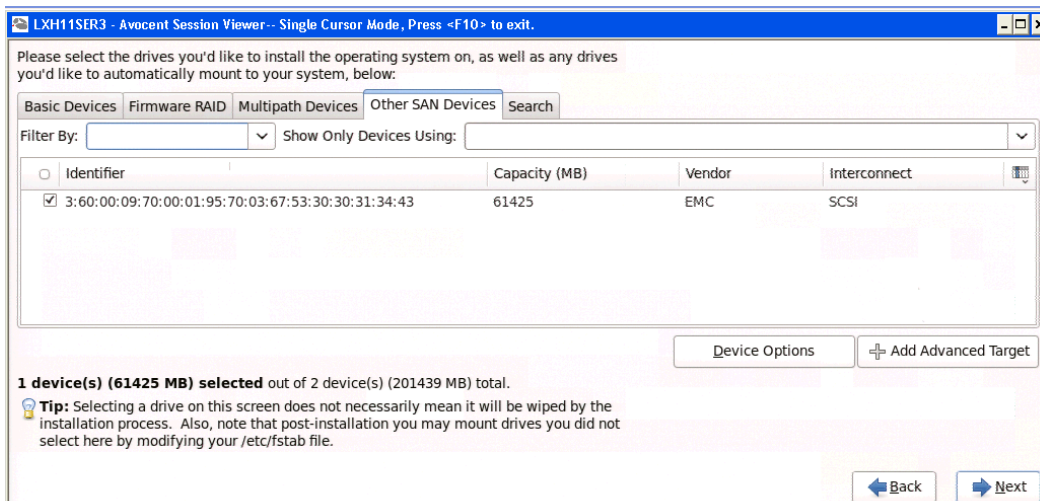
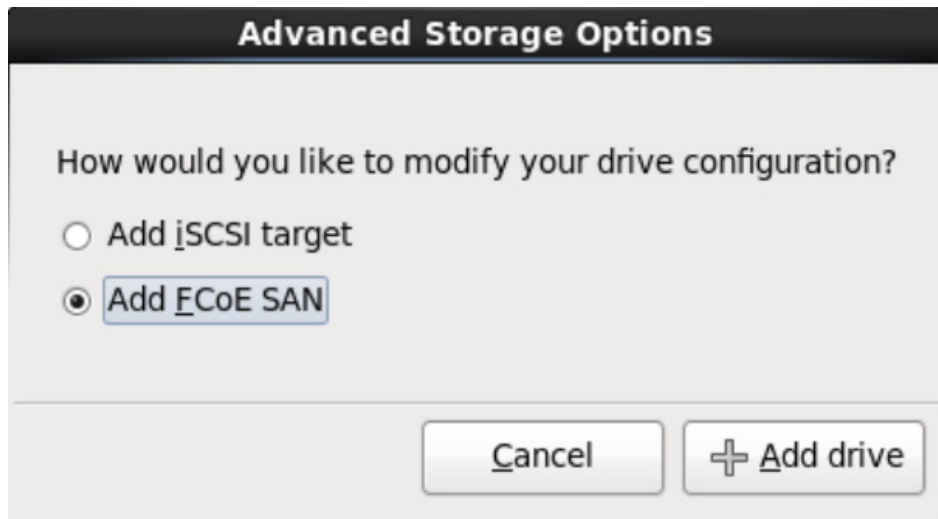


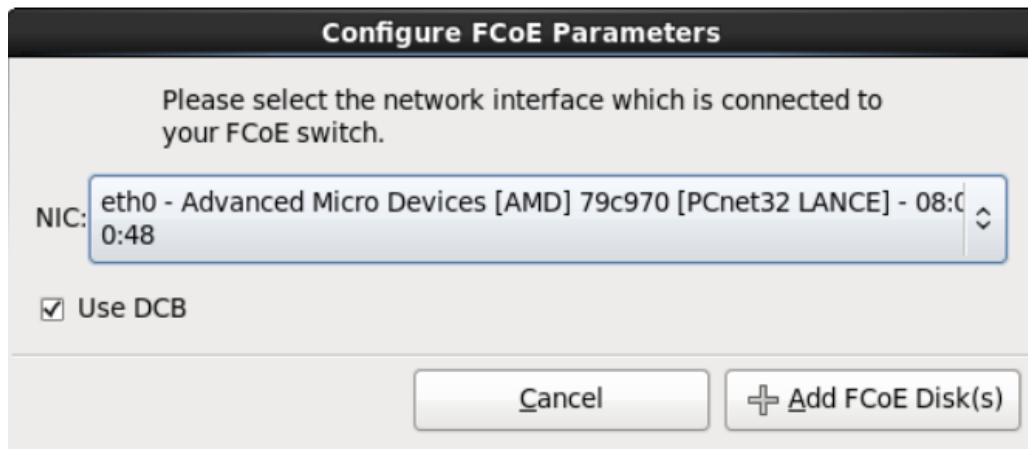
Figure 45 Selecting Other SAN devices

4. In the **Advanced Storage Options** dialog box, select **Add FCoE SAN**, and then click **add drive**, as shown in [Figure 46](#).



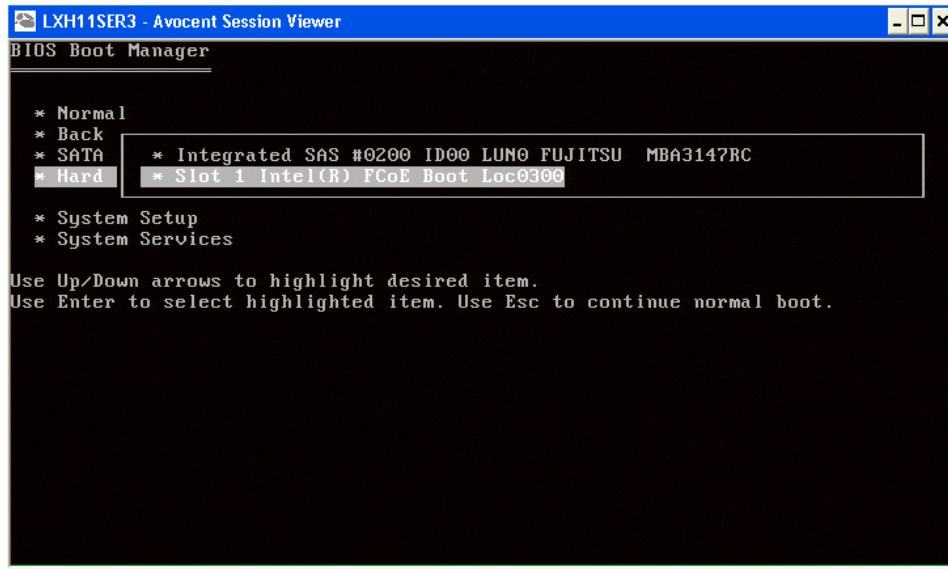
**Figure 46** Adding an FCoE device

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)**, as shown in [Figure 47](#).



**Figure 47** Configuring FCOE parameters

6. After adding the FCoE disks, you can complete the installation. After rebooting, boot from the related remote drive, as shown in Figure 48.



**Figure 48** Booting from the remote device

# Multipath booting from SAN

## IMPORTANT

Since the Linux OS will be installed on external boot device, any SCSI hard disks 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 OS version from PP 5.7.x. Booting from a SAN is not supported on Oracle VM Server 3.x.x. Refer to 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 or CLARiiON storage in PNR mode only.

Support for ALUA devices in the VNX series, VNXe series, Unity series, or CLARiiON 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 in order to ensure persistent device naming
- ☒ Use of Logical Volumes LVM2 for creation of all other required partitions

For steps on how to configure boot devices with PowerPath, refer to the [EMC PowerPath for Linux Installation and Administration Guide](#), on Dell EMC Online Support.

When using DM-MPIO to boot from a SAN environment, the detail configuration steps may vary depending on Linux distributions and OS versions. Refer to 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 OS 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 OS installation**—To, install the OS with only a single path SAN device (for example `/dev/sda`), and after the host is booted up 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 OS distribution to boot with the new DM-MPIO device name.
  - Re-make the initial `ramdisk` image.
  - Reboot the host to boot up from the DM-MPIO SAN device.

### **IMPORTANT**

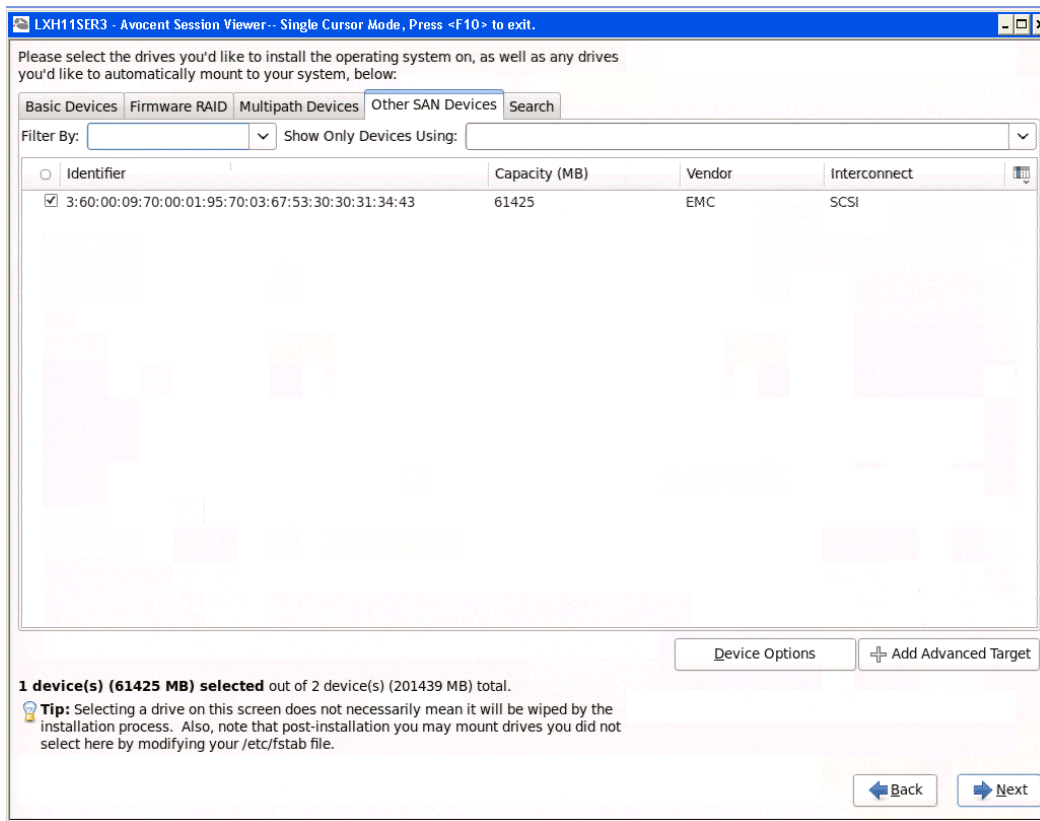
Installation differs in the following ways for Linux OS on an external boot device with normal installation on internal hard disks attached to the server:

- ☒ **Partitions where boot image and OS 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:

- ☒ To enable DM-MPIO at installation time, 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 OS to install on. **Figure 49** shows an example of a RHEL installation that uses the **linux mpath** option.



**Figure 49** 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, refer to 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.



---

**Notes:**

- ☒ The above steps listed are for RHEL 6, and the commands are different for RHEL5 and RHEL 7. For more details, refer to the Red Hat administration guide, in the "Storage/DM Multipath" chapter, in the "Moving root file systems from a single path device to multipath device" section, which is available on the [Redhat website](#).
- 
- ☒ When using DM-MPIO to boot SLES from a SAN environment:
    - To enable DM-MPIO at installation time, on the **YaST Installation Settings** page, click **Partitioning** to open the 'YaST Partitioner', select 'Customer Partitioning (for experts)', Select the 'configure Multipath' to start multipath. Make sure 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 make sure 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\_MODULES. Then re-run /sbin.mkinitrd to update the `initrd` image and reboot the server.
    - These general steps apply to SLES 11. The commands are different for SLES 12 and later. For more details, refer to the *SUSE SLES Storage Administration Guide* on the [SUSE website](#).
- 

**Best practices for ALUA mode**

This section contains ALUA boot from SAN configuration on:

- ☒ [“Configuring SLES 11 SP2” on page 149](#)
- ☒ [“Configuring RHEL 5.8” on page 150](#)
- ☒ [“Configuring RHEL 6.3” on page 150](#)
- ☒ [“Configuring RHEL 7.0” on page 151](#)

**Configuring SLES 11 SP2**

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, Unity, and CLARiiON is **1 emc**.
3. Copy a default multipath.conf file.

```
#cp /usr/share/doc/packages/multipath-tools/multipath.conf.synthetic /etc/multipath.conf
```

4. Referring to the Host Connectivity Guide 'MPIO configuration for VNX Unified Storage and CLARiiON > SuSE Linux Enterprise Server (SLES) > ALUA > ALUA SLES 10 SP2/SLES11', add the following stanza to the /etc/multipath.conf file.

```
devices {
    # Device attributed for EMC CLARiiON and 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 RHEL 5.8

1. Install RHEL 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 RHEL 6.3

1. Install RHEL 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=latacyrheb-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 RHEL 7.0

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. Because RHEL 7 start to grub2 to manage boot loader, the procedure is different as before:

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 you want to add after this comment. Be careful not to change the `exec tail` line above.

```

menuentry 'Red Hat Enterprise Linux Server, with Linux BFS ' {
    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-127cc7ec1967
    fi
    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 the 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 co-existence 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`.

---

**Note:** Refer to the appropriate *EMC PowerPath for Linux Installation and Administration Guide* for your operating system, which is available at [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, VNX series, VNXe series, Unity series, or CLARiiON

This section provides guidelines for booting from Symmetrix, XtremIO, VNX series, VNXe series, Unity series, or CLARiiON.

## 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 required to install the Linux operating system. Refer to 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 display the error message "Anaconda died after receiving signal 6" on RHEL 6.1 (and higher) with VCM/ACLX set to write enable. This could be the result of the VCM/ACLX device having been labeled as another UNIX device (e.g., SUN). Therefore, instead of using RHEL Basic Storage, select the **Specialized Storage**

option and choose your intended boot device. Dell EMC recommends that you deselect 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, Unity series, or CLARiiON-specific guidelines**

- ⊗ Prior to the installation on a VNX series or CLARiiON 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, Unity series, or CLARiiON system in order to manually register the boot server.

### **XtremIO-specific guidelines**

In order to install the OS 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 OS must be mapped to the host.

During the installation procedure, we recommend 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 above, volumes are numbered by default starting from the LUN ID 1 (and 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.



# CHAPTER 6

## Path Management

This chapter provides information on path management, including:

☒ Introduction.....	158
☒ PowerPath .....	159
☒ Veritas Dynamic Multipathing.....	160
☒ Device-mapper multipath I/O (DM-MPIO).....	161

# 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, as well as 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 in case of 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 which are instances of a single device and 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 will be discussed briefly:

- ☒ [“PowerPath” on page 159](#)
- ☒ [“Device-mapper multipath I/O \(DM-MPIO\)” on page 161](#)
- ☒ [“Veritas Dynamic Multipathing” on page 160](#)

# 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. In addition, support is 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, thereby, enhanced by automatically and intelligently optimizing data access across all available paths. PowerPath is unique in the industry in providing this intelligent multipath load balancing capability.

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

## Automatic path failover feature

PowerPath's automatic path-failover and dynamic-recovery feature permits data access to be dispersed to an alternate data path in the event of a failure. This eliminates the possibility of disrupting an application due to the failure of an adapter, cable, or channel controller. In the event of a path failure, all outstanding and subsequent I/O requests are automatically directed to the alternative path.

PowerPath's intelligent path management extends beyond failover. Instead of losing an active HBA in the event of a path failure, PowerPath fails over to an alternative path. Mission-critical applications continue without interruption, performance is optimized, and storage assets are maximized. PowerPath's 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.

### Documentation

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 via a round-robin algorithm. DMP is supported on hosts attaching to Dell EMC Fibre Channel and or iSCSI connected storage arrays. Refer to 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. Refer to the [Veritas website](#) for detailed technical and installation information.

Veritas DMP I/O load balancing is performed via 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, the use of PowerPath to control one set of storage array devices, and DMP to control another set of storage array devices within the same Linux server is not supported. This holds true for Dell EMC Fibre Channel and or iSCSI connected storage arrays. If PowerPath is installed, all storage array devices must be controlled by PowerPath.

Dell EMC supports standalone DMP failover functionality for the purposes of controlling non-VxVM volume managers. Support requires Storage Foundation 5.1 SP1 or later. Refer to the [Dell EMC Simple Support Matrix](#) for supported Linux operating system versions.

---

**Note:** On a VMAX series, the C-bit must be enabled on the array port attached to the host for DMP usage.

---

---

**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 the 2.6 kernel distributions or later, the multipathing feature has been introduced as part of the operating system. Native operating system Multipathing is also referred to as Device Mapper Multipath I/O (DM-MPIO). For supported operating systems in conjunction with DM-MPIO, refer to 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.

---

**Note:** The co-existence of DM-MPIO and PowerPath on the same host is not supported on Linux.

---

For more details, refer to [Chapter 7, "Native Multipath Failover."](#)

Operating system distributions also provide information on configuring DM-MPIO on the host, usually available in the directory `‘/usr/share/doc/’`.

Refer to the [Dell EMC Simple Support Matrix](#) and the documentation listed above for the appropriate version required on the host.



# CHAPTER 7

## Native Multipath Failover

This chapter contains the following information:

☒ Storage arrays and code revisions .....	164
☒ Supported host bus adapters .....	168
☒ Supported operating systems.....	169
☒ Server platforms.....	170
☒ DM-MPIO on IBM zSeries .....	170
☒ Configuration requirements .....	171
☒ Useful utilities .....	172
☒ Known issues .....	173
☒ MPIO configuration for VMAX series .....	178
☒ MPIO configuration for Unity storage, VNX Unified Storage, and CLARiiON 180	
☒ MPIO configuration for Dell EMC Invista or VPLEX virtualized storage	194
☒ MPIO configuring for XtremIO storage.....	197
☒ Changing the path selector algorithm .....	199
☒ Configuring LVM2.....	201
☒ Disabling Linux Multipath .....	203

## Storage arrays and code revisions

The Dell EMC storage systems listed in [Table 19](#) are supported in conjunction with Native Multipathing. [Table 19](#) also lists the required code revisions and types of multipathing support.

**Table 19** Required code revisions (page 1 of 2)

Storage array	Array code minimum requirements	Type of Multipath supported		
		Active/Active	PNR	ALUA <sup>1</sup>
VMAX All Flash 250F/FX,450F/FX,850F/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
CLARiiON CX200, CX400, CX600 CX300, CX500, CX700 CX300i, CX500i	FLARE 19	N/A	Yes, per host	Yes, per host. FLARE 26 and later CX300, CX500, CX700 CX300i, CX500i only
CLARiiON CX3-20(c), CX3-40 (c), CX3-80	FLARE 22	N/A	Yes, per host	Yes, per host FLARE 26 and later
CLARiiON CX3-10c, CX3-20f, CX3-40f	FLARE 24	N/A	Yes, per host	Yes, per host FLARE 26 and later
CLARiiON CX4-120(C8), CX4-240(C8), CX4-480(C8), CX4-960(C8)	FLARE 28	N/A	Yes, per host	Yes, per host FLARE 28 and later



**Table 19** Required code revisions (page 2 of 2)

Storage array	Array code minimum requirements	Type of Multipath supported		
		Active/Active	PNR	ALUA <sup>1</sup>
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 Management (OPM) feature
XtremIO	XIOS 2.2.1	Yes	N/A	N/A

1. Before deploying an array using Asymmetric Logical Unit Access (ALUA) feature on either Fibre Channel or iSCSI check the following:
  - Your Linux host operating system environment supports ALUA. See [Table 20 on page 166](#). Inconsistent device behaviors and or other host failures are expected if this feature is enabled on a host that does not support ALUA.
  - A failover software package is still required to make a logical unit accessible through the available host target ports. When a path failure is detected by the failover software, it is expected that the failover software will automatically reconfigure the target device to make it accessible using other available target ports.

**Note:** Always refer to the [Dell EMC Simple Support Matrix](#) for the latest supported configurations.

## VMAX series behavior

The VMAX series arrays present active accessible LUNs on all paths 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. VMAX3 and VMAX All Flash support ALUA access on Mobility ID devices from HYPERMAX OS 5977.811.784 and later.

## Unity series, VNX series, and CLARiiON behavior

Unity series and VNX series systems support active/passive and ALUA access as described below.

Prior to FLARE release 26, the CLARiiON arrays present configured LUNs as:

- ☒ *Active* on the paths connected to the service processor that is the current owner of the LUN
- ☒ *Passive* on the paths 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 will attempt 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.

With the advent of FLARE release 26, the CLARiiON array supports two types of asymmetric logical unit path management types. (Refer to [Table 20 on page 166](#) 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 with explicit asymmetric logical unit access management are capable of setting the target port group asymmetric access state of each target port group using the **SCSI Set Target Port Groups** command.

☒ **Implicit Asymmetric Logical Unit Access**

SCSI target devices with implicit asymmetric logical unit access management are capable of setting the target port group asymmetric access state of each target port group using mechanisms other than the **SCSI Set Target Port Groups** command.

The CLARiiON ALUA feature allows one port to differ from other ports connected to the same target device. The implicit access implementation allows target devices with multiple target ports to be implemented in separate physical groupings, each having designated but differing levels of access to the target ports.

The implicit 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 will be treated like any non-ALUA CLARiiON 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 non-preferred port to access a target device is that the performance may not be optimal. In addition, when a logical unit is being accessed through a non-preferred port for an extended period of time, the array may automatically attempt to trespass the logical unit so that the port is on the non-preferred SP.

Table 20 lists ALUA support in the Linux operating systems:

**Table 20** ALUA supported Linux operating systems

Linux distribution	Implicit ALUA support 1	Explicit ALUA support 2
Red Hat Enterprise Linux	RHEL 5.1 or later RHEL 6.0 or later RHEL7.0 and later	RHEL 6.0 RHEL 7.0 or 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	OL6.1 and later OL7.0 and later

1. Does not require any alteration to the `/etc/multipath.conf` file.
2. 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.

All target ports in a target port group that support asymmetric access to logical units shall be in one of the following target port asymmetric access states with respect to the ability to access a particular logical unit:

☒ **Active/Active**

While in an active/active state the target port group should be capable of accessing 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, especially those involving data transfer or caching, may operate with lower performance than they would if the target port group were in the active/non-optimized state.

☒ **Active/Standby**

While in the standby state all target ports in a target port group are capable of performing a limited set of commands. The standby state is intended to provide a state from which it should be possible to provide a higher level of accessibility, should this become necessary for any reason, to a logical unit by transitioning to either the active/active or active/enabled states.

Commands that operate in the standby state are those 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 remove logical unit or element reservations
- Testing service delivery subsystem

The policy used to handle these and other CLARiiON-specific scenarios, the *group\_by\_prio* policy, is used in conjunction with the priority being 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 fashion.

Use of the recommended Dell EMC parameters in `/etc/multipath.conf` automatically handles implicit/explicit ALUA and other policy decisions. Therefore, unless specific site conditions require alterations, the default values should be used.

## XtremIO behavior

The XtremIO presents active, accessible LUNs on all paths 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.

## 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 refer to the Linux "Base Connectivity" section of the [Dell EMC Simple Support Matrix](#) for supported HBA models and refer to 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 refer to the Linux "Base Connectivity" section of the [Dell EMC Simple Support Matrix](#) for supported CNA models and refer to 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 NIC iSCSI software initiator are supported with Dell EMC iSCSI storage arrays. Refer to 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 (RHEL) 5 and newer updates
- ☒ Red Hat Enterprise Linux (RHEL) 6 and newer updates
- ☒ Red Hat Enterprise Linux (RHEL) 7 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
- ☒ Oracle Linux (OL) Release 5 and newer updates
- ☒ Oracle Linux (OL) Release 6 and newer updates
- ☒ Oracle Linux (OL) Release 7 and newer updates
- ☒ Asianux 3.0 and newer service packs
- ☒ Asianux 4.0 and newer service packs

The following 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 refer to the subsections containing Red Hat Enterprise Linux (RHEL) equivalent release versions. Because Scientific Linux and CentOS are built from RHEL, 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 packaged with a distributed OS.

## Server platforms

The following system architectures are supported with native MPIO currently:

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

Always refer to 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 since RHEL 5 and SLES 10. DM-MPIO is also supported to run as a guest operating system starting from RHEL5 and SLES10 on z/VM 5.1 and 5.2 on the IBM zSeries. Only VMAX volumes are supported 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 a FCP device. Disks not used for the OS 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 set forth in [“Operating system limits and guidelines”](#) on page 14.
- ⊗ Boot from SAN of a multipathed LUN was introduced and supported by Dell EMC with the release of RHEL 5.1 and SLES 10 SP2. For more information, refer to [Chapter 5, “Booting From SAN.”](#)
- ⊗ Currently, 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.

---

**Note:** The `/etc/multipath.conf` configurations on the following pages are Dell EMC defaults used in the process of DM-MPIO qualification with Dell EMC storage. The built-in defaults in DM-MPIO are based on Dell EMC defaults as well.

Dell EMC recognizes that not every SAN environment will be similar to the one used in Dell EMC qualification and each customer site may need to deviate from these defaults to tune DM-MPIO for their environment. If you should deviate from these settings, Dell EMC may request you to return to the default settings as part of the root cause analysis.

---

## Useful utilities

Table 21 lists system utilities that are useful when using DM-MPIO. Refer to the man pages for detailed information and usage instructions.

**Table 21** Useful utilities

Command name	Purpose (From the respective man pages)
<b>dmsetup</b>	dmsetup manages logical devices that use the device-mapper driver.
<b>lvm</b>	lvm provides the command-line tools for LVM2.
<b>multipath</b>	multipath is used to detect multiple paths to devices for fail-over or performance reasons and coalesces them.
<b>udev</b>	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.
<b>udevinfo</b>	udevinfo queries the udev database for device information 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.
<b>udevmonitor</b>	udevmonitor listens to the kernel uevents and events sent out 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.
<b>iostat</b>	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.
<b>hotplug</b>	hotplug is a program which is used by the kernel to notify user mode software when some significant (usually hardware related) events take place.
<b>devmap_name</b>	devmap_name queries the device-mapper for the name for the device specified by major and minor number.
<b>kpartx</b>	This tool reads partition tables on specified device and creates device maps over partitions segments detected.
<b>scsi_id</b>	scsi_id queries a SCSI device via the SCSI INQUIRY vital product data (VPD) page 0x80 or 0x83 and uses the resulting data to generate a value that is unique across all SCSI devices that properly support page 0x80 or page 0x83.
<b>lsblk</b>	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

Table 22 lists known issues in native multipath failover, along with workarounds.

Table 22 Known issues (page 1 of 4)

Problem description	Environments	Workaround / Fix
Device-mapper (dm) names and sd names may not be persistent across reboots.	All	Persistence can be achieved by: <ul style="list-style-type: none"> <li>☒ Use of LVM on top of the DM names.</li> <li>☒ Use of scsi_id based names.</li> <li>☒ Use of user-friendly multipath names.</li> <li>☒ Use of device aliases defined in /etc/multipath.conf.</li> </ul>
During a host reboot having unfractured SnapView clones in a host's storage group, the server may hang.	First seen in SLES 10	The workaround is to fracture or remove any unfractured cloned LUNs from the host storage group before that host is rebooted.
In a cluster configured with a VNX series, VNXe series, Unity series, or CLARiiON using ALUA, there is no follow-over capability; therefore, one host may fail the LUN back to the original SP after the host with the failing path has trespassed the LUN to the secondary SP.	All RedHat, Asianux, Oracle OL versions; also, all SLES 9 and all SLES10	See Red Hat BZ #437107 and Novell BZ #370057 and #502626.  As a workaround in the VNX series, VNXe series, Unity series, or CLARiiON storage group, change the default owner of the LUN to a storage processor accessible by all hosts attached to the storage group.  This is resolved in RHEL 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, Unity series, or CLARiiON systems resulting in lower performance.	All RedHat, Asianux, Oracle OEL versions; also SLES 9 and SLES10 SP1, SP2, and SP3	In the VNX series, VNXe series, Unity series, or CLARiiON storage group, change the default owner of the LUN to a storage processor accessible by all hosts attached to the storage group under non-failure conditions.  This is resolved in RHEL 5.5, OEL 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.
Multipath configuration with user-friendly names results in multiple distinct devices being addressed with the same "mpath" name.	SLES 10	Edit the file /var/lib/multipath/bindings and delete all duplicate occurrences of the same user-friendly name. Execute the multipath command for reconfiguring all devices.

Table 22 Known issues (page 2 of 4)

Problem description	Environments	Workaround / Fix
During an out of family code load (NDU) on a Symmetrix array's directors to the LUN may go off line at the same time. This will prevent access to the LUN and IO will be failed immediately when using the default DM-MPIO settings for Symmetrix.	Refer to Dell EMC Knowledge base solution emc212937.	<p>To work-around this issue configure no_path_retry to queue. For example:</p> <pre> devices { ## Device attributes for EMC SYMMETRIX     device {         vendor          "EMC"         product         "SYMMETRIX"         no_path_retry   queue     } } </pre> <p>This will cause IO to pend until the path is restored. Another solution is to queue for a number of retrys such as the following:</p> <pre> devices { ## Device attributes for EMC SYMMETRIX     device {         vendor          "EMC"         product         "SYMMETRIX"         path_checker    tur         polling_interval 5         no_path_retry   6     } } </pre> <p>These are tunable parameters and can be adjusted to accommodate the particular environment for the results you wish to achieve.</p>
In SLES 11, when a path fails, all references to the path and the devices are completely removed from the kernel OS.	SLES 11	This is a change in the behavior of the Linux kernel. Expect that if a path fails it will be removed from the <b>multipath -l</b> command's output.
When 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 causes I/O to hang.	SLES 11- multipath-tools-0.4.8-40.4.1	<p>Bz 518007 - Re-insertion of the cables will not re-enable the paths. Cable pulls with greater than 10 seconds apart do not exhibit this behavior.</p> <p>Fixed in multipath-tools-0.4.8-40.5</p>

Table 22 Known issues (page 3 of 4)

Problem description	Environments	Workaround / Fix
Dell EMC Knowledgebase solution #emc227052  Dell EMC Invista requires its own specific configuration stanza in /etc/multipath.conf	All supported releases	Add: <pre>devices { ## Device attributes for EMC Invista     device {         vendor          "EMC"         product         "Invista"         path_checker    tur         no_path_retry   5         product_blacklist "LUNZ"     } }</pre> Resolved as a default in RHEL 5.5 and higher, SLES 10 SP3 and higher, and SLES 11 SP1 and higher, thereby not requiring the addition of this stanza to support Invista.
VPLEX requires its own specific configuration stanza in /etc/multipath.conf	All supported releases	Add: <pre>devices { ## Device attributes for EMC Invista     device {         vendor          "EMC"         product         "Invista"         path_checker    tur         no_path_retry   5         product_blacklist "LUNZ"     } }</pre> Resolved as a default in RHEL 5.5 and higher, SLES 10 SP3 and higher, and SLES 11 SP1 and higher, thereby not requiring the addition of this stanza to support VPLEX.
Novell BZ # 254644 - Root device not managed by DM-MPIO on VNX series or CLARiiON system.	SLES 10	This can occur if you are attempting to configure a large amount of LUNs attached to the server and you run out of file descriptors (fd). Currently the limit for the maximum number of open files is 1024 (ulimit -n).  As multipath and multipathd requires one fd for each path you have to increase this limit to roughly 2 times the number of paths. You will need to edit the script /etc/init.d/multipathd and change the variable MAX_OPEN_FDS to a number greater than it currently uses.
Red Hat BZ #509095 - inconsistent multipath maps following storage addition.	RHEL 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 filesystem, this option should be set to a path on the same device as the root filesystem, for example /etc/multipath_bindings.  By setting this, multipath will use the same bindings during boot as it does during normal operation.

Table 22 Known issues (page 4 of 4)

Problem description	Environments	Workaround / Fix
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 <a href="#">“Supported operating systems” on page 169</a> .	Use this device stanza: <pre>device {     vendor "EMC"     product "Celerra"     path_grouping_policy "multibus"     path_checker "tur"     no_path_retry "30" }</pre> 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 # 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 RHEL 6.3  ( <a href="#">Redhat Customer Portal 6.3. device-mapper -multipath technical notes</a> )  The multipathd daemon did not have a failover method to handle switching of path groups when multiple nodes were using the same storage. Consequently, 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 followover failback method to device-mapper-multipath. If the followover failback method is set, multipathd does not fail back to the preferred path group, unless it just came back online. When multiple nodes are using the same storage, a path failing on one machine now no longer causes the path groups to continually switch back and forth.
SUSE BZ #802456 - Dell EMC active/active ALUA wasn't driving IO down all paths.	First seen in SLES 11; not seen in RHEL.	Fixed in SLES 11 SP2 patch multipath-tools 8339 (multipath-tools-0.4.9-0.70.72.1 & kpartx-0.4.9-0.70.72.1).

---

**Note:** On VNX series, VNXe series, Unity series, or CLARiiON systems, 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.

---



---

**Note:** VMAX series Layered Applications are supported in conjunction with native multipathing. Include any and all 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 required.

---

---

**Note:** The VNX series or CLARiiON Layered Application SnapView snapshot and the admsnap host utility are supported in Linux native multipathing environments when performed with CLARiiON Release 29 FLARE and layered applications. Support in Linux native multipathing environments is also available for MirrorView, SAN Copy, and SnapView clones.

---

Unfractured SnapView clones are *not* supported in active storage groups in the Linux native multipathing environment. During a host reboot having unfractured SnapView clones in a host's storage group may cause the host to hang. Dell EMC is currently working with the DM-MPIO development community to resolve this problem. Until this can be resolved the workaround is to fracture or remove any unfractured cloned LUNs from the host storage group before that host is rebooted.

## MPIO configuration for VMAX series

This section discusses methods for configuring the VMAX series arrays for the following:

- ☒ “RedHat Enterprise Linux (RHEL)” on page 178
- ☒ “Oracle Linux and VM server” on page 179
- ☒ “SuSE Linux Enterprise server” on page 179

### RedHat Enterprise Linux (RHEL)

The following sections detail the procedure for configuring native multipath failover for Symmetrix family of arrays in a RHEL host.

The RHEL native MPIO already contains default configuration parameters for 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.

This is a sample output from `multipath -ll`:

```
mpath15 (360060480000190101965533030423744) EMC,SYMMETRIX
[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]
```

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 VCM database, as in this example. The VCM database is a read-only device that is used by the array. By blacklisting it you will eliminate any error messages that may occur due to its presence.

```
blacklist {
    wwid 360060480000190101965533030303230
    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 or Oracle enhanced Red Hat kernel use the same configurations as Red Hat Enterprise Linux for the VMAX series arrays because they share the same kernel.

## SuSE Linux Enterprise server

This section details the procedure for configuring native multipath failover (MPIO) for the VMAX series of arrays on a SuSE Linux Enterprise Server (SLES) host.

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-multipath` package.

This is a sample output from `multipath -ll`:

```
# multipath -ll
mpath27 (360060480000190100501533031353831) EMC,SYMMETRIX
[size=468M][features=0][hwhandler=0]
\_ round-robin 0 [prio=4][undef]
\_ 11:0:0:39 sdb 68:80 [undef][ready]
\_ 11:0:1:39 sdcc 69:0 [undef][ready]
\_ 10:0:0:39 sdl 8:176 [undef][ready]
\_ 10:0:1:39 sdw 65:96 [undef][ready]
```

1. If you want multipath devices to be created as `/dev/mapper/mpathn`, set Value of `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 VCM database, as in this example. The VCM database is a read-only device that is used by the array. By blacklisting it you will 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"
```

# MPIO configuration for Unity storage, VNX Unified Storage, and CLARiiON

This section contains the following information for configuring the Unity series, VNX (Unified Storage) series, and CLARiiON storage arrays.

- ☒ “Blacklisting the Unity series, VNX series, or CLARiiON LUNZ” on page 180
- ☒ “Failover mode” on page 180
- ☒ “Red Hat Enterprise Linux (RHEL)” on page 181
- ☒ “Oracle Linux and VM Server” on page 188
- ☒ “SuSE Linux Enterprise Server (SLES)” on page 188

## Blacklisting the Unity series, VNX series, or CLARiiON LUNZ

When zoned to a Unity series, VNX series, or CLARiiON, a host may not boot if a LUNZ is exposed to the system. This has been fixed by allowing DM-MPIO to blacklist and, thus, skip over these devices. In order 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 RHEL 4.7 and later and RHEL 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, refer to 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, VNX series, and CLARiiON support Asymmetrical Logical Unit Access (ALUA) when configured for failover mode 4. Refer to [“Unity series, VNX series, and CLARiiON behavior” on page 165](#) for more information on this functionality.

---

**Note:** Dell EMC does not currently support having PNR and ALUA devices attached to the same host for ALUA devices in a Unity series, VNX series, and CLARiiON storage systems.

---

## Red Hat Enterprise Linux (RHEL)

The section discusses methods for configuring the Unity series, VNX (Unified Storage) series, and CLARiiON storage arrays. RHEL 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 following device stanza is the default PNR stanza that is automatically included in RHEL 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 CLARiiON and 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
}
}
```

This is sample output from `multipath -ll`:

```
mpath3 (36006016005d01800b207271bb8ecda11) DGC,RAID 5
[size=3.1G][features=1 queue_if_no_path][hw_handler=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 CLARiiON and 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"
}
}

```

This is sample output from `multipath -ll` after these changes have been made:

```

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 Linux 6.0 (and point releases)

Only RHEL6.5 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 CLARiiON and 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_selector "round-robin 0"
    path_checker emc_clariion
    features "1 queue_if_no_path"
    hardware_handler "1 emc"
    prio emc
    failback immediate
    rr_weight unifor
    no_path_retry 60
  }
}

```

This is a sample output from `multipath -ll` after these changes have been made:

```

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_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
  }
}
```

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_selector "round-robin 0"

path_checker emc_clariion

features "1 queue_if_no_path"

hardware_handler "1 alua"

prio tpg_pref

failback immediate

no_path_retry 300
}
}

```

This is a sample output from `multipath -ll` after these changes have been made:

```

# 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 Linux 7.0 (and point releases)

For RHEL7.0 and later releases, if you have PNR, make the following updates to the `multipath.conf` file:

---

**Note:** The default configuration parameters for the Unity series, VNX series, and CLARiiON arrays are included as part of the multipath package; hence they do not need to be separately defined.

---

### IMPORTANT

The following stanza is only an example. Consult Red Hat's documentation to ensure the correct syntax is followed for your release. MPIO continues to evolve with each release of Red Hat.

---

```

## 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 output of `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_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
}
}

```

Here is sample output from `multipath -ll` after these changes have been made:

```

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

```

## RHEL7.2 and later

Since RHEL7.2, the default stanza parameter `retain_attached_hw_handler` `yes` can recognize PNR and ALUA automatically and does not need changes to the `multipath.conf` file. However, the manual configuration of PNR and ALUA will still work.

### IMPORTANT

The following stanza is only an example. Consult Red Hat's documentation to ensure the correct syntax is followed for your release. MPIO continues to evolve with each release of Red Hat.

```

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 that use the stock Red Hat kernel or Oracle enhanced Red Hat kernel use the same configurations as Red Hat Enterprise Linux. Refer to RHEL configuration on the previous page.

## SuSE Linux Enterprise Server (SLES)

The section discusses methods for configuring the Unity series, VNX (Unified Storage) series, and CLARiiON storage arrays.

### SLES 10 (and point releases)

If you have PNR, make the following updates to the `multipath.conf` file.

On a SLES 10 system, the default configuration parameters for the Unity series, VNX series, and CLARiiON arrays are included as part of the multipath package, so that they do not need to be separately defined.

### IMPORTANT

The following stanza is only an example. Consult SuSE documentation to ensure the correct syntax is followed for your release. MPIIO continues to evolve with each release of SLES.

```

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
}

```

Here is sample output from `multipath -ll`:

```

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 support in SLES 10 and SLES 10 SP1. Only implicit ALUA mode is supported. Refer to [Table 20 on page 166](#) for details.

☒ **SLES10SP2 and later:**

Explicit ALUA is only supported with SLES 10 SP2 and greater, refer to [table 27](#). You only need to perform the following modifications for explicit ALUA support.

```
defaults {
    user_friendly_names    yes
}
devices {
# Device attributes for EMC CLARiion and 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
    }
}
```

This is sample output from `multipath -ll` after these changes have been made:

```
mpathf (3600601601bf024006e16cd89575fde11) dm-1 DGC,RAID 5
[size=5.0G][features=1 queue_if_no_path][hw_handler=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 PNR, make the following updates to the `multipath.conf` file:

On a SLES 11 system, the default configuration parameters for the Unity series, VNX series, and CLARiiON arrays are included as part of the multipath package, so that they do not need to be separately defined.

**IMPORTANT**

The following stanza is only an example. Consult SuSE documentation to ensure the correct syntax is followed for your release. MPIO continues to evolve with each release of SLES.

```
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_selector         "round-robin 0"
    path_grouping_policy  group_by_prio
    failback              immediate
    rr_weight             uniform
    no_path_retry         60
    rr_min_io             1000
    path_checker          emc_clariion
    prio                 emc
}
```

This is a sample output from `multipath -ll`:

```
mpathb (360060160782918000be69f3182b2da11) DGC,RAID 3
[size=3G][features=1 queue_if_no_path][hw_handler=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:

**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 CLARiiON and VNX series ALUA
    device {
        vendor          "DGC"
        product         "*"
        hardware_handler "1 alua"
        prio            ""alua""
    }
}
```

This is sample output from `multipath -ll` after these changes have been made:

```
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 PNR, make the following updates to the `multipath.conf` file:

**Note:** On a SLES 12 system, the default configuration parameters for the Unity series, VNX series, and CLARiiON arrays are included as part of the `multipath` package, so they do not need to be separately defined.

### IMPORTANT

The following stanza is only an example. Consult SuSE documentation to ensure the correct syntax is followed for your release. MPIIO continues to evolve with each release of SLES.

```
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
dev_loss_tmo 0
}

```

This is sample output from `multipath -ll`:

```

Mpathd (36006016021003500e04ad9cc785fe411) undef DGC,VRAID
size=20G features='1 queue_if_no_path' hwandler='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_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
    dev_loss_tmo 0
}

```

Here is sample output from `multipath -ll` after these changes have been made:

```

mpathgd (36006016021003500e04ad9cc785fe411) dm-4 DGC,VRAID
size=20G features='1 queue_if_no_path' hwandler='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
```

## MPIO configuration for Dell EMC Invista or VPLEX virtualized storage

This section discusses methods for configuring Invista and VPLEX virtualized storage.

- ☒ “Red Hat Enterprise Linux (RHEL)” on page 194
- ☒ “Oracle Linux and VM Server” on page 195
- ☒ “SuSE Linux Enterprise Server (SLES)” on page 195
- ☒ “OPM” on page 196

### Red Hat Enterprise Linux (RHEL)

The Linux native MPIO since RHEL5.5 already contains default configuration parameters for Invista and 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.

#### **IMPORTANT**

The following 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 Invista/VPLEX
    device {
        vendor            "EMC"
        product           "Invista"
        path_checker      tur
        no_path_retry     5
        product_blacklist "LUNZ"
    }
}
```

This is sample output from `multipath -ll`:

```
# multipath -v2 -d
create: mpath15 (360060480000190101965533030423744) EMC,INVISTA
[size=8.4G][features=0][hw_handler=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 or Oracle enhanced Red Hat kernel use the same configurations as Red Hat Enterprise Linux for Invista or VPLEX virtualized storage. Refer [“Red Hat Enterprise Linux \(RHEL\)” on page 194](#) configuration for more information.

## SuSE Linux Enterprise Server (SLES)

Since SLES10 SP3, the Linux native MPIO already contains default configuration parameters for Invista and 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.

### **IMPORTANT**

The following is only an example. Consult SuSE documentation to ensure the correct syntax is followed for your release. MPIO continues to evolve with each release of SLES.

```
##
defaults {
    user_friendly_names    yes
}

devices {
    ##      Device attributes for EMC SYMMETRIX
    device {
        vendor              "EMC  "
        product             "INVISTA"
        path_grouping_policy    multibus
        path_checker        tur
        no_path_retry       5
        product_blacklist   "LUNZ"
    }
}
}
```

This is the `multipath.conf` output:

```
mpath27 (360060480000190100501533031353831) EMC,INVISTA
[size=468M][features=0][hwhandler=0]
\_ round-robin 0 [prio=4][undef]
\_ 11:0:0:39 sdb 68:80 [undef][ready]
\_ 11:0:1:39 sdcc 69:0 [undef][ready]
\_ 10:0:0:39 sdl 8:176 [undef][ready]
\_ 10:0:1:39 sdw 65:96 [undef][ready]
```

## OPM

Starting with VPLEX 5.5 Acropolis, Optimal Path Management (OPM) was introduced to improve VPLEX performance. OPM utilizes an ALUA mechanism to spread loads across VPLEX directors while gaining cache locality. Refer to [“Optimal-Path-Management \(OPM\) feature” on page 282](#) for details.



## 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 with the following parameters:

### Red Hat Enterprise Linux (RHEL)

**Note:** Device Mapper Multipath support and default configuration for an XtremIO storage array is provided with RHEL 7, `device-mapper-multipath` version: 0.4.9-77.el7.x86\_64 and later. All previous versions of RHEL, `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
}
```

The 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. Please refer to RHEL configuration on the previous page.

## SuSE Linux Enterprise Server (SLES)

### IMPORTANT

The following is only an example. Consult SuSE documentation to ensure the correct syntax is followed for your release. MPIO continues to evolve with each release of SLES.

```
device {
    vendor "XtremIO"
    product "XtremApp"
    path_grouping_policy "multibus"
    path_selector "queue-length 0"
    path_checker "tur"
    features "1 queue_if_no_path"
    hardware_handler "0"
    prio "const"
    failback "immediate"
    rr_weight "uniform"
}
```

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 sdfs 68:96 active ready running
```

## Changing the path selector algorithm

With the release of RHEL 6.0, two new path selector algorithms were added to DM-MPIO. This can be set as the default path selector algorithm to use for all your connected storage by setting it in the *defaults* stanza or on an individual storage device in the *device* stanza. The default path selector is *round-robin*.

There are three selector algorithms.

- ☒ `round-robin 0`

Loop through every path in the path group, sending the same amount of IO to each.

- ☒ `queue-length 0`

Send the next bunch of IO down the path with the least amount of outstanding IO.

- ☒ `service-time 0`

Choose the path for the next bunch of IO based on the amount of outstanding IO to the path and its relative throughput.

Dell EMC storage has the default round-robin applied to their default settings; therefore, to change to an alternate path selector you will need to create a *device* stanza for the Dell EMC product you want to alter. If you change the path selector in the *defaults* stanza, it will not take effect. Refer to the following Symmetrix, VNX series, or CLARiiON, and Invista/VPLEX examples.

### Symmetrix:

```
device {
    vendor "EMC"
    product "SYMMETRIX"
    path_grouping_policy multibus
    getuid_callout "/lib/udev/scsi_id --page=pre-spc3-83 --whitelisted --device=/dev/%n"
    path_selector "service-time 0"
    path_checker directio
    features "0"
    hardware_handler "0"
    prio const
    rr_weight uniform
    rr_min_io 1000
}
```

### VNX series or CLARiiON:

```
device {
    vendor "DGC"
    product ".*"
    product_blacklist "LUNZ"
    path_grouping_policy group_by_prio
    getuid_callout "/lib/udev/scsi_id --whitelisted --device=/dev/%n"
    path_selector "service-time 0"
    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
    rr_min_io 1000
}
```

**Invista/VPLEX:**

```

device {
    vendor                "EMC"
    product               "Invista"
    product_blacklist    "LUNZ"
    getuid_callout       "/lib/udev/scsi_id --whitelisted --page=pre-spc3-83"
--device="/dev/%n"
    features              "0"
    hardware_handler     "0"
    path_selector        "queue-length 0"
    path_grouping_policy multibus
    rr_weight             uniform
    no_path_retry        5
    rr_min_io            1000
    path_checker         tur
    prio                 const
}

```

## Configuring LVM2

When using LVM, it is recommended that Logical Volumes be created on DM-MPIO devices instead of SCSI devices for the following reasons:

- ⊗ In a multipath environment, more than one SCSI sd device points 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 specifically tied to particular sd devices instead of the multipath infrastructure. This may result in multipath infrastructure not providing failover capabilities in the event of a SAN failure or device unavailability.
- ⊗ The SCSI device (sd) names are not persistent across reboots or SAN changes.

By default, LVM2 does not scan for multipath devices. In addition, 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 and the device mapper devices need to be included as part of the volume scan operation. The procedures for RHEL 4, RHEL 5, SLES 9, and SLES 10 are outlined in the following sections.

---

**Note:** The following sections will provide sample filters that may be used to configure LVM. Your environment may differ and therefore require a different filter.

---

### Configuring LVM2 for DM-MPIO on RHEL

To configure LVM2 for DM-MPIO on RHEL:

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 out 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 will filter out the boot device if it is under LVM control. If your configuration is booted under LVM, use [Step 3](#).

3. If there are Logical Volumes on devices that are not controlled by multipath, such as the boot device, then enable selective scanning of those devices. For instance in the below example, the partition `sda2` contains `/boot` and the root filesystem, and 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 `/etc/lvm/lvm.conf`.
5. Execute 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:

1. Add the following line to `/etc/lvm/lvm.conf` file, to enable scanning of device-mapper block devices.

```
types = [ "device-mapper", 1]
```

2. Replace the default filter with the following line. This filters out all `sd` devices from the system and scans for multipath devices in the `/dev/disk/by-name` persistent directory generated by `udev`.

- 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 there are Logical Volumes on devices that are not controlled by multipath, 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.

- 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 `/etc/lvm/lvm.conf`. Execute the command `lvmdiskscan` and ensure that the required SCSI devices are scanned and that the LVM volume groups and partitions are available.

## Disabling Linux Multipath

If you decide to upgrade to PowerPath for path management of your servers' storage, it is necessary to disable Linux Multipath *before* installing PowerPath or both multipathing applications will hold locks on the same devices, making the server unstable for use.

*Prior* to installing PowerPath for Linux, perform the following steps:

1. As root, edit the `/etc/multipath.conf` file by commenting out all its present entries by inserting a `#` symbol at the beginning of each line entry in the configuration file that needs to be commented out, 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](#).





# CHAPTER 8

## Virtualization

This chapter contains the following information on storage virtualization:

☒ Linux virtualization .....	206
☒ Xen Hypervisor.....	207
☒ Kernel-based Virtual Machine (KVM) .....	216
☒ Citrix XenServer .....	225
☒ Oracle VM Server.....	233

## Linux virtualization

Linux-based virtualization technologies have gained extensive attention since their introduction. The two representative offerings are Xen virtualization technology and KVM (Kernel Virtual Machine) 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. There are numerous benefits to virtualization. Those benefits include (but are *not* limited to) the following:

- ☒ Creates an open storage environment capable of easily supporting the introduction of new technologies
- ☒ Significantly 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. Originally developed by Cambridge University, Xen technology has been adopted by major Linux operating system vendors, and emerged as a main stream Linux virtualization product.

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. Refer to the [Dell EMC Simple Support Matrix](#) for the latest supported configurations.

This section contains the following information:

- ☒ [“Virtualization modes” on page 207](#)
- ☒ [“Virtual machine installation and management” on page 208](#)
- ☒ [“Storage management” on page 214](#)
- ☒ [“Connectivity and path management software” on page 214](#)

## Virtualization modes

Xen kernel implements virtualization of operating system in two modes, each discussed briefly in this section:

- ☒ [“Paravirtualization” on page 207](#)
- ☒ [“Full virtualization” on page 207](#)

Also included is a brief description of the following:

- ☒ [“Virtual machine-specific drivers” on page 208](#)
- ☒ [“Cluster or pool environment” on page 208](#)

### Paravirtualization

*Paravirtualization* transfers the virtualized operating systems' 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.

<b>Virtual machine-specific drivers</b>	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.
<b>Cluster or pool environment</b>	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

The Virtual Machine (VM) can be installed and monitored via both graphical mode and command lines.

Both SuSE and Red Hat offer graphical user interface, GUI, for installation and administration of Virtual Machines.

---

**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 OS is installed.

---

Xen kernel hypervisor and virtualization management GUI need to be installed separately using different packages. Ensure 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 up 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 [Figure 50 on page 209](#) through [Figure 55 on page 213](#). 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 figure shows three overlapping screenshots of the 'Create a new virtual machine' wizard interface, illustrating steps 1, 2, and 3 of the process.

**Step 1: Virtual Machine Name**  
 The first screenshot shows the 'Virtual Machine Name' step. It prompts the user to 'Please choose a name for your virtual machine:'. The 'Name' field contains 'new\_VM'. An example name 'system1' is shown below the field.

**Step 2: Virtualization Method**  
 The second screenshot shows the 'Virtualization Method' step. It prompts the user to 'You will need to choose a virtualization method for your new virtual machine:'. Two options are available:
 

- Paravirtualized: Lightweight method of virtualizing machines. Limits operating system choices because the OS must be specially modified to support paravirtualization, but performs better than fully virtualized.
- Fully virtualized: Involves hardware simulation, allowing for a greater range of virtual devices and operating systems (does not require OS modification).

 Below the options, the 'CPU architecture' is set to 'x86\_64' and the 'Hypervisor' is set to 'kvm'.

**Step 3: Installation Method**  
 The third screenshot shows the 'Installation Method' step. It prompts the user to 'Please indicate where installation media is available for the operating system you would like to install on this virtual machine:'. Three options are available:
 

- Local install media (ISO image or CDROM)
- Network install tree (HTTP, FTP, or NFS)
- Network boot (PXE)

 Below the options, it prompts the user to 'Please choose the operating system you will be installing on the virtual machine:'. The 'OS Type' and 'OS Variant' are both set to 'Generic'. A note states: 'Not all operating system choices are supported by Red Hat. Please see the link below for supported configurations:'. A link is provided: [Red Hat Enterprise Linux 5 virtualization support](#).

Figure 50 Virtual machine installation wizard interface steps 1 – 3

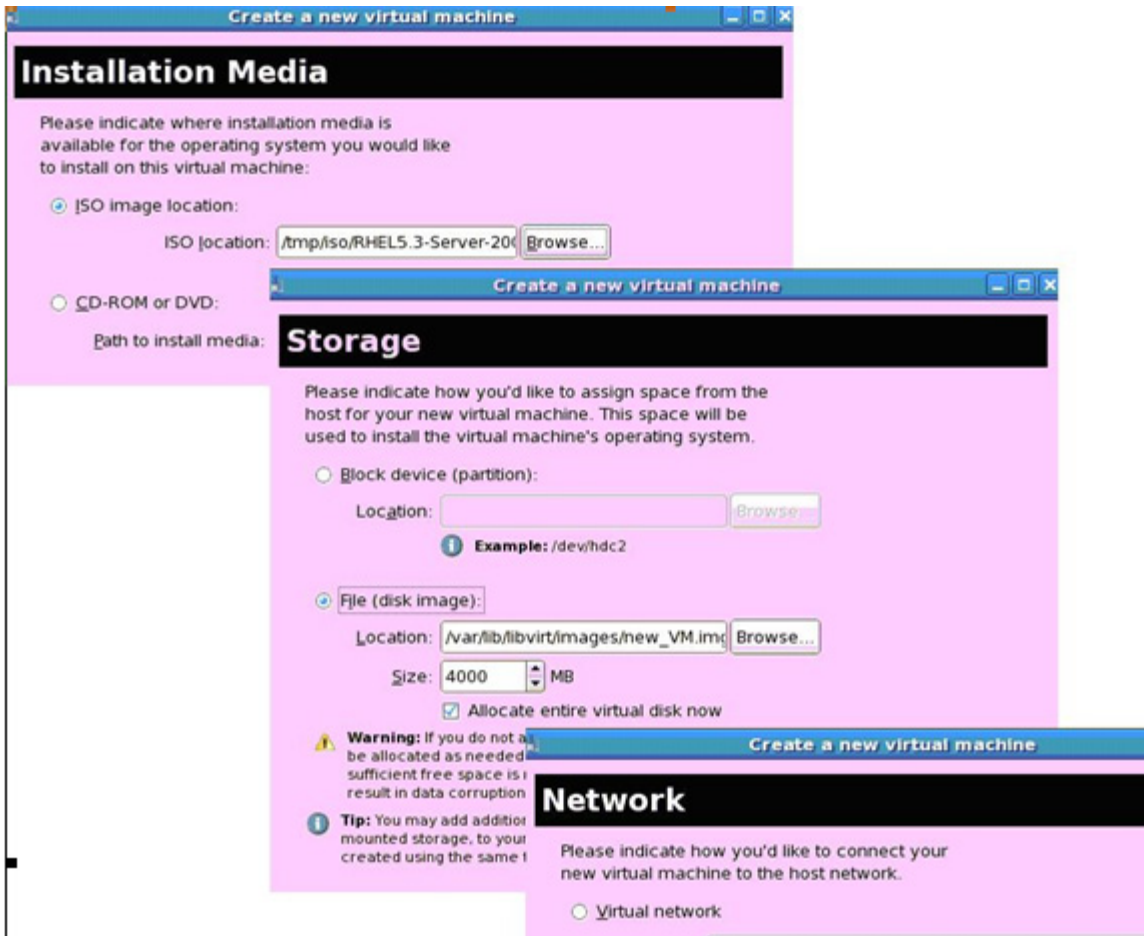


Figure 51 Virtual machine installation wizard interface steps 4 – 6

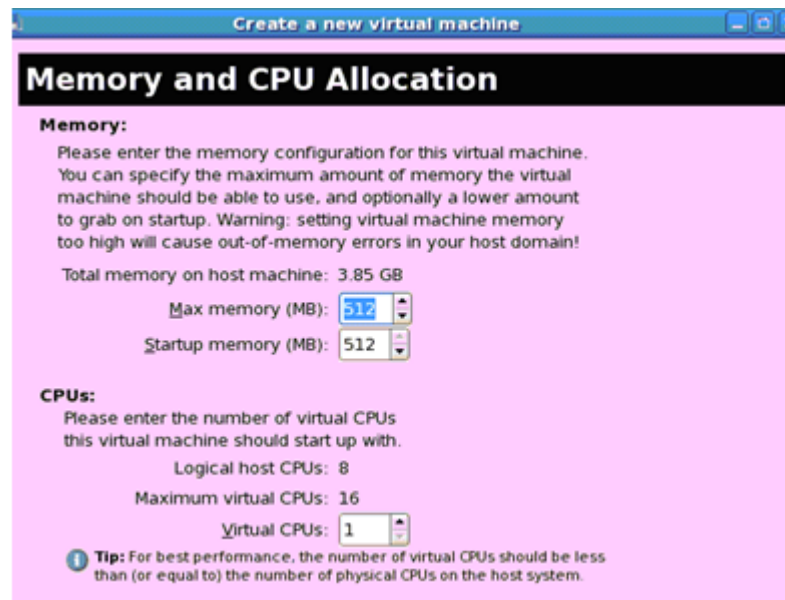


Figure 52 Virtual machine installation wizard interface step 7

After all parameters are properly indicated, a summary screen will display showing the selected parameters.

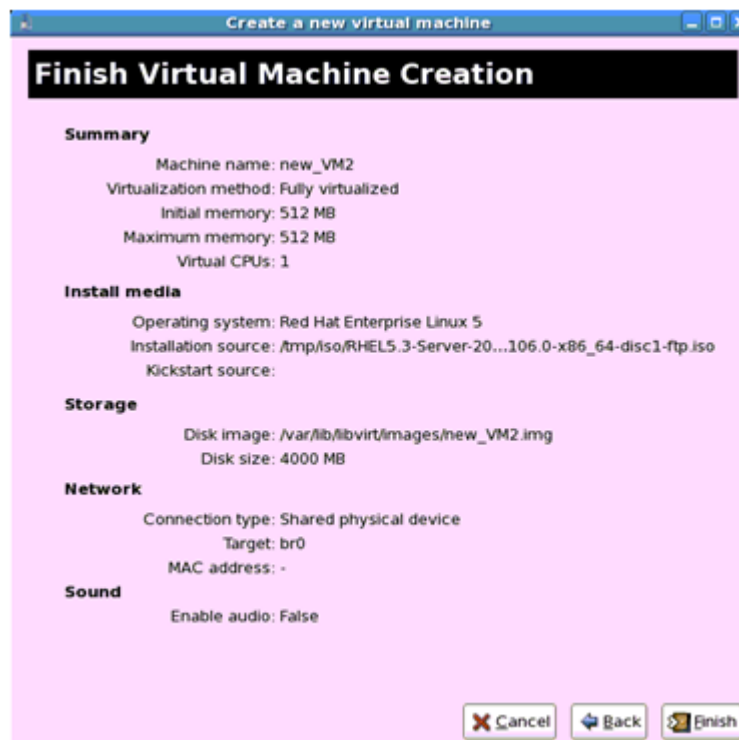
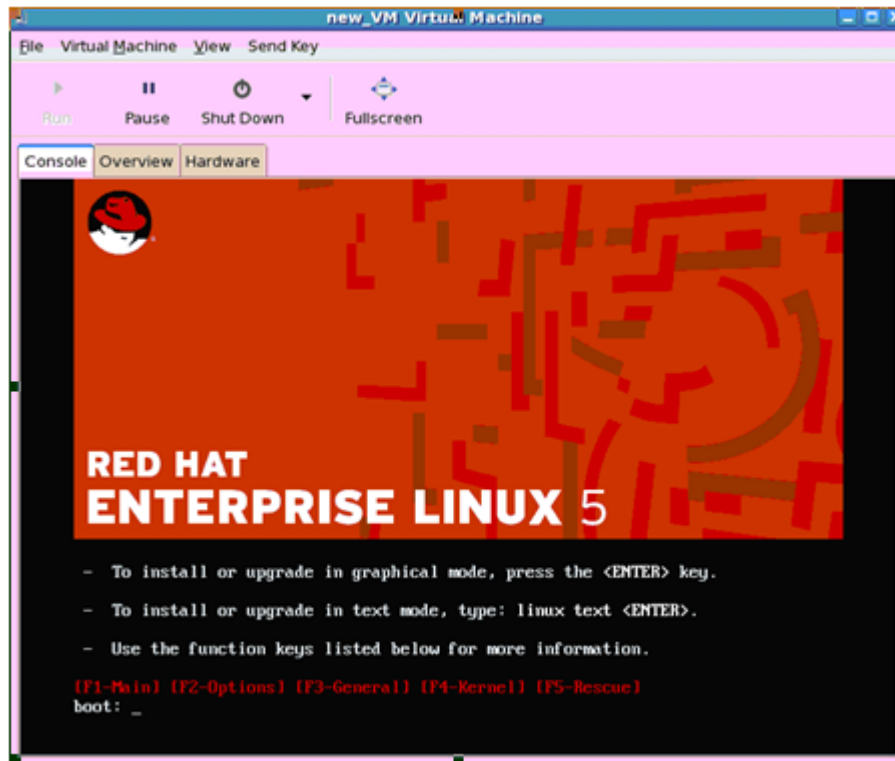


Figure 53 Summary screen showing virtual machine installation parameters

When the parameters are confirmed, the installation process can be started.

A **VNC Viewer** window connecting to the virtual machine's console will display and the installation can be continued in the same way as installing a physical machine.



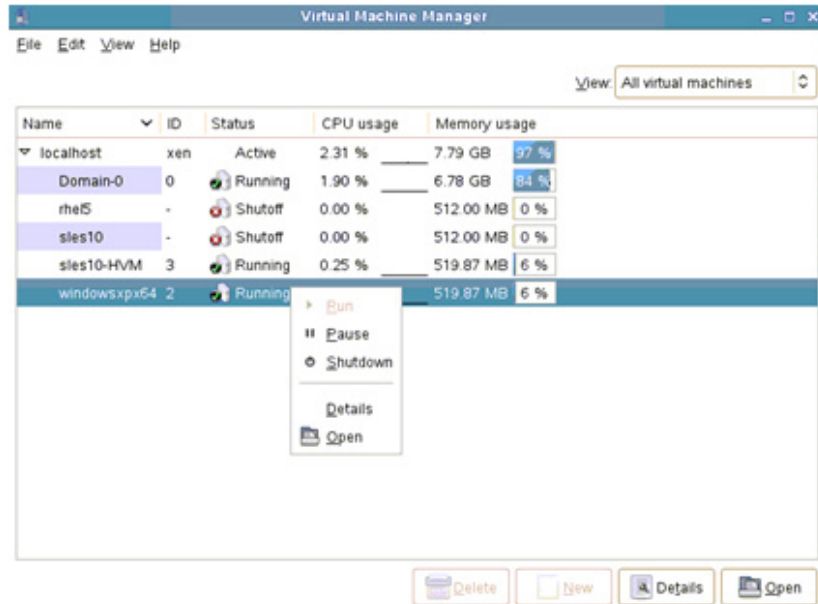
**Figure 54** VNC Viewer window connecting to the virtual machine's console

For more information on the virtualization package installation and configuration, and virtual machine installation, refer to:

- ☒ [SuSE on the Micro Focus \(Novell\) website](#)
  
- ☒ [Red Hat website](#)

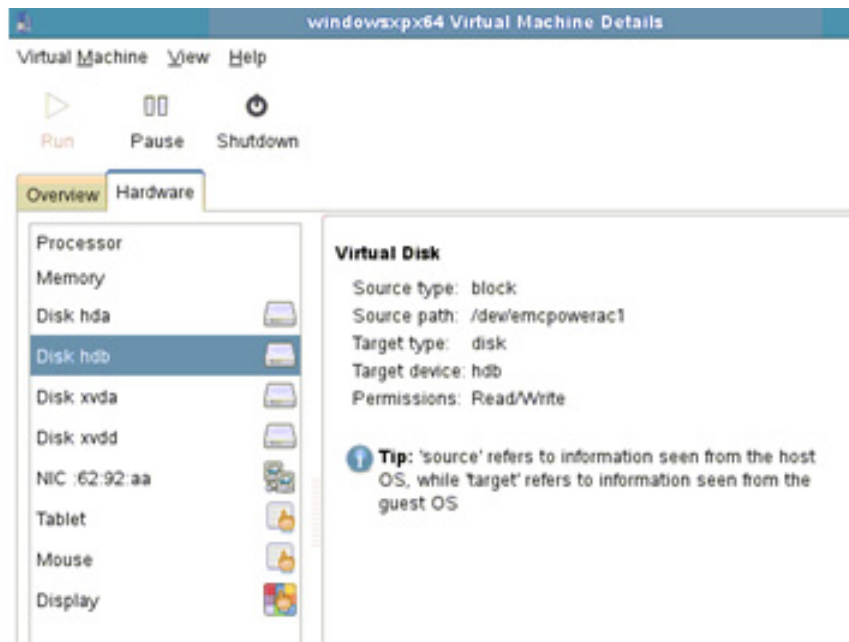


After installation, the VMs will appear as new domains under the control of the Xen kernel. Users can access the graphical console of the VMs directly using the VNC Viewer function provided by the virtual machine manager GUI. The virtual machines can be shut down, started, paused, or destroyed using virtualization management utility, or command lines. Their status and hardware utilization status are also monitored by the management utility. An example is shown in Figure 55.



**Figure 55** 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 Figure 56.



**Figure 56** 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 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 associated with Dom0 is the coordination and distribution of virtual I/O from the physical server to the guest operating system via 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 will have no access to the underlying hardware except thru 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 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. Refer to the Dell EMC Simple Support Matrix Linux OS 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 (KVM)

The Kernel-based Virtual Machine (KVM), is a Linux kernel virtualization infrastructure implemented as a kernel module, as opposed to a separate integrated kernel, such as Xen.

This section contains the following information:

- ☒ [“Introduction” on page 216](#)
- ☒ [“Implementing KVM” on page 216](#)
- ☒ [“Installing and managing the virtual machine” on page 220](#)
- ☒ [“Storage management” on page 223](#)
- ☒ [“Connectivity and multipathing functionality” on page 223](#)

### 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.

### Implementing KVM

This section contains the following information to implement KVM:

- ☒ [“Prerequisites” on page 216](#)
- ☒ [“SLES 11 and 12” on page 217](#)
- ☒ [“RHEL 5” on page 217](#)
- ☒ [“RHEL 6 and 7” on page 218](#)

**Prerequisites** There are some packages that need to be installed in order to run KVM.

#### **IMPORTANT**

These packages will not be automatically installed during OS installation unless manually specified during package selection.

These packages include:

- ☒ libvirt-0.7.6-1.9.8
- ☒ virt-utils-1.1.2-0.2.17
- ☒ 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

**SLES 11 and 12** In the case of SLES systems, users are allowed to install additional applications after OS installation is completed. SLES YaST service serves this purpose. Therefore, users can implement KVM technology when the need arises after the OS is installed, in the same manner as adding other applications to a running SLES system.

**RHEL 5** In the case of RHEL 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 OS installation. [Figures 57 through 59](#) show the process of installing the KVM package in RHEL 5.x series systems.



**Figure 57** Enter the Installation Number



Figure 58 Select the Virtualization package

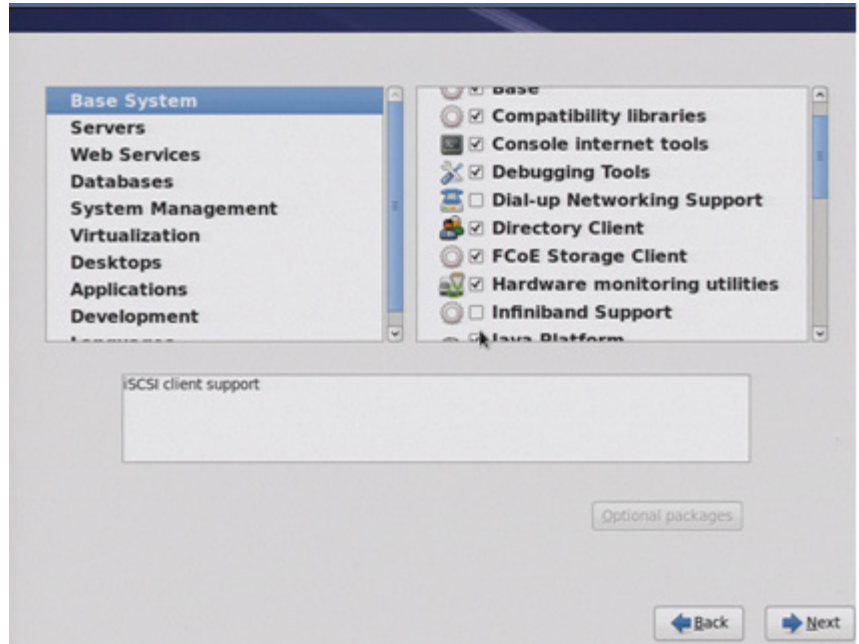


Figure 59 Select KVM, Virtualization

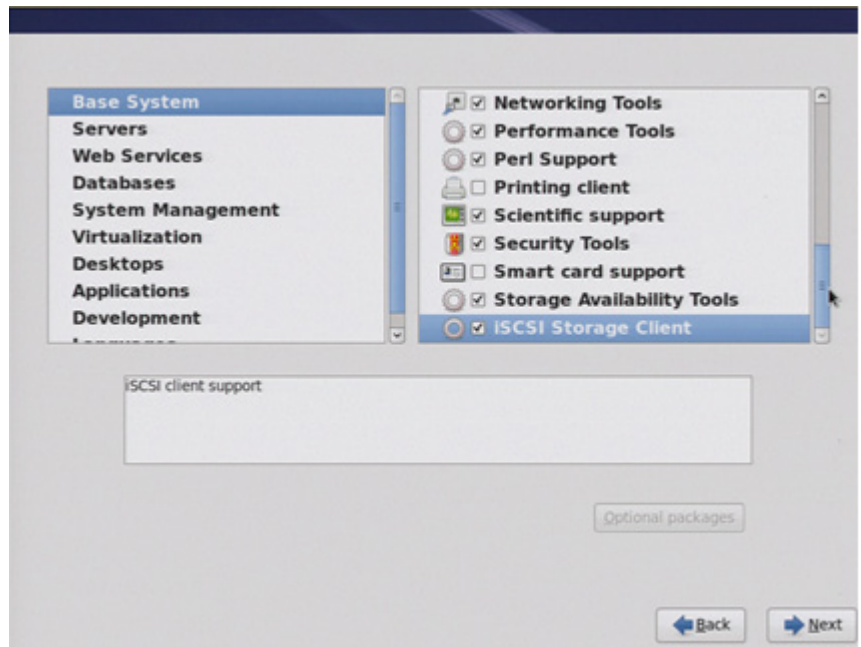
**RHEL 6 and 7**

In the RHEL 6 and 7 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

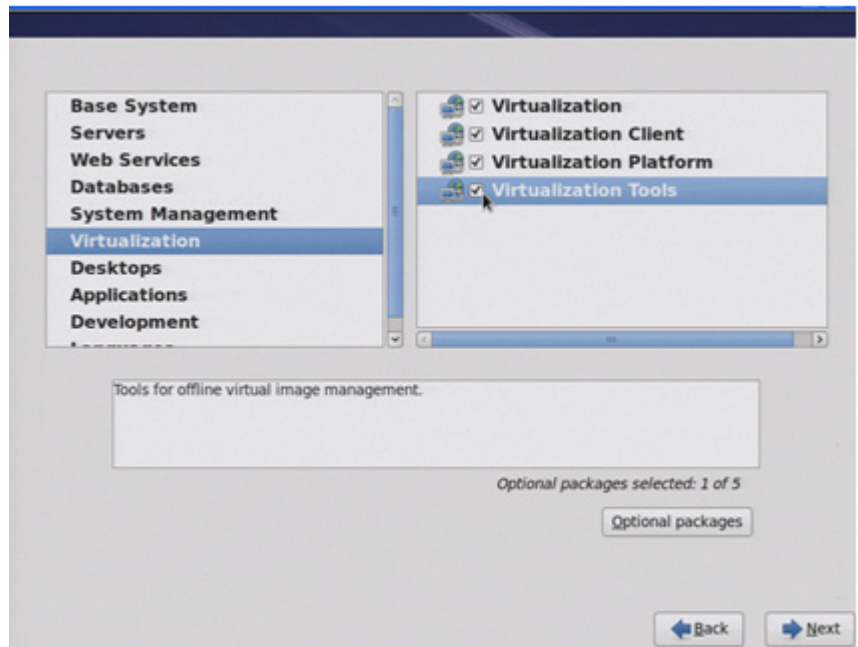
Figures 60 through 62 show the process of installing the KVM package in RHEL 6 series systems.



**Figure 60** Customize the Base System



**Figure 61** Customize the Base System (cont.)



**Figure 62** 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.

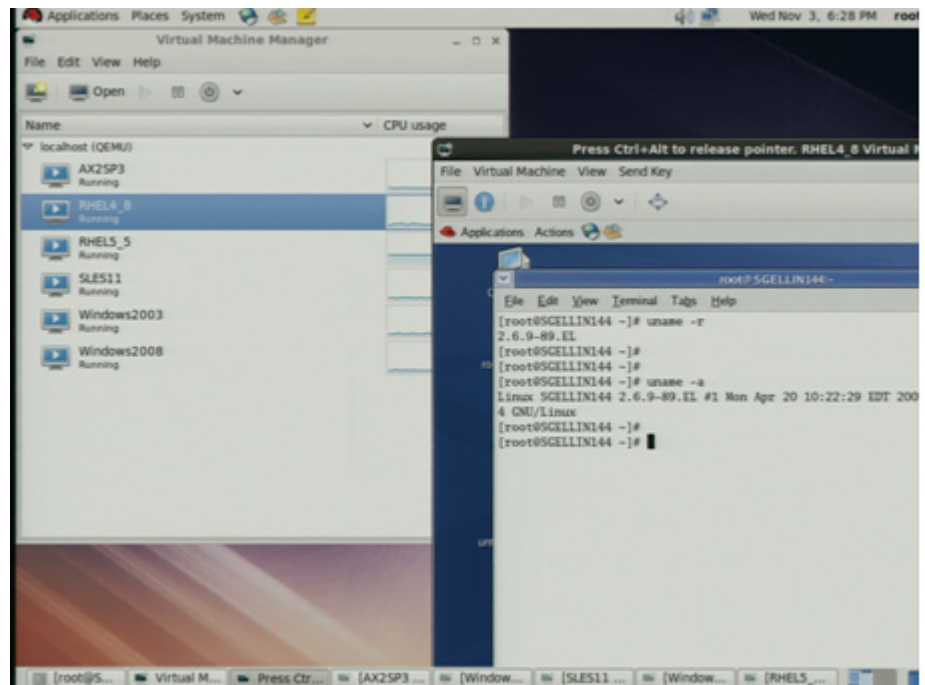
In both RHEL and SLES environments, the management GUI can be evoked by typing **virt-manager** on the command prompt. In SLES, the GUI can also be accessed from the suite of utilities consolidated under YaST.

Before creating virtual machines on SLES hosts running KVM, make sure 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.

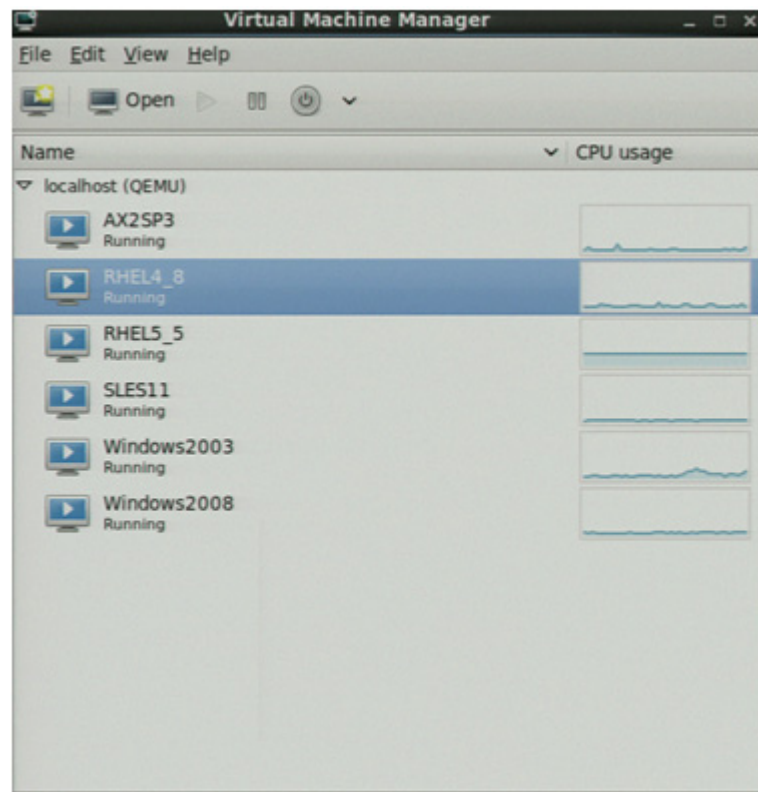
For detailed virtual machine installation procedures, refer to the [“Virtual machine installation and management”](#) on page 208.



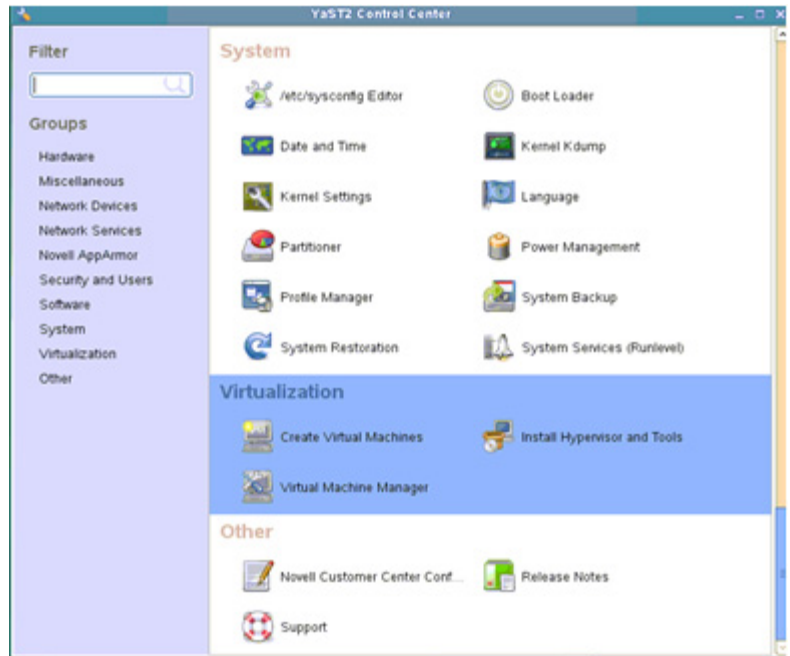
Figures 63 through 68 show how to install the virtual machine.



**Figure 63** Virtual Machine Manager Utility on RHEL



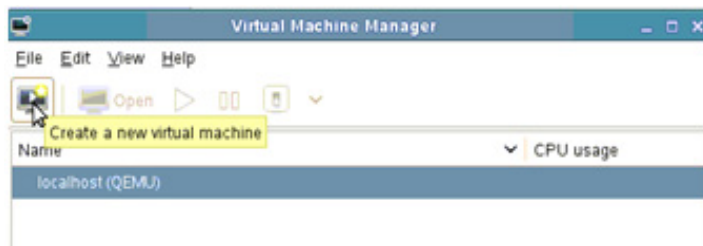
**Figure 64** Virtualization Machine Manager virtual machine statuses



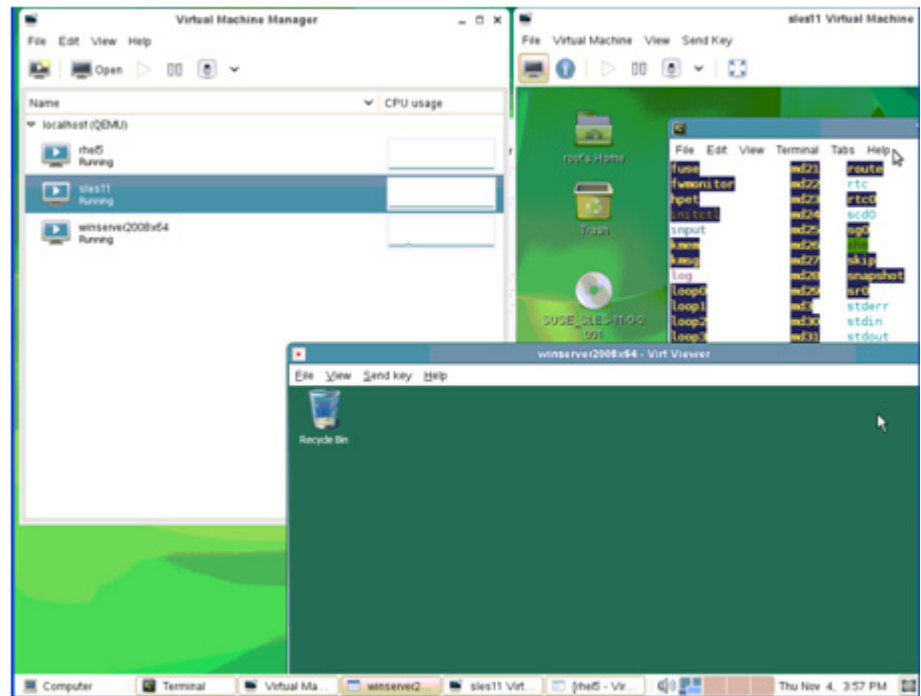
**Figure 65** Procedures to evoke virtualization management GUI utility in SLES



**Figure 66** Virtualization Machine Manager



**Figure 67** Create a new virtual machine



**Figure 68** Virtual Machine Manager Utility on SuSE with VMs running

## Storage management

As the administration of KVM is very 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 OS vendor on KVM as Guest OS. By default, the guest OS is a fully-virtualized machine. The development community and OS 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 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. Refer to the [Dell EMC Simple Support Matrix](#) Linux OS footnotes in the base connectivity section for supported configurations.

## Creating a VM on PowerPath

### **IMPORTANT**

For KVM on RHEL 6.2 and previous releases, Dell EMC does not Support creating a VM (Virtual Machine) directly on PowerPath devices. You must pass PowerPath devices through LVM to KVM as logical volumes.

To create a VM on PowerPath devices on RHEL 6.3 and later, turn off cgroups as follows:

1. Edit `/etc/libvirt/qemu.conf` and put the following line in the `cgroup_controllers` section, as shown next:

```
cgroup_controllers = [ "cpu", "memory" ]
```

After editing, it should look like the following example:

```
# where they are located.
#
# cgroup_controllers = [ "cpu", "devices", "memory", "blkio",
"cpuset", "cpuacct" ]
cgroup_controllers = [ "cpu", "memory" ]
# This is the basic set of devices allowed / required by
# all virtual machines.
```

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 on KVM, refer to the following URL:

[http://www.linux-kvm.org/page/Main\\_Page](http://www.linux-kvm.org/page/Main_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 virtual machine (VM) management, storage management, and backup management.

This section contains the following information:

- ☒ “XenServer overview” on page 225
- ☒ “Connectivity and path management software” on page 225
- ☒ “Live VDI Migration” on page 226

## XenServer overview

The basic setup of a standard XenServer implementation consists of a server with XenServer OS installed, and a workstation with XenCenter management suite installed. Only a personal computer is needed for the installation of 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 together to form a server pool and facilitate VM live migration across hosts and other advanced operations.

XenServer utilizes 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 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 refer to 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, refer to the [Dell EMC Simple Support Matrix](#).

For more information regarding the virtual machine installation, refer to the [Citrix website](#).

## Connectivity and path management software

Dell EMC supports storage 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. Refer to the Dell EMC Simple Support Matrix Linux OS specific footnotes in the base connectivity section of [Dell EMC E-Lab Navigator](#) for supported configurations.

## Live VDI Migration

This section contains the following information:

- ☒ [“Live VDI Migration overview” on page 226](#)
- ☒ [“Moving virtual disks using XenCenter” on page 226](#)
- ☒ [“Limitations and caveats” on page 227](#)

### Live VDI Migration overview

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.

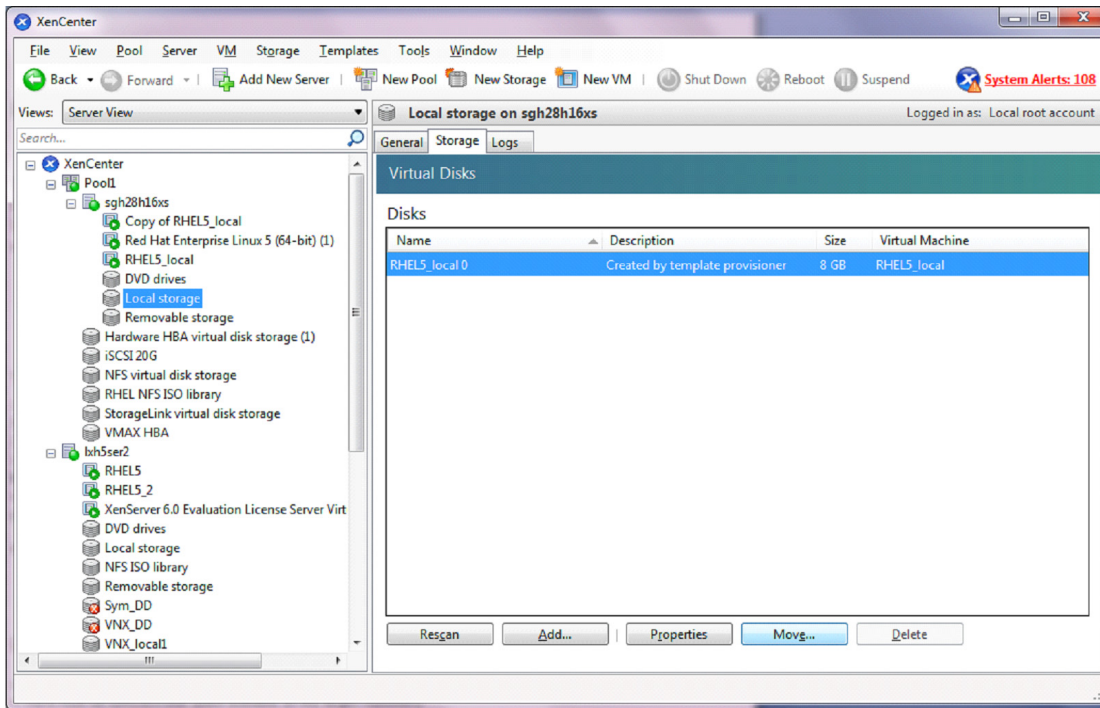
This enables administrative operations such as:

- ☒ 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

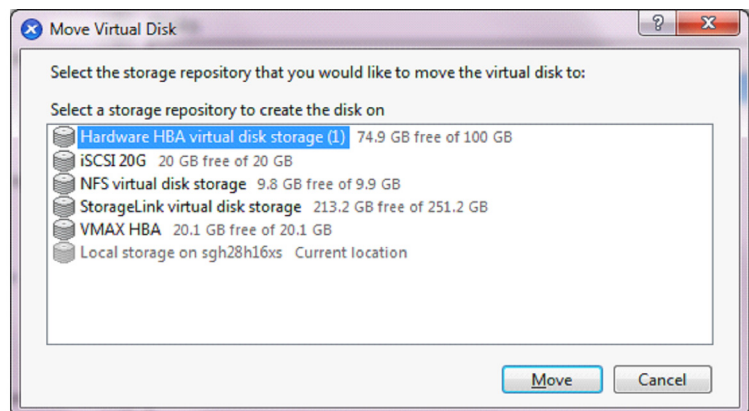
To move virtual disks, complete the following steps:

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.

**Note:** Make sure 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

This section contains the following information:

- ☒ [“XenMotion overview” on page 228](#)
- ☒ [“Compatibility requirements” on page 228](#)
- ☒ [“Migrating the virtual machine from one host to another host” on page 229](#)
- ☒ [“Limitations and caveats” on page 232](#)

### 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 shared between the two hosts. As a result, a virtual machine 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 will fail 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.

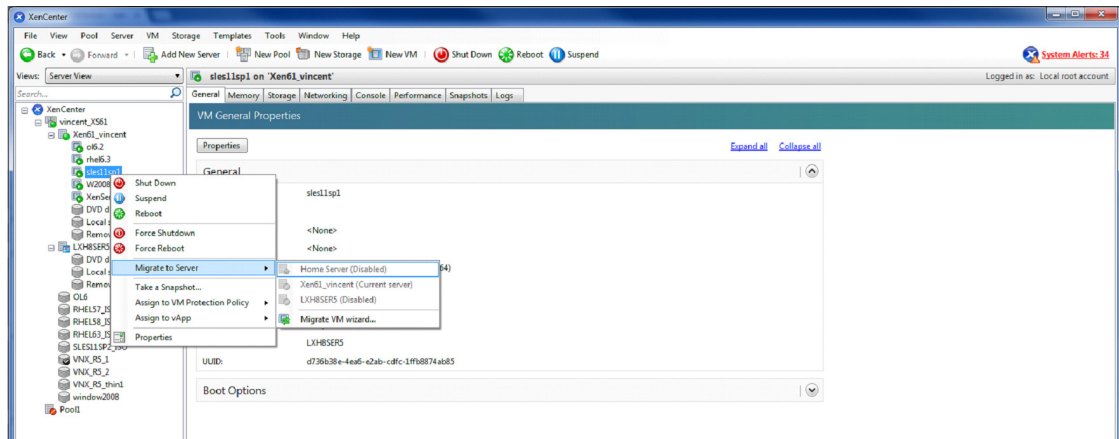


- The target storage must have enough free disk space (for the virtual machine and its snapshot) available for the incoming virtual machines. If there is not enough space, the migration will fail to complete.

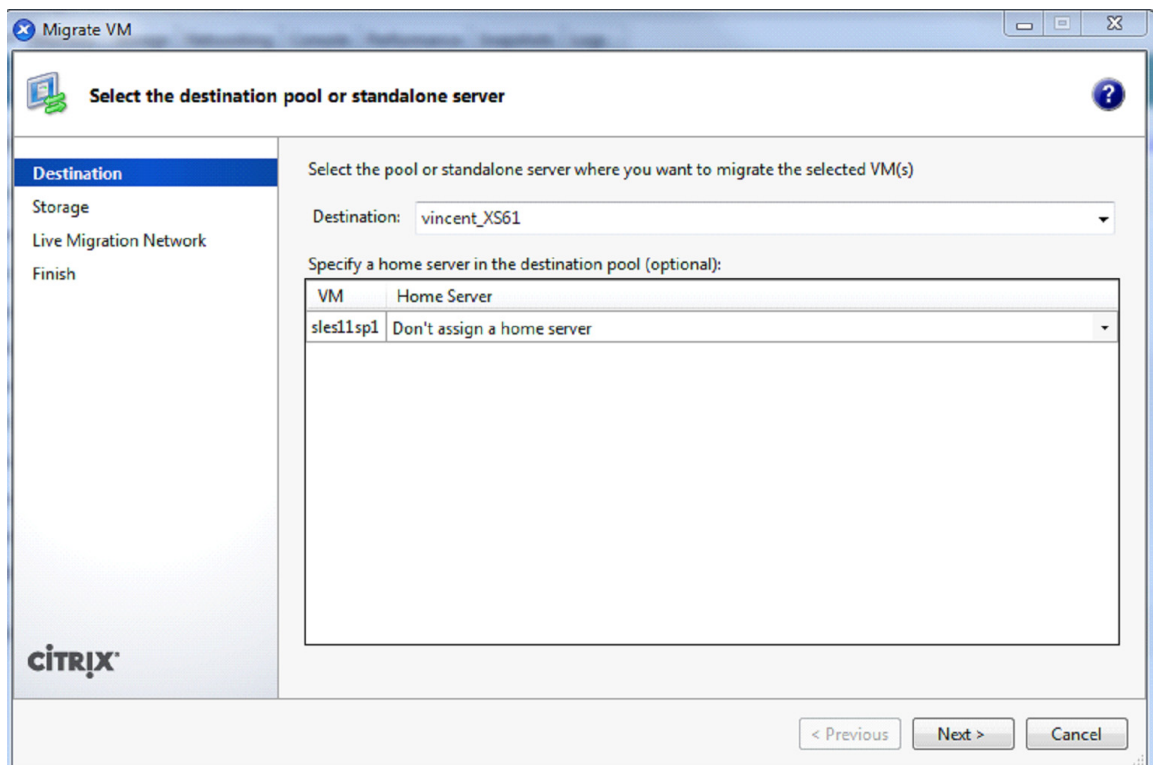
## Migrating the virtual machine from one host to another host

To migrate the virtual machine from one host to another host, complete the following steps.

1. Right-click the virtual server from the XenCenter pane. In the drop-down menu, select **Migrate to Server > Migrate VM wizard**.



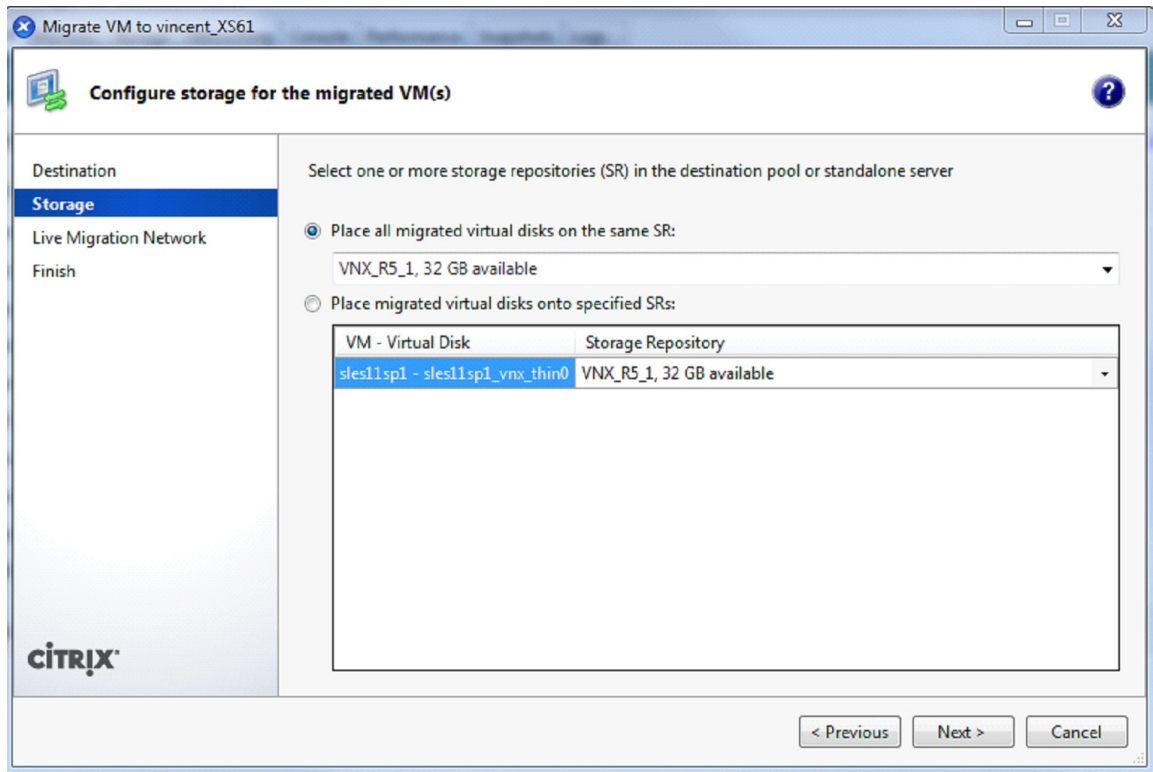
The **Destination** window displays.



2. In the **Destination** window, complete the following fields.

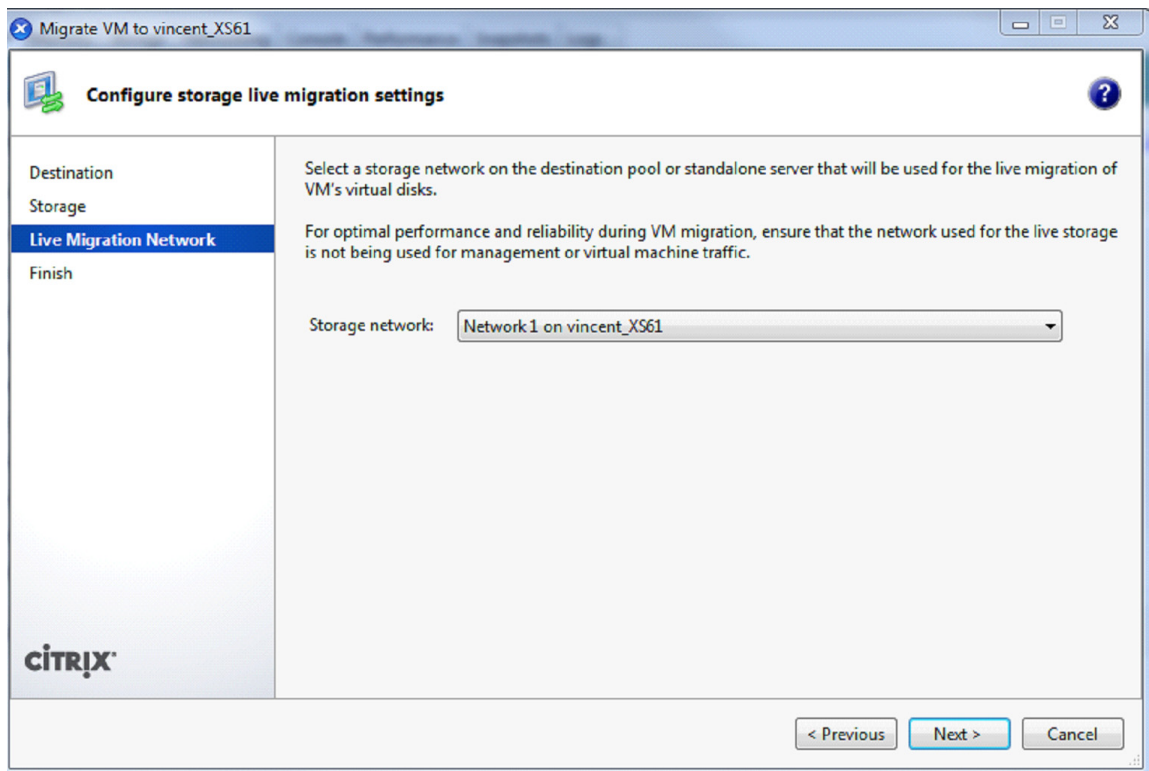
- a. In the **Destination** field, choose the destination in the drop-down menu where you want to migrate the virtual machine.

- b. Select a standalone server or a pool from the drop-down menu in the **Specify a home server in the destination pool** box.
- c. Click **Next**. The **Storage** window displays.



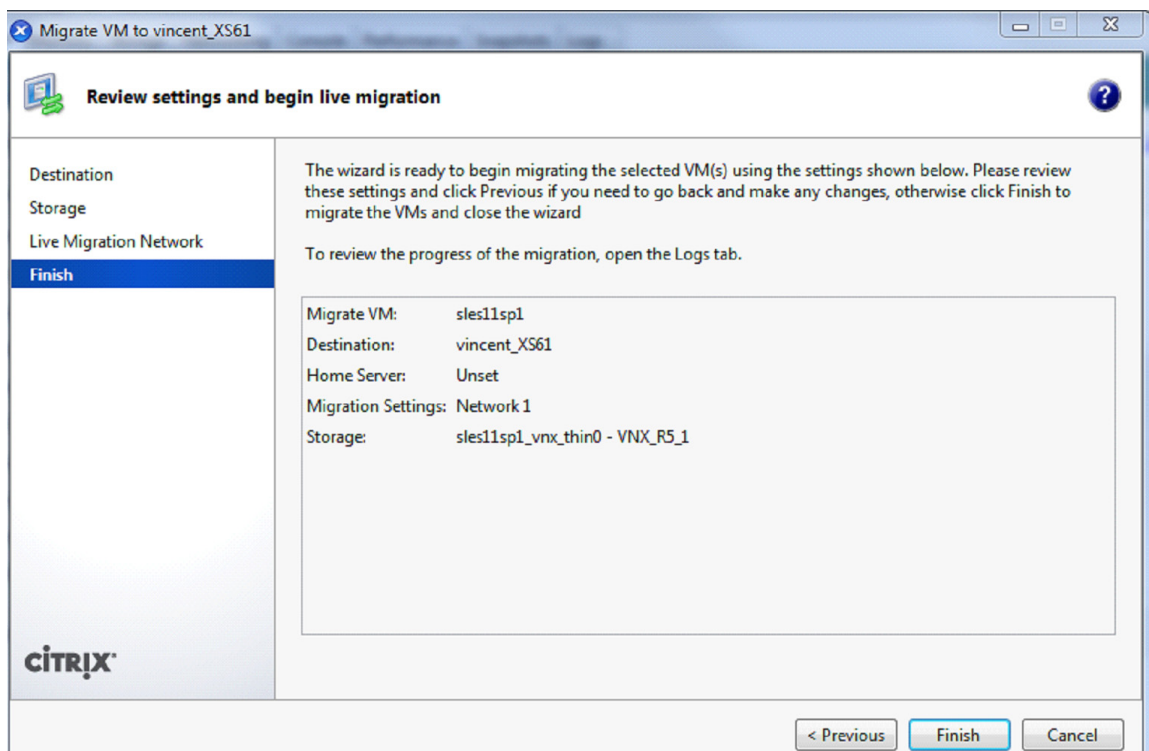
3. In the **Storage** window, complete the following fields.
  - a. Specify the storage repository where you would like to place the migrated virtual machine's virtual disks.
  - b. Click **Next**.

The **Live Migration Network** window displays.



- 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.



- Review the configuration settings and click **Finish** to start migrating the virtual machine.

**Notes** Note the following:

- ☒ 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.

## Limitations 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 depend 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 VM Server

Oracle VM Server is another Linux Xen kernel-based virtualization product. Oracle VM Server for x86 incorporates the free and open-source Xen hypervisor technology, supports Windows, Linux, and Oracle Solaris x86 guests, and includes an integrated Web-based management console. The components included in an Oracle VM configuration include hosts with Oracle VM Server installed and another host with Oracle VM Manager management suite installed.

This section contains the following information:

- ☒ [“OVM overview” on page 233](#)
- ☒ [“Connectivity and multipathing functionality” on page 233](#)

### OVM overview

Oracle VM requires Oracle VM Manager to be installed on a server. There must be a Linux enterprise-level operating system installed before Oracle VM Manager can be installed. The list of supported Linux enterprise level OS supported by Oracle VM Manager varies between different Oracle VM Manager versions. Consult Oracle for the compatible Linux operating system and Oracle VM Manager combinations.

For Oracle VM Server 3.0 and later, it may be required to implement Oracle VM Server environment in a cluster setting. Therefore, minimally two hosts with Oracle VM Server 3.0 installed are required before VMs can be created. Consult Oracle documentation for installation requirements.

Oracle VM uses a storage repository concept. For Oracle VM 3.0 and later, a storage repository can only be created by the Oracle VM Manager.

Because cluster setting is compulsory for Oracle VM 3.0 series, a cluster policy disk, which is used to store the cluster configuration information (similar to the function of quorum disk) is needed. This policy disk must be visible to all nodes inside the cluster.

### Connectivity and multipathing functionality

Dell EMC supports storage provisioned to hosts running OVM through the following connectivity:

- ☒ Fibre Channel
- ☒ iSCSI
- ☒ Fibre Channel over Ethernet (FCoE)

Dell EMC can support both Linux native multipathing (DM-MPIO) and PowerPath with OVM technology. Refer to Chapter 7, [“Native Multipath Failover” on page 163](#), for Linux native multipath configuration, and refer to the *PowerPath Installation and Administration Guide* available on [Dell EMC Online Support](#) for additional information about PowerPath installation and administration on an OVM. Because an Oracle VM server uses a Oracle Linux kernel, the PowerPath version to be installed should use the one matching the equivalent version of Oracle Linux. Install PowerPath on the Oracle VM Server (hypervisor), and not the virtual machine. The virtual machine is then spared the burden of implementing load balancing and failover utilities.



# CHAPTER 9

## Virtual Provisioning

This chapter provides information about Virtual Provisioning and Linux.

---

**Note:** For further information regarding the correct implementation of Virtual Provisioning, refer to the *Symmetrix Virtual Provisioning Implementation and Best Practices Technical Note*, available at [Dell EMC Online Support](#).

---

☒ Virtual Provisioning on VMAX series.....	236
☒ Virtual Provisioning on VNX, Unity, or CLARiiON .....	238
☒ Virtual Provisioning on XtremIO .....	239
☒ Space reclamation.....	240
☒ Implementation considerations.....	241
☒ Operating system characteristics .....	245

## Virtual Provisioning on VMAX series

VMAX series Virtual Provisioning improves storage capacity utilization and simplifies storage management by allowing storage to be allocated and accessed on demand from a pool of storage that services one or many applications.

This type of storage has multiple benefits:

- ☒ Enables LUNs to be increased over time to meet increased demand with no impact to the host or application as space is added to the thin pool.
- ☒ Delivers space from the thin pool on demand.
- ☒ Provides wide striping for a thin pool.
- ☒ Relieves the storage administrator's efforts of physical device/LUN configuration.

The Virtual Provisioning feature introduces three new concepts: thin devices, data devices, and thin pools. Thin devices can be created with an inflated capacity, because the actual storage space for the data written to the thin devices is on the data devices. In this way, when additional storage is needed, more data devices can be created in the thin pool.

Virtual Provisioning simplifies data layout, with automated wide striping that provides equivalent or potentially better performance than standard provisioning. Virtual Provisioning is appropriate for all storage types in a tiered storage environment and supports both local and remote replication with SRDF and TimeFinder.

Virtual provisioning provides the ability to persistently preallocate space. Extents that are persistently preallocated are not reclaimed by a standard reclaim operation.

## Terminology

This section provides common terminology and definitions for VMAX series virtual provisioning

<b>VP compression</b>	5876 Q4 2012 SR introduces VP compression, allowing thin device data to be compressed within a thin pool. Data may be compressed manually for an individual device or group of devices, via Solutions Enabler or Unisphere for VMAX series. Alternatively, inactive data may be compressed automatically for thin devices that are managed by FAST VP. In order for data to be compressed, the thin pool containing the data must be enabled for compression. Only allocated extents are compressed. VP compression is supported on FBA and CKD 3390 devices.
<b>Thin devices (TDEVs)</b>	Thin devices, also known as VP devices (FBA and CKD), have no storage allocated to them when they are created; storage is instead allocated on-demand from a "bound" thin pool. The first write to a location in a thin device results in space being allocated on a data device from the bound pool.
<b>Data devices (TDATs)</b>	Data devices are grouped in a thin pool and are dedicated to the purpose of providing the actual physical storage used by thin devices. As with thin pools, data devices must have identical device emulation type, reside on identical drive technologies, and use the same RAID protection types, and drive technology.
<b>Thin pool</b>	A thin pool, also known as a VP pool, contains thin devices of identical emulation and protection type, all of which reside on disks of the same technology type and speed.



When a write is performed to a portion of the thin device, the VMAX system allocates a minimum allotment of physical storage from the pool and maps that storage to a region of the thin device including the area targeted by the write. These storage allocation operations are performed in small units of storage called thin device extents.

VMAX systems balance the allocation of extents across all the data devices in the pool that are enabled and that have remaining unused capacity. When a read is performed on a thin device, the data being read is retrieved from the appropriate data device in the thin pool to which the thin device is bound. Reads directed to an area of a thin device that has not been mapped do not trigger allocation operations. The result of reading an unmapped block is that a block in which each byte is equal to zero will be returned. When more storage is required to service existing or future thin devices, data devices can be added to existing thin storage pools. New thin devices can also be created and associated with existing thin pools.

A thin device can be presented for host use before all of the reported capacity of the device has been mapped. If the reported capacities of the thin devices using a given pool exceeds the pool available storage capacity, the thin device configuration is said to be oversubscribed.

### **Oversubscribed thin pools**

Oversubscribing of thin pools allows the presenting of larger than needed devices to hosts and applications without having enough physical drives to fully allocate all of the space represented by the thin devices.

---

**Note:** Unlike earlier VMAX series, VMAX3 arrays are pre-configured at the factory with Virtual Provisioning (VP) pools ready for use. DATA devices (TDATs) are provisioned/pre-configured /created while the host addressable storage devices TDEVs are created by either the customer or customer support, depending on the environment. VMAX3 arrays support only thin devices

---

For more details about thin provisioning on a VMAX series, refer to the individual product guide that is available on Dell EMC Online Support.

## Virtual Provisioning on VNX, Unity, or CLARiiON

Virtual Provisioning enables organizations to reduce storage costs by increasing capacity utilization, simplifying storage management, and reducing application downtime. Virtual Provisioning also helps companies to reduce power and cooling requirements and delay capital expenditures.

One of the biggest challenges facing storage administrators is balancing how much storage space will be required by the various applications in their data centers. Administrators are typically forced to allocate space based on anticipated storage growth. They do this to reduce the management expense and application downtime required to add storage later on. This generally results in the over-provisioning of storage capacity, which then leads to higher costs; increased power, cooling, and floor space requirements; and lower capacity utilization rates. Even with careful planning, it may be necessary to provision additional storage in the future. This may require application downtime depending on the operating systems involved.

To address these concerns, the CLARiiON CX4 introduced thin LUN technology with FLARE release 28.5. This technology works with Unity, VNX, or CLARiiON to provide powerful, cost-effective, flexible solutions. Unity, VNX, or CLARiiON thin LUNs can present more storage to an application than is physically available. Storage managers are freed from the time-consuming administrative work of deciding how to allocate drive capacity. Instead, an array-based mapping service builds and maintains all of the storage structures based on a few high-level user inputs. Disk drives are grouped into storage pools that form the basis for provisioning actions. Physical storage is automatically allocated only when new data blocks are written.

Thin provisioning improves storage capacity utilization and simplifies storage management by presenting an application with sufficient capacity for an extended period of time. When additional physical storage space is required, disk drives can be nondisruptively added to the central storage pool. This reduces the time and effort required to provision additional storage, and avoids provisioning storage that might not be needed.

For more information on Thin Provisioning for Unity, VNX and CLARiiON, refer to the following white papers on EMC.com:

- ☒ For VNX5200, VNX5400, VNX5600, VNX5800, VNX7600, and VNX8000 VNX Virtual Provisioning for the new VNX Series - Applied Technology:

*Virtual Provisioning for the EMC VNX2 Series White Paper*

- ☒ For VNX5100, VNX5300, VNX5500, VNX5700 and VNX 7500 VNX Virtual Provisioning VNX5100, VNX5300, VNX5500, VNX5700 and VNX 7500 - Applied Technology:

*EMC VNX Virtual Provisioning White Paper*

- ☒ For EMC CLARiiON Virtual Provisioning - Applied Technology:

*EMC CLARiiON Virtual Provisioning White Paper*

- ☒ For EMC UNITY: FAST Technology Overview:

*EMC Unity: Fast Technology Overview White Paper*

## Virtual Provisioning on XtremIO

XtremIO storage is natively thin provisioned, using a small internal block size. This provides fine-grained resolution for the thin provisioned space.

All volumes in the system are thin provisioned, meaning that the system consumes capacity only when it is actually needed. XtremIO determines where to place the unique data blocks physically inside the cluster after it calculates their fingerprint IDs. Therefore, it never pre-allocates or thick-provisions storage space before writing.

As a result of XtremIO's content-aware architecture, blocks can be stored at any location in the system (and only metadata is used to refer to their locations) and the data is written only when unique blocks are received.

Therefore, unlike thin provisioning with many disk-oriented architectures, with XtremIO there is no space creeping and no garbage collection. Furthermore, the issue of volume fragmentation over time is not applicable to XtremIO (as the blocks are scattered all over the random-access array) and no defragmentation utilities are needed.

XtremIO's inherent thin provisioning also enables consistent performance and data management across the entire life cycle of the volumes, regardless of the system capacity utilization or the write patterns to the system.

You can monitor XtremIO cluster thin-provisioning savings by using the storage pane that is shown in the dashboard workspace in [Figure 69](#). The dashboard compares used disk space to allocated disk space.



Storage	
Overall Efficiency	9.4:1
Data Reduction Ratio	3.7:1
Deduplication	2.8:1
Compression	1.3:1
Thin Provisioning Saving	61%

**Figure 69** Thin-provisioning savings dashboard

**Note:** Refer to the *EMC XtremIO Storage Array User Guide* on Dell EMC Online Support for more details about XtremIO thin provisioning properties.

## Space reclamation

One feature of the SCSI standard for thin provisioning is the ability for a storage device to reclaim unused space, extents, and provide them back to the pool for use by another user or application.

Dell EMC storage with Virtual Provisioning includes support for the SCSI UNMAP and WRITE\_SAME unmap commands. These commands allow for operating systems and applications to communicate a range of logical block addresses (LBAs) that are no longer needed. If the LBA range covers an entire Virtual Provisioning extent it can be reclaimed by the Dell EMC storage array and added back to the pool of free space.

The following are supported forms of reclamation with Linux hosts and Dell EMC storage, each discussed briefly in this section.

- ☒ “Veritas Storage Foundation” on page 240
- ☒ “Linux filesystem” on page 240

### Veritas Storage Foundation

Veritas Storage Foundation offers a reclamation facility that utilizes the WRITE\_SAME unmap command specification inherent in Dell EMC storage with Virtual Provisioning.

Dell EMC supports the following at a minimum:

- ☒ VxVM 5.1 for Thin Reclamation on VNX series, VNXe series, Unity series, or CLARiiON systems
- ☒ VxVM 5.1 SP1 on VMAX 400K/200K/100K and VMAX All Flash, VMAX 40K, VMAX 20K/VMAX, VMAX 10K (Systems with SNxxx987xxxx), VMAX 10K (Systems with SN xxx959xxxx, VMAXe)
- ☒ VxVM 6.0.1 for Thin Reclamation on XtremIO system.

PowerPath is not currently supported with Thin Reclamation. Consult the Symantec document, *Automating Thin Storage Reclamation with Veritas Storage Foundation*, on the [Symantec website](#).

### Linux filesystem

The Linux filesystem is an ordered tree-like hierarchical structure 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 can further contain other subdirectories or files.

Each file is described by an inode, which holds location (LBAs) and other important information of the file. Therefore, when a file or directory is deleted or modified, the range of LBAs in use by the filesystem changes, and extents on the Dell EMC storage may be released.

With the release of Red Hat's RHEL 6.3, VMAX 400K/200K/100K, VMAX All Flash, VMAX 40K, 20K/VMAX, VMAX 10K(SN xxx987xxx)/VMAX 10K(SN xxx959xxx), and VMAXe starting from code 5876.159.102 and with SPC-3 enable this feature on an ext4 filesystem and use the following mount command:

```
#> mount -t ext4 -o discard /dev/emcpowera /my_filesystem
```

## Implementation considerations

When implementing Virtual Provisioning, it is important that realistic utilization objectives are set. Generally, organizations should target no higher than 60 percent to 80 percent capacity utilization per pool. A buffer should be provided for unexpected growth or a “runaway” application that consumes more physical capacity than was originally planned for. There should be sufficient free space in the storage pool equal to the capacity of the largest unallocated thin device.

Organizations also should balance growth against storage acquisition and installation timeframes. It is recommended that the storage pool be expanded before the last 20 percent of the storage pool is utilized to allow for adequate striping across the existing data devices and the newly added data devices in the storage pool.

Thin devices can be deleted once they are unbound from the thin storage pool. When thin devices are unbound, the space consumed by those thin devices on the associated data devices is reclaimed.

---

**Note:** Users should first replicate the data elsewhere to ensure it remains available for use.

---

Data devices can also be disabled and/or removed from a storage pool. Prior to disabling a data device, all allocated tracks must be removed (by unbinding the associated thin devices). This means that all thin devices in a pool must be unbound before any data devices can be disabled.

This section contains the following information:

- ☒ [“Over-subscribed thin pools” on page 241](#)
- ☒ [“Thin-hostile environments” on page 242](#)
- ☒ [“Pre-provisioning with thin devices in a thin hostile environment” on page 242](#)
- ☒ [“Host /boot, / \(root\), /swap, and /dump devices positioned on Symmetrix Virtual Provisioning \(tdev\) devices” on page 243](#)
- ☒ [“Cluster configuration” on page 244](#)

### Over-subscribed thin pools

It is permissible for the amount of storage mapped to a thin device to be less than the reported size of the device. It is also permissible for the sum of the reported sizes of the thin devices using a given thin pool to exceed the total capacity of the data devices comprising the thin pool. In this case the thin pool is said to be *over-subscribed*. Over-subscribing allows the organization to present larger-than-needed devices to hosts and applications without having to purchase enough physical disks to fully allocate all of the space represented by the thin devices.

The capacity utilization of over-subscribed pools must be monitored to determine when space must be added to the thin pool to avoid out-of-space conditions.

Not all operating systems, filesystems, Logical Volume managers, multipathing software, and application environments will be appropriate for use with over-subscribed thin pools. If the application, or any part of the software stack underlying the application, has a tendency to produce dense patterns of writes to all available storage, thin devices will tend to become fully allocated quickly. If thin devices belonging to an over-subscribed pool are used in this type of

environment, out-of-space and undesired conditions may be encountered before an administrator can take steps to add storage capacity to the thin data pool. Such environments are called *thin-hostile*.

## Thin-hostile environments

There are a variety of factors that can contribute to making a given application environment thin-hostile, including:

- ⊗ One step, or a combination of steps, involved in simply preparing storage for use by the application may force all of the storage that is being presented to become fully allocated.
- ⊗ If the storage space management policies of the application and underlying software components do not tend to reuse storage that was previously used and released, the speed in which underlying thin devices become fully allocated will increase.
- ⊗ Whether any data copy operations (including disk balancing operations and de-fragmentation operations) are carried out as part of the administration of the environment.
- ⊗ If there are administrative operations, such as bad block detection operations or file system check commands, that perform dense patterns of writes on all reported storage.
- ⊗ If an over-subscribed thin device configuration is used with a thin-hostile application environment, the likely result is that the capacity of the thin pool will become exhausted before the storage administrator can add capacity unless measures are taken at the host level to restrict the amount of capacity that is actually placed in control of the application.

## Pre-provisioning with thin devices in a thin hostile environment

In some cases, many of the benefits of pre-provisioning with thin devices can be exploited in a thin-hostile environment. This requires that the host administrator cooperate with the storage administrator by enforcing restrictions on how much storage is placed under the control of the thin-hostile application.

For example:

- ⊗ The storage administrator pre-provisions larger than initially needed thin devices to the hosts, but only configures the thin pools with the storage needed initially. The various steps required to create, map, and mask the devices and make the target host operating systems recognize the devices are performed.
- ⊗ The host administrator uses a host Logical Volume manager to carve out portions of the devices into Logical Volumes to be used by the thin-hostile applications.

- ⊗ The host administrator may want to fully preallocate the thin devices underlying these Logical Volumes before handing them off to the thin-hostile application so that any storage capacity shortfall will be discovered as quickly as possible, and discovery is not made by way of a failed host write.
- ⊗ When more storage needs to be made available to the application, the host administrator extends the Logical Volumes out of the thin devices that have already been presented. Many databases can absorb an additional disk partition non-disruptively, as can most file systems and Logical Volume managers.
- ⊗ Again, the host administrator may want to fully allocate the thin devices underlying these volumes before assigning them to the thin-hostile application.

In this example it is still necessary for the storage administrator to closely monitor the over-subscribed pools. This procedure will not work if the host administrators do not observe restrictions on how much of the storage presented is actually assigned to the application.

## Host /boot, / (root), /swap, and /dump devices positioned on Symmetrix Virtual Provisioning (tdev) devices

A boot, root, swap, or dump (/boot, / (root), /swap, and /dump) device positioned on Symmetrix Virtual Provisioning (thin) device(s) is supported with Enginuity 5773 and later. However, some specific processes involving boot, root, swap, or dump (/boot, / (root), /swap, and /dump) devices positioned on thin devices should not have exposure to encountering the out-of-space condition.

Host-based processes such as kernel rebuilds, swap, dump, save crash, and Volume Manager configuration operations can all be affected by the thin provisioning out-of-space condition. This exposure is not specific to Dell EMC's implementation of Thin Provisioning. Dell EMC strongly recommends that the customer avoid encountering the out-of-space condition involving boot, root, swap, or dump (/boot, / (root), /swap, and /dump) devices positioned on Symmetrix VP (thin) devices using the following recommendations:

- ⊗ It is strongly recommended that Virtual Provisioning devices utilized for boot, root, swap, or dump volumes must be fully allocated<sup>1</sup> or the VP devices must not be oversubscribed<sup>2</sup>.

Should the customer use an over-subscribed thin pool, they should understand that they need to take the necessary precautions to ensure that they do not encounter the out-of-space condition.

---

**Note:** This warning applies to both the host server operating system and the virtual machines' operating systems that may reside on it if the server is configured for virtual machines.

---

1. A fully allocated Symmetrix VP (thin) device has 100% of the advertised space mapped to blocks in the data pool that it is bound to. This can be achieved by use of the Symmetrix VP pre-allocation mechanism or host-based utilities that will enforce pre-allocation of the space (such as, host device format.)
2. An over-subscribed Symmetrix VP (thin) device is a thin device, bound to a data pool, that does not have sufficient capacity to allocate for the advertised capacity of all the thin devices bound to that pool.

- ⊠ It is not recommended to implement space reclamation, available with Enginuity 5874 and later, with pre-allocated or over-subscribed Symmetrix VP (thin) devices that are utilized for host boot, root, swap, or dump volumes. Although not recommended, Space reclamation is supported on the listed types of volumes.

Should the customer use space reclamation on this thin device, they need to be aware that this freed space may ultimately be claimed by other thin devices in the same pool and may not be available to that particular thin device in the future.

## Cluster configuration

When using high availability in a cluster configuration, it is expected that no single point of failure exists within the cluster configuration and that one single point of failure will not result in data unavailability, data loss, or any significant application becoming unavailable within the cluster. Virtual provisioning devices (thin devices) are supported with cluster configurations; however, over-subscription of virtual devices may constitute a single point of failure if an out-of-space condition should be encountered. To avoid potential single points of failure, appropriate steps should be taken to avoid under-provisioned virtual devices implemented within high availability cluster configurations.



## Operating system characteristics

Most host applications will behave in a similar manner, in comparison to the normal devices, when writing to thin devices. This same behavior can also be observed as long as the thin device written capacity is less than the thin device subscribed capacity. However, issues can arise when the application writes beyond the provisioned boundaries.

This section describes operating system characteristics when applications write data beyond the provisioned boundary (out-of-space or pool exhaustion). The following were examined and will be further discussed in this section:

- ☒ “Thin pool exhaustion” on page 245
- ☒ “Filesystem utilities” on page 245
- ☒ “Filesystems” on page 246
- ☒ “Volume Managers” on page 247
- ☒ “Path Management” on page 248
- ☒ “Pre-provisioned thin devices” on page 250

### Thin pool exhaustion

With the current behavior of the Linux operating system, file systems, and multipathing software supported in the [Dell EMC Simple Support Matrix](#), the exhaustion of the thin pool causes undesired results.

Specifically, the Symmetrix will return a check condition on a *write(2)* when this occurs with an **04/44/00** check condition, **Hardware Error/Internal target failure**, returned in the sense code. This resulted in the following messages being seen in `/var/log/messages` when the default reporting is used.

- ☒ RHEL

```
Jan 18 09:43:33 182bi054 kernel: lost page write due to I/O error on emcpowerd1
Jan 18 09:43:34 182bi054 kernel: Buffer I/O error on device emcpowerd1, logical block 103877
```

- ☒ SLES

```
Jan 18 10:23:21 182bi228 kernel: sd 2:0:1:197: SCSI error: return code = 0x08000002
Jan 18 10:23:21 182bi228 kernel: sdbn: Current: sense key: Hardware Error
Jan 18 10:23:21 182bi228 kernel: Additional sense: Internal target failure
Jan 18 10:23:21 182bi228 kernel: end_request: I/O error, dev sdbn, sector 0
Jan 18 10:23:21 182bi228 kernel: Buffer I/O error on device sdbn, logical block 0
Jan 18 10:23:21 182bi228 kernel: lost page write due to I/O error on sdbn
Jan 18 10:23:23 182bi228 kernel: SCSI device sdbn: 7680000 512-byte hdwr sectors
```

### Filesystem utilities

The following filesystem utilities were examined. Findings include:

- fdisk(8)** **fdisk(8)** will report no errors while creating the partition tables if there is no space left to write one.
- mke2fs(8)** **mke2fs(8)** produced no errors while creating filesystems when space was exhausted and while mounting the filesystems created under these conditions. However, there was an error reported by `ext3fs`.

```
Jan 17 16:05:57 182bi189 kernel: JBD: no valid journal superblock found
Jan 17 16:05:57 182bi189 kernel: EXT3-fs: error loading journal.
```

**mkreiserfs(8)** **mkreiserfs(8)** produced errors while creating filesystem when space was exhausted.

```
Jan 17 16:19:04 182bi189 kernel: ReiserFS: dm-1: found reiserfs format "3.6" with standard
journal
Jan 17 16:19:04 182bi189 kernel: ReiserFS: dm-1: using ordered data mode
Jan 17 16:19:04 182bi189 kernel: reiserfs: using flush barriers
Jan 17 16:19:04 182bi189 kernel: ReiserFS: dm-1: journal params: device dm-1, size 8192, journal
first block 18, max trans len 1024, max batch 900, max commit age 30, max trans age 30
Jan 17 16:19:04 182bi189 kernel: ReiserFS: dm-1: checking transaction log (dm-1)
Jan 17 16:19:04 182bi189 kernel: reiserfs: disabling flush barriers on dm-1
Jan 17 16:19:04 182bi189 kernel: SCSI error : <3 0 3 196> return code = 0x8000002
Jan 17 16:19:04 182bi189 kernel: Current sdbb: sense key Hardware Error
Jan 17 16:19:04 182bi189 kernel: Additional sense: Internal target failure
Jan 17 16:19:04 182bi189 kernel: end_request: I/O error, dev sdbb, sector 66064
Jan 17 16:19:04 182bi189 kernel: ReiserFS: dm-1: warning: journal-837: IO error during journal
replay
Jan 17 16:19:04 182bi189 kernel: ReiserFS: dm-1: warning: Replay Failure, unable to mount
Jan 17 16:19:04 182bi189 kernel: ReiserFS: dm-1: warning: sh-2022: reiserfs_fill_super: unable
to initialize journal space
```

**mkfs.xfs(8)** **mkfs.xfs(8)** produced no errors while creating filesystems when space was exhausted but did fail to mount.

```
Jan 17 15:36:44 182bi189 kernel: SGI XFS with ACLs, security attributes, realtime, large block
numbers, dmapi support, no debug enabled
Jan 17 15:36:44 182bi189 kernel: SGI XFS Quota Management subsystem
Jan 17 15:36:44 182bi189 kernel: XFS mounting filesystem dm-1
Jan 17 15:36:44 182bi189 kernel: XFS: totally zeroed log
Jan 17 15:36:44 182bi189 kernel: Ending clean XFS mount for filesystem: dm-1
Jan 17 15:40:30 182bi189 kernel: SCSI error : <3 0 3 196> return code = 0x8000002
Jan 17 15:40:30 182bi189 kernel: Current sdbb: sense key Hardware Error
Jan 17 15:40:30 182bi189 kernel: Additional sense: Internal target failure
Jan 17 15:40:30 182bi189 kernel: end_request: I/O error, dev sdbb, sector 4792736
Jan 17 15:40:30 182bi189 kernel: I/O error in filesystem ("dm-1") meta-data dev dm-1 block
0x492020 ("xlog_iodone") error 5 buf count 1536
Jan 17 15:40:30 182bi189 kernel: xfs_force_shutdown(dm-1,0x2) called from line 966 of file
fs/xfs/xfs_log.c. Return address = 0xf9b3e87c
Jan 17 15:40:30 182bi189 kernel: Filesystem "dm-1": Log I/O Error Detected. Shutting down
filesystem: dm-1
Jan 17 15:40:30 182bi189 kernel: Please umount the filesystem, and rectify the problem(s)
Jan 17 15:41:33 182bi189 kernel: xfs_force_shutdown(dm-1,0x1) called from line 353 of file
fs/xfs/xfs_rw.c. Return address = 0xf9b3e87c
```

## Filesystems

When a filesystem runs out of space the user is immediately notified when a **write(2)** command is issued that the filesystem is out of space and they cannot continue. The **write(2)** does not get any further than this, no data is sent to page cache. This happens because the filesystem is tracking the space it has and how much it has used.

However, this is not the case with thin devices. In an oversubscribed under provisioned scenario the filesystem does not realize that there is no more space available and the Symmetrix returns an **04/44/00** check condition – **Hardware Error/Internal target failure**. Depending on when this error is returned, it has differing effects on the application, filesystem, and presentation to the user.

**Ext2fs** The ext2 filesystem is a non-journal filesystem. The filesystem showed no reaction to an application writing data to a device that could not allocate any more extents. The filesystem believes the space is available and passes the data to page cache. Once the page cache fault is encountered, the error is passed up to the application and is at the discretion of the application. However, if the application has closed out, no notification will reach it.

**Ext3fs** The ext3 filesystem is a journal filesystem. The filesystem showed no reaction to an application writing data to a device that could not allocate any more extents. The filesystem believes the space is available and passes the data to page cache. Once the page cache fault is encountered the error is passed up to the application and is at the discretion of the application. However, if the application has closed out, no notification will reach it.

However, this was not the case when it updates its journal. Once the error occurs during the commit the filesystem is remounted as READ ONLY.

```
Jan 18 09:57:38 182bi054 kernel: lost page write due to I/O error on emcpowerd1
Jan 18 09:57:38 182bi054 kernel: Aborting journal on device emcpowerd1.
Jan 18 09:57:38 182bi054 kernel: __journal_remove_journal_head: freeing b_committed_data
Jan 18 09:57:41 182bi054 kernel: ext3_abort called.
Jan 18 09:57:41 182bi054 kernel: EXT3-fs error (device emcpowerd1): ext3_journal_start_sb:
Detected aborted journal
Jan 18 09:57:41 182bi054 kernel: Remounting filesystem read-only
```

**Reiserfs** The Reiser filesystem is a journal filesystem. The filesystem showed no reaction to an application writing data to a device that could not allocate any more extents. The filesystem believes the space is available and passes the data to page cache. Once the page cache fault is encountered the error is passed up to the application and is at the discretion of the application. However, if the application has closed out, no notification will reach it.

However, this was not the case when it updates its journal.

```
Jan 17 12:22:42 182bi189 kernel: SCSI error : <3 0 2 194> return code = 0x8000002
Jan 17 12:22:42 182bi189 kernel: Current sdaj: sense key Hardware Error
Jan 17 12:22:42 182bi189 kernel: Additional sense: Internal target failure
Jan 17 12:22:42 182bi189 kernel: end_request: I/O error, dev sdaj, sector 334472
Jan 17 12:22:42 182bi189 kernel: REISERFS: abort (device dm-0): Journal write error in
flush_commit_list
Jan 17 12:22:42 182bi189 kernel: REISERFS: Aborting journal for filesystem on dm-0
```

**Xfs** The xfs filesystem is a journal filesystem. This filesystem responded to system errors immediately.

```
Jan 17 15:47:05 182bi189 kernel: I/O error in filesystem ("dm-1") meta-data dev dm-1 block
0x492020 ("xlog_iodone") error 5 buf count 1024
Jan 17 15:47:05 182bi189 kernel: xfs_force_shutdown(dm-1,0x2) called from line 966 of file
fs/xfs/xfs_log.c. Return address = 0xf9b3e87c
Jan 17 15:47:05 182bi189 kernel: Filesystem "dm-1": Log I/O Error Detected. Shutting down
filesystem: dm-1
Jan 17 15:47:05 182bi189 kernel: Please umount the filesystem, and rectify the problem(s)
```

**VxFS** The VxFS filesystem is an integral part of the Veritas volume manager. The filesystem responded immediately to the system errors.

```
Jan 24 10:09:59 182bi228 kernel: VxVM vxio V-5-0-2 Subdisk dg301-01 block 38608: Uncorrectable
write error
Jan 24 10:09:59 182bi228 kernel: vxfs: msgcnt 1 msg 038: V-2-38: vx_dataioerr -
/dev/vx/dsk/dg3/dg3vol01 file system file data write error in dev/block 0/19304
```

## Volume Managers

The following Volume Managers were examined. Findings include:

**Linux Native Volume Manager** LVM did not appear to be affected in this scenario.

**Veritas Volume Manager** The process of initializing the physical disks will fail if there is no space left in the pool.

```
Jan 23 10:27:33 182bi228 kernel: Buffer I/O error on device VxDMP2, logical block 959984
Jan 23 10:27:33 182bi228 kernel: lost page write due to I/O error on VxDMP2
```

## Path Management

The following Path Management software were examined. Findings include:

**PowerPath** 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 appearing in RHEL 6, SLES 10 SP4, and SLES 11 that shows an altered behavior where the path is faulted when an under provisioned pool is exhausted.

Figure 70 through Figure 73 show examples of SLES.

```
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 70 Filesystem mpathq-part1 made from thin LUN

```
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 71 Provisioned size: 200G; actual allocated size:17G

```

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 72 I/O error being seen in /var/log/messages when pool being exhausted

```

SUSE10SP4:~ # multipath -l
mpathq (3600601608a903200882fdeb05b6e211) dm-6 DGC,VRAID
[size=200G][features=1 queue_if_no_path][hw_handler=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][hw_handler=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][hw_handler=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][hw_handler=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 (3600601608a9032002a6c340ecea211) dm-0 DGC,RAID 5
[size=2.1T][features=1 queue_if_no_path][hw_handler=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][hw_handler=1 emc]
  \_ round-robin 0 [prio=0][active]

```

Figure 73 Thin LUN paths both marked as failed

## **Pre-provisioned thin devices**

In the conditions where the host is exposed to pre-provisioned thin devices that had not been bound to the thin pool, no negative effects have been noted.

# CHAPTER 10

## VPLEX

This chapter describes VPLEX-specific configuration in the Linux environment and contains support information.

☒ VPLEX overview.....	252
☒ Prerequisites.....	253
☒ Host Configuration for Linux: Fibre Channel HBA Configuration....	254
☒ Provisioning and exporting storage .....	263
☒ Storage volumes.....	265
☒ System volumes .....	267
☒ Required storage system setup.....	268
☒ Host connectivity .....	270
☒ Exporting virtual volumes to hosts .....	271
☒ Front-end paths .....	274
☒ Configuring Linux hosts to recognize VPLEX volumes .....	276
☒ Linux native cluster support .....	277
☒ Optimal-Path-Management (OPM) feature .....	282

## VPLEX overview

For detailed information about VPLEX, refer to the documentation available at [Dell EMC Online Support](#).

## VPLEX documentation

Refer to 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 at [Dell EMC E-Lab Navigator](#)

For the most up-to-date support information, always refer to the [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:** Always refer to 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**.

### **IMPORTANT**

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 seconds 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 Veritas 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 explains FC HBA related configuration details that must be addressed when using Fibre Channel with VPLEX.

The values provided are required and optimal for most scenarios. However, in extreme scenarios the values might 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 on how to monitor VPLEX performance refer to 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 and execution throttle for QLogic

---

**Note:** Changing the HBA queue depth is designed for advanced users. Increasing the queue depth might cause hosts to over-stress the arrays that are connected to the Linux host, resulting in performance degradation while communicating with them.

---

- ☒ **Execution throttle** setting—This setting controls the amount of outstanding I/O requests per HBA port. Set the HBA execution throttle to the QLogic 65535 default value. You can do this on the HBA firmware level by using the HBA BIOS or QConvergenceConsole 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 refer to the HBA vendor documentation.

To set the queue depth on the Qlogic FC HBA:

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 hex (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.

---

**Note:** The value is in hexadecimal: **20 hex (32 decimal)**. If the execution throttle is less than the queue depth, then the execution throttle overrides the queue depth setting.

---

```
# cat /sys/module/qla2xxx/parameters/ql2xmaxqdepth
20
```

To set the execution throttle on the Qlogic FC HBA:

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
2. 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 Assigned WWN : N/A
Press <Enter> to continue:

```

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

---

**Note:** Changing the HBA queue depth is designed for advanced users. Increasing the queue depth may cause hosts to over-stress other arrays connected to the Linux host, resulting in performance degradation while communicating with them.

---

- ☒ **HBA Queue depth**—This setting controls the amount 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**

---

**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 refer to the HBA vendor documentation.

To set the queue depth on the Emulex FC HBA:

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 the new queue depth 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 for the following:

- ☒ “VPLEX with GeoSynchrony v4.x” on page 263
- ☒ “VPLEX with GeoSynchrony v5.x” on page 264
- ☒ “VPLEX with GeoSynchrony v6.x” on page 264

### 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:

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. Refer to the EMC VPLEX Management Console Help or the *EMC VPLEX CLI Guide*, located at [Dell EMC Online Support](#), for more information.

[Figure 74 on page 264](#) illustrates the provisioning and exporting process.

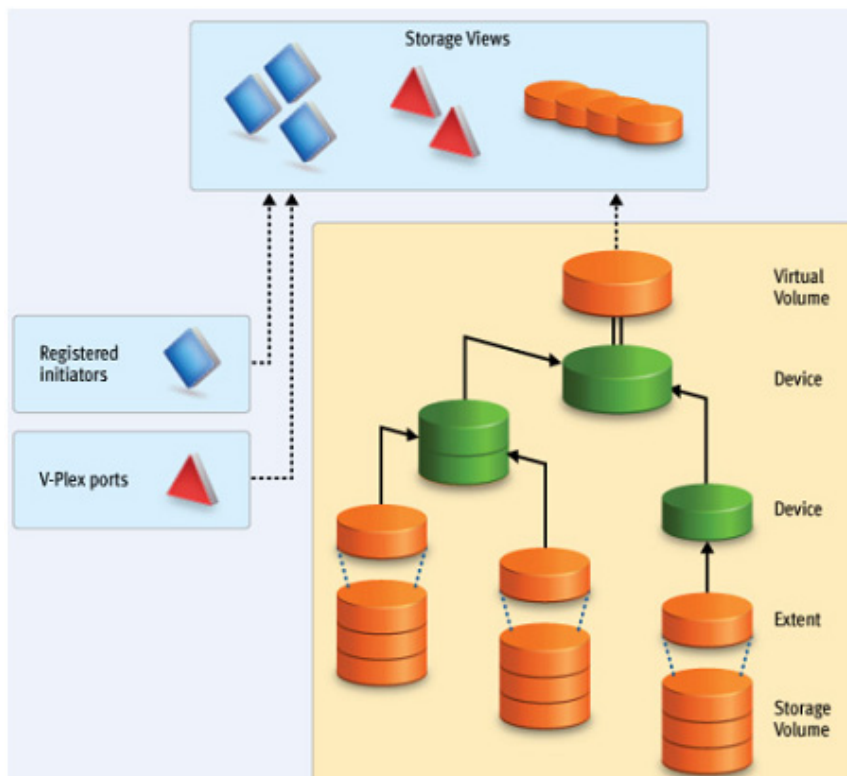


Figure 74 VPLEX provisioning and exporting storage process

## 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 of 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, refer to the *EMC VPLEX with GeoSynchrony 5.0 Product Guide*, located at [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, refer to the *EMC VPLEX with GeoSynchrony 6.0 Product Guide* 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.

---

**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 helps during troubleshooting and performance analysis.

---

This section contains the following information:

- ☒ [“Claiming and naming storage volumes ” on page 265](#)
- ☒ [“Extents ” on page 265](#)
- ☒ [“Devices ” on page 265](#)
- ☒ [“Distributed devices” on page 266](#)
- ☒ [“Rule sets” on page 266](#)
- ☒ [“Virtual volumes ” on page 266](#)

### Claiming and naming storage volumes

You must claim storage volumes before you can use them in the cluster (with the exception of the metadata volume, which is created from an unclaimed storage volume). Only after claiming a storage volume can you 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. A device's storage capacity 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 therefore are used only in multi-cluster 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 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, refer to the *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 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 on VPLEX Witness, refer to the *EMC VPLEX with GeoSynchrony 5.0 Product Guide*, located at [Dell EMC Online Support](#).

## Virtual volumes

Virtual volumes are created on devices or distributed devices and presented to a host via 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 created from storage devices. There are two types of system volumes. Each is briefly discussed in this section:

- ☒ “Metadata volumes” on page 267
- ☒ “Logging volumes” on page 267

## 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 as well as any distributed configuration information shared between clusters.

For more information about metadata volumes for VPLEX with GeoSynchrony v4.x, refer to the *EMC VPLEX CLI Reference Guide* on Dell EMC Online Support.

For more information about metadata volumes for VPLEX with GeoSynchrony v5.x, refer to the *EMC VPLEX with GeoSynchrony 5.0 and Point Releases Product Guide* on Dell EMC Online Support.

For more information about metadata volumes for VPLEX with GeoSynchrony v6.x, refer to the *EMC VPLEX with GeoSynchrony 6.0 Product Guide* 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, refer to the *EMC VPLEX CLI Guide*, available at [Dell EMC Online Support](#).

For more information about logging volumes for VPLEX with GeoSynchrony v5.x, refer to the *EMC VPLEX with GeoSynchrony 5.0 and Point Releases Product Guide* on Dell EMC Online Support.

For more information about metadata volumes for VPLEX with GeoSynchrony v6.x, refer to the *EMC VPLEX with GeoSynchrony 6.0 Product Guide* on Dell EMC Online Support.

## Required storage system setup

Symmetrix, VNX series, VNXe series, Unity series, CLARiiON, and XtremIO product documentation and installation procedures for connecting a Symmetrix, VNX series, VNXe series, CLARiiON, or XtremIO storage system to a VPLEX instance are available at [Dell EMC Online Support](#).

## Required Symmetrix FA bit settings

For Symmetrix-to-VPLEX connections, configure the Symmetrix Fibre Channel directors (FAs) as shown in [Table 23](#).

**Note:** Dell EMC recommends that you download the latest information before installing any server.

**Table 23** Required Symmetrix FA bit settings

Set <sup>1</sup>	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 <sup>2</sup>

1. For the Symmetrix 8000 series, the flags should be Unique Worldwide Name (UWN), Common Serial Number, and Enable Point-to-Point (PP).
2. 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, Unity series, or CLARiiON system, 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, Unity series, or CLARiiON system, 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 [Table 23](#) are for connectivity of VPLEX to a Symmetrix array only. For Host to Symmetrix FA bit settings, refer to 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.

Refer to the *VPLEX Procedure Generator*, available at [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

Refer to the *VPLEX Procedure Generator*, available at [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 on qualified switches, hosts, host bus adapters, and software, always consult the [Dell EMC Simple Support Matrix](#), available through E-Lab Interoperability Navigator (ELN), under the **PDFs and Guides** tab, 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.

---

**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 included in the storage view will be able to access the virtual volume.

The virtual volumes 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 together 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.

**Note:** When performing encapsulation with Citrix XenServer, apply Citrix Hot Fix XS62E003 to XenCenter.

Exporting virtual volumes consists of the following tasks:

1. Creating a storage view, as shown in Figure 75.

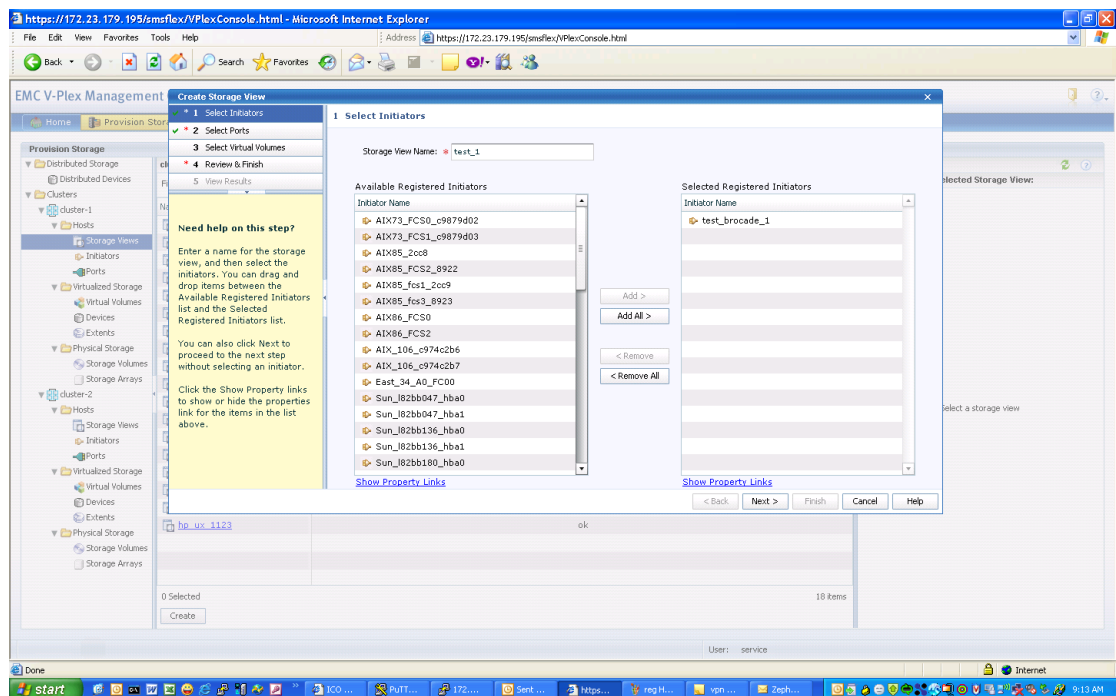


Figure 75 Create storage view

2. Registering initiators, as shown in Figure 76.

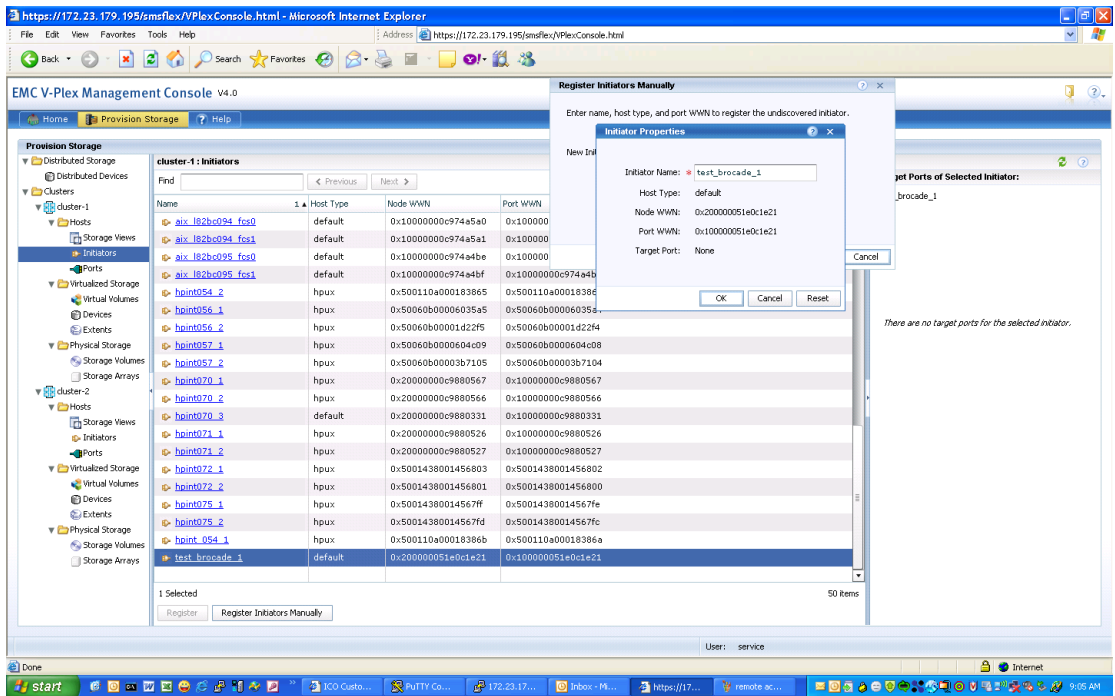


Figure 76 Register initiators

**Note:** When initiators are registered, you can set their type as indicated in Table 24 on page 275.

3. Adding ports to the storage view, as shown in Figure 77.

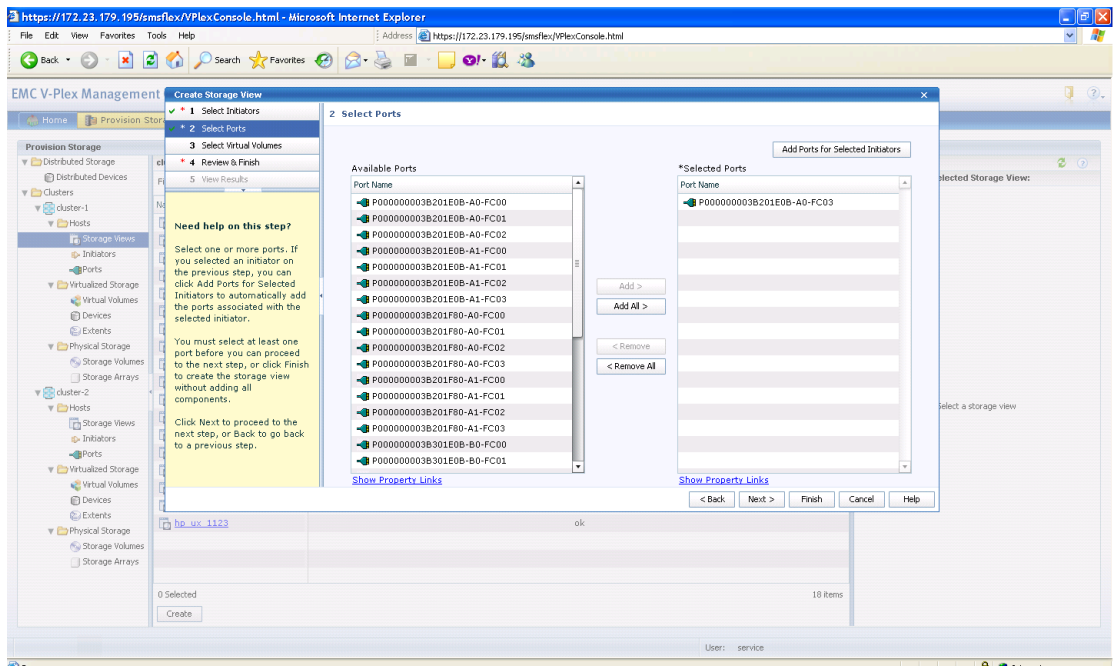


Figure 77 Add ports to storage view

4. Adding virtual volumes to the storage view, as shown in Figure 78.

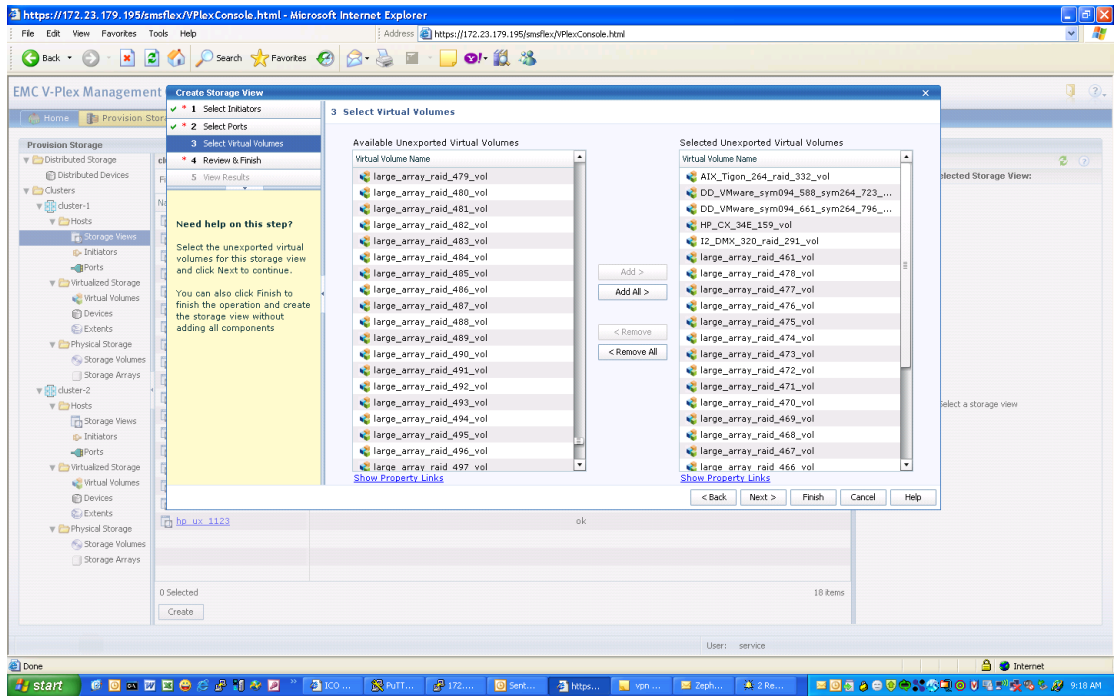


Figure 78 Add virtual volumes to storage view

## Front-end paths

This section contains the following information:

- ☒ “Viewing the World Wide Name for an HBA port” on page 274
- ☒ “VPLEX ports” on page 274
- ☒ “Initiators” on page 274

### 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 Solution Enabler
- ☒ **syminq** command (Symmetrix users)
- ☒ The Linux sysfs filesystem; for example

```
# cat /sys/class/fc_host/host1/port_name
0x10000000c93dc171
#
```

### VPLEX ports

The virtual volumes 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 listed in [Table 24](#). When initiators are registered, you can set their type, also indicated in [Table 24](#).

**Table 24** Supported hosts and initiator types

<b>Host</b>	<b>Initiator type</b>
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. Refer to [“LUN scanning mechanisms”](#) on page 26 for information on 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. Figure 79 shows an example of RCHS in a non-cross-cluster connected configuration.

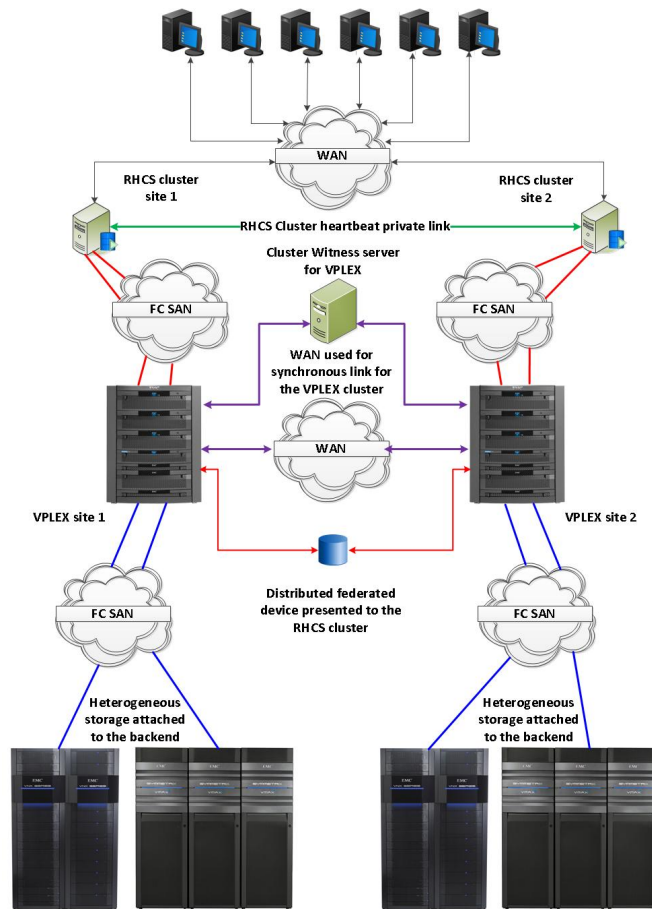


Figure 79 Red Hat RCHS in a non-cross-cluster connected configuration

Figure 80 shows an example of HAE in a cross-cluster connected configuration.

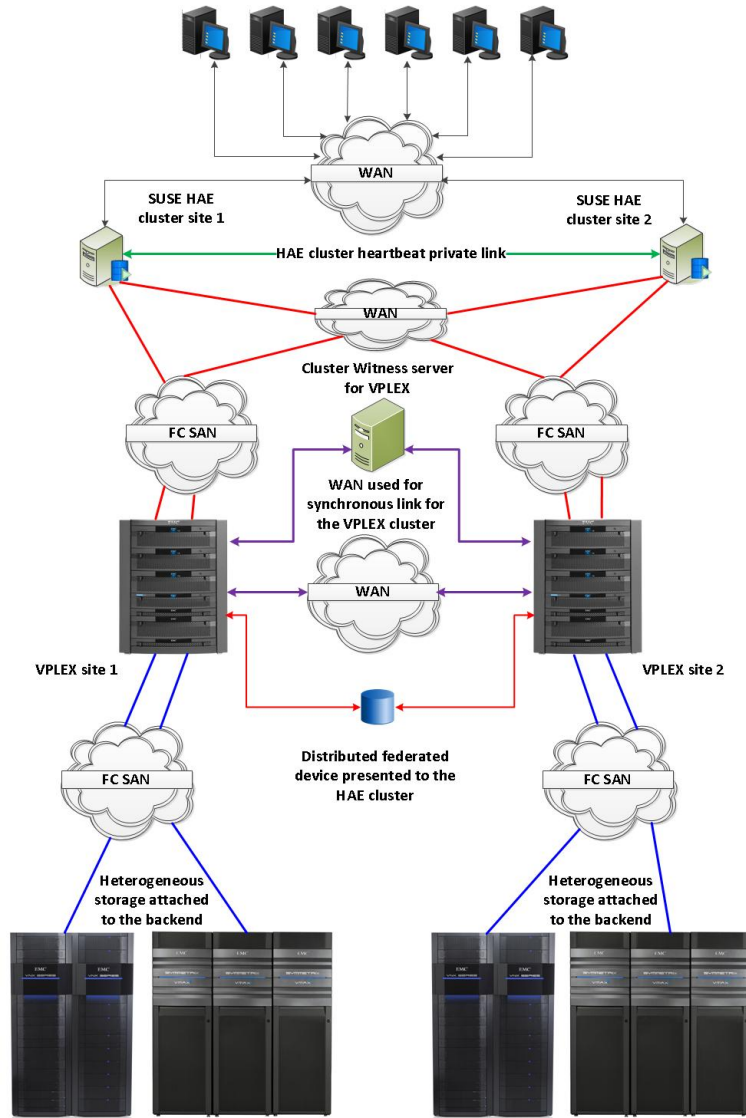


Figure 80 SUSE HAE in a cross-cluster connected configuration

## Supported Red Hat RHCS configurations and best practices

Consider the following support information and best practices.

- Supported**
- ☒ RHEL RHCS 5.5 and later, and RHEL RHCS 6.x
  - ☒ Minimum GeoSynchrony 5.0.1
  - ☒ Active/passive or active/active cluster
  - ☒ Stretched storage cluster without cross-connect

- ☒ Stretched storage cluster with cross-cluster connect
  - < 1ms round trip time latency for connections between the two sites
  - PowerPath only
- ☒ PowerPath or Linux native DM-MPIO
- ☒ RHEL GFS and GFS2 clustered filesystem
- ☒ RHCS quorum disk
- ☒ RHCS fence\_scsi() and sg3util sg\_persist()

### Best practices

- ☒ Stretched storage cluster without cross-connect
  - For Red Hat's stretch cluster support, refer to Red Hat Knowledgebase Article ID: 58412, [Support for Red Hat Enterprise Linux High Availability Cluster Stretch and Multi-Site Architectures](#).
- ☒ Server cluster heartbeat limitation per Red Hat specification (< 2.0ms RTT)
  - Refer to Red Hat Knowledgebase Article ID: 58412, [Support for Red Hat Enterprise Linux High 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 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 330000 ms. 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.

- ☒ 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;
}
}

```

☒ 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);
}

```

## 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 cluster 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 filesystem
  - ☒ STONITH block device (sbd)
- Best practices**
- ☒ Reference *SUSE Linux Enterprise High Availability Extension* documentation
  - ☒ It is recommended that SUSE technical support be 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 seconds
    - 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 utilized ([http://www.linux-ha.org/wiki/SBD\\_Fencing](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 filesystem 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.

## Optimal-Path-Management (OPM) feature

Starting with VPLEX 5.5 Acropolis, Optimal-Path-Management (OPM) was introduced to improve VPLEX performance. OPM utilizes the ALUA mechanism to spread the IO load across VPLEX directors while gaining cache locality.

### VPLEX OPM feature overview

The OPM feature in VPLEX FE uses the SCSI ALUA (Asymmetric Logical Unit Access) 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.

Figure 81 and Figure 82 show the cache coherency overhead reductions on the local site, first without OPM, and then with OPM.

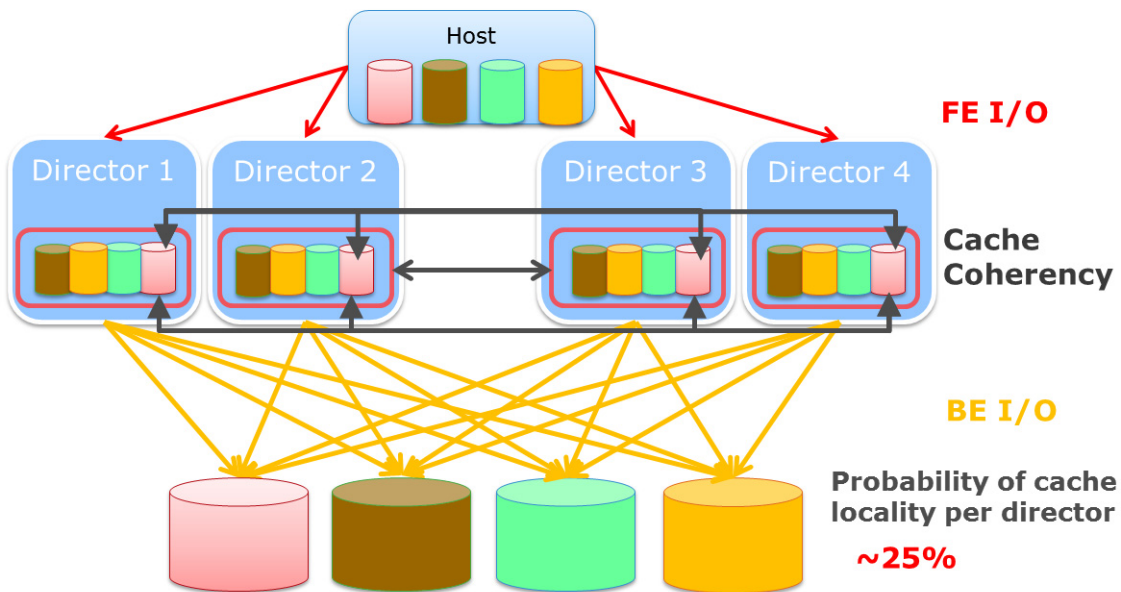
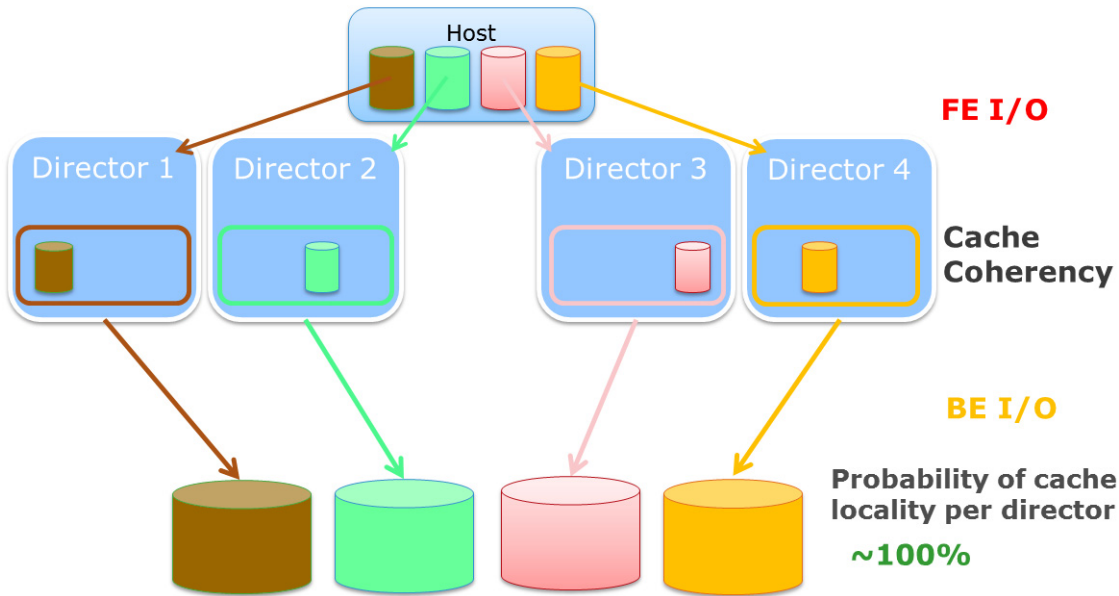


Figure 81 Cache coherency overhead reduction without OPM multiple volumes



**Figure 82** 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 provisioned to a host, you can check out its optimal director using VPLEX CLI. [Figure 83](#) shows an example of a volume in a metro cluster.

```

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]  []
    
```

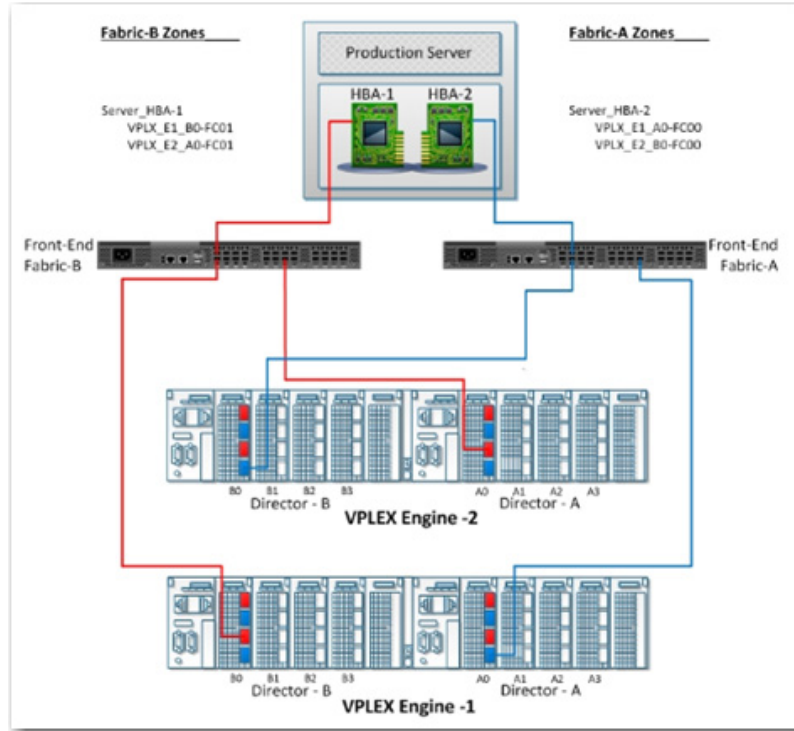
**Figure 83** Volume in a metro cluster example

## 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.

Figure 84 shows the best practice host connectivity for dual VPLEX engines.



**Figure 84** Host connectivity across dual VPLEX engines

Figure 85 shows the best practice host connectivity quad engines in a VPLEX cluster.



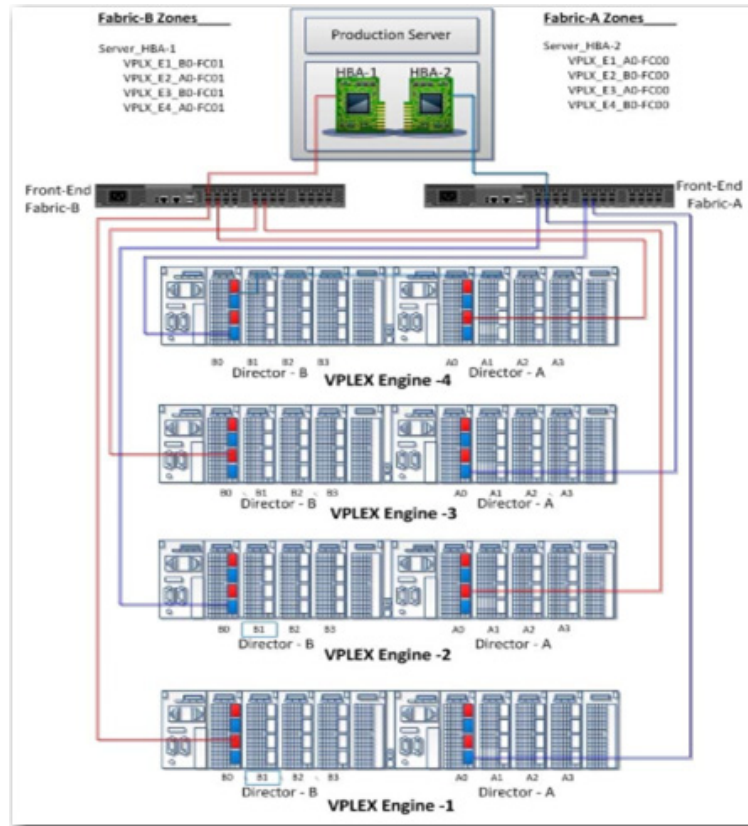


Figure 85 Host connectivity across quad engines in a VPlex cluster

## 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.

**Note:** Veritas DMP does not support OPM ALUA.

**PowerPath** PowerPath versions that support VPlex OPM are summarized in the following table.

Table 25 PowerPath versions that support VPlex OPM

	PP 6.0 and later	PP 5.9.x	PP 5.8.x	PP 5.7.x	PP 5.5.x
Linux	Yes	N/A	N/A	No	No

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 Figure 86.

```
Pseudo name=emcpower6a
VPLEX ID=PMTELABR000001
Logical device ID=6000144000000010F00333BB962AC0D8 [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
=====
```

Figure 86 Output from the powermt command

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 Figure 87.

```
Administrator@A50T3234 ~
$ powermt display alua dev=all
Pseudo name=harddisk1
VPLEX ID=FNM00141000019
Logical device ID=60001440000000107043F71D5C566D82
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
=====
```

Figure 87 Output from the powermt command

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 Figure 88.

```

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                   P00000000477043F7-B0-FC03.0  director-1-2-B
CL1-0C                   P00000000477043F7-B0-FC02.0  director-1-2-B
CL1-0D                   P0000000047704419-B0-FC03.0  director-1-1-B
CL1-0A                   P00000000476043F7-A0-FC02.0  director-1-2-A
CL1-0B                   P0000000047604419-A0-FC03.0  director-1-1-A
CL1-05                   P0000000047704419-B0-FC01.0  director-1-1-B
CL1-06                   P00000000476043F7-A0-FC01.0  director-1-2-A
CL1-03                   P0000000047604419-A0-FC01.0  director-1-1-A
CL1-04                   P00000000477043F7-B0-FC00.0  director-1-2-B
CL1-09                   P0000000047704419-B0-FC02.0  director-1-1-B
CL1-07                   P0000000047604419-A0-FC02.0  director-1-1-A
CL1-08                   P00000000477043F7-B0-FC01.0  director-1-2-B
CL1-01                   P0000000047704419-B0-FC00.0  director-1-1-B
CL1-02                   P00000000476043F7-A0-FC00.0  director-1-2-A
CL1-00                   P0000000047604419-A0-FC00.0  director-1-1-A
    
```

**Figure 88** Output from the export storage-view command

Map each **Storage Interface** value to its corresponding director name, as shown in Table 26. The optimal paths detected in the host lead to director-1-1-A.

**Table 26** mapping Storage Interface values to corresponding director name

Interface	Director	Path ALUA state	Interface	Director	Path ALUA state
CL1-08	Director-1-1-B	Non-optimized	CL1-04	Director-1-2-B	Non-optimized
CL1-06	Director-1-2-A	Non-optimized	CL1-02	Director-1-2-A	Non-optimized
CL1-05	Director-1-1-B	Non-optimized	CL1-01	Director-1-1-B	Non-optimized
CL1-03	Director-1-1-A	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 Figure 89, the ID matches the volume name **rc1\_lr0\_0000\_vol**.

```

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  -
VPD83T3:6000144000000107043f71d5c566d82
rc1_lr0_0001_vol  ok  ok  unexported  4194304  4K  16G  local  rc1_lr0_0001  synchronous  true  0B  -
VPD83T3:6000144000000107043f71d5c566d83
    
```

**Figure 89** Output from the ll command

In the VPLEX CLI, run the `opm virtual-volume show-state [volume name]` command, as shown in the example in Figure 90.

```

VPlexcli:/> opm virtual-volume show-state rc1_lr0_0000_vol
cluster-1
Virtual Volumes      Optimal Directors      Override Directors
-----
rc1_lr0_0000_vol    [director-1-1-A]      []
    
```

**Figure 90** 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 seconds.
- ☒ In the host, run the `powermt display perf dev=[ dev name | all]` command. For example, as shown in the example in Figure 91.

```

Administrator@A50T3234 ~
$ powermt display perf dev=harddisk1
Timestamp = 15:14:05 UTC, 19 Jul 2015
Sample Interval = 60
Pseudo name=harddisk1
VPLEX_ID=FNW00141000019
Logical device ID=6000014400000000107043F71D5C566D82
state=alive; policy=ADaptive; queued-IOS=1

Read bytes/s      KB<=4      4<KB<=8      8<KB<=128      KB>128      All
Write bytes/s     3.58M      -              -              -           3.58M
Total bytes/s     1.76M      -              -              -           1.76M

Read Avg Response ms 0.313      -              -              -           0.313
Write Avg Response ms 0.418      -              -              -           0.418
All Avg Response ms  0.348      -              -              -           0.348

=====
### HW Path      Host      I/O Paths      Metrics      Reads--Writes      Retry Error
                        delta delta
=====
  2 port2\path0\tgt0\lun0 c2t0d0
    Low Latency (ms) 0.063 0.192 0 0
    High Latency (ms) 14.3 13.9
  2 port2\path0\tgt1\lun0 c2t1d0
    Low Latency (ms) - - 0 0
    High Latency (ms) - -
  2 port2\path0\tgt2\lun0 c2t2d0
    Low Latency (ms) - - 0 0
    High Latency (ms) - -
  2 port2\path0\tgt3\lun0 c2t3d0
    Low Latency (ms) - - 0 0
    High Latency (ms) - -
  3 port3\path0\tgt0\lun0 c3t0d0
    Low Latency (ms) 0.068 0.191 0 0
    High Latency (ms) 14.2 23.8
  3 port3\path0\tgt1\lun0 c3t1d0
    Low Latency (ms) - - 0 0
    High Latency (ms) - -
  3 port3\path0\tgt2\lun0 c3t2d0
    Low Latency (ms) - - 0 0
    
```

**Figure 91** Output from the `powermt` command

Note 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.

---

**Note:** Refer to each of the operating system distributor's documentation for more 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
3600014400000010f00333bb8640ace9 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 `multipathd` service for the change to take effect. Verify the syntax with your Red Hat release.

### RHEL 6 & 7

```
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 Virtual Machine (OVM) / Oracle Unbreakable Enterprise Kernel 6 (UEK 6)

```

device {
    vendor "EMC"
    product "Invista"
    product_blacklist "LUNZ"
    path_grouping_policy group_by_prio
    getuid_callout "/lib/udev/scsi_id --whitelisted
--device=/dev/%n"
    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 1000
    rr_min_io_rq 1
    failback immediate
}

[root@lhx9ser5 ~]# 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

```

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

```

**Note:** For XenServer, starting with version 6.5, the above 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 `multipathd` 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@lhx9ser5 ~]# iostat -t 2 | grep sd[dwj]q
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 RHEL 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@lhx5ser3 ~]# 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 [Figure 92](#).

VPD ID	Name	VPD ID	VPD ID
VPD83T3:600014400000010f00333bb862ac37d	 	DD-Symm3050_0587__Symm0292_1714-Linux_vol	

**Figure 92** Storage view in VPLEX GUI manager

```

VPlexcli:/> opm virtual-volume show-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 at this time. 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, located at [Dell EMC E-Lab Navigator](#), for the current list of supported configurations.



# CHAPTER 11

## Native Clusters

This chapter provides information on native clusters.

☒ Supported clusters.....	294
☒ Red Hat Cluster Suite (RHCS) .....	295
☒ Heartbeat.....	299
☒ High Availability Extension (HAE).....	301

## 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 (RHCS)

Red Hat defines the Red Hat Cluster Suite (RHCS) 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 filesystem. 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. Always refer to the [Dell EMC Simple Support Matrix \(ESM\)](#) 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 filesystem
- ☒ RHCS quorum disk
  - With PowerPath, support began with RHEL 5.5 and continued with RHEL 6.0; see [“Best practices and additional installation information” on page 296](#).
- ☒ RHCS fence\_scsi() and sg3util sg\_persist()
  - With PowerPath, support began with RHEL 5.5 and continued with RHEL 6.0 and RHEL 7.0; see [“Best practices and additional installation information” on page 296](#).

RHCS ensures high availability and manageability of critical network resources including data, applications, and services. It is a multinode 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 (GFS)

Global File System (GFS) is an open source cluster filesystem developed by Red Hat. It operates within the framework provided by Red Hat Cluster Suite and enables shared storage and filesystems 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 filesystem, whereas, GFS2 can be used on a standalone system, like any other filesystem, or as a shared disk filesystem in a cluster.

While RHCS provides the infrastructure for configuring cluster nodes, maintaining cluster health, enabling services, etc., 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. Refer to “GFS” on page 336 for additional details regarding GFS.

## Best practices and additional installation information

- ☒ Server cluster heartbeat limitation per Red Hat specification (< 2.0ms RTT)
  - Refer to Red Hat Knowledgebase Article ID: 58412, *Support for Red Hat Enterprise Linux High Availability Cluster Stretch and Multi-Site 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;
}
}

```

- 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](#).

# Heartbeat

Heartbeat is an open source clustering software for Linux. It had its beginnings as the Linux-HA project with its first release in 1998; limited to two nodes and very simple takeover semantics, and no resource monitoring. In 2005 the project was expanded by the release of version 2 which added n-node clusters, resource monitoring, dependencies, and policies.

Dell EMC supports the Novell 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 multinode 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
- ☒ Multi-node 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
- ☒ Although shared disk systems are supported, they are not required
- ☒ Support for cluster aware logical volume managers like EVMS

## Heartbeat cluster components

The following components make up a Heartbeat version 2 cluster found in Novell'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 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 NIC 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 Novell's implementation of Heartbeat or to configure a cluster, go to the [Micro Focus \(Novell\) website](#).



## High Availability Extension (HAE)

Novell'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 Novell 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)
- ☒ Multi-node 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 Novell'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 NIC 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 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 Novell's implementation of High Availability Extension or to configure a cluster, go to the [Micro Focus \(Novell\) website](#).

# CHAPTER 12

## Reference: Supported Linux features and limitations

This chapter provides information on general features and limitations packaged as part of the Linux support on Dell EMC arrays.

☒ Filesystems and feature limitations .....	304
☒ Linux volume managers .....	307
☒ LUN limits .....	308
☒ PATH limits.....	309

## Filesystems and feature limitations

This section provides the following information:

- ☒ "Filesystem support" on page 304
- ☒ "Features and limitations" on page 305
- ☒ "Linux filesystems" on page 306

"Filesystem" is the general name given to the host-based logical structures and software routines that are used to control storage, organization, manipulation, and retrieval of data. Filesystems 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 filesystem is an ordered tree-like hierarchical structure 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 filesystem is made available to users by mounting it to a specific mounting point.

### Filesystem support

Dell EMC qualifies and supports a growing list of the more popular Linux filesystems, as listed in [Table 27](#).

**Table 27** Supported filesystems

	RHEL	OL	SuSE
Ext2	Yes	Yes	Yes
Ext3	Yes	Yes	Yes
Ext4	≥RHEL 5.3	Yes	Yes
ReiserFS	Yes	Yes	Yes
GFS2	RHEL 5.2	Yes	No
XFS	≥RHEL 5.5	≥OL6	Yes
OCFS2	≥RHEL 5.2	Yes	≥SLES 10 SP2
NSS	No	No	OES
VxFS	Yes	Yes	Yes
Btrfs	≥RHEL 7	≥OL7	≥SLES 12

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. Oracle Linux

The Red Hat Scalable File System Add-on is a solution which incorporates the Linux XFS filesystem and is available, for an additional cost per socket-pair, with the Red Hat Enterprise Linux Server subscription. 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 higher.

## Features and limitations

Table 28 and Table 29 summarize filesystem features and limitations.

**Table 28** Local filesystem features and limitations

Local filesystem	Compatibility	Capacity		Data structure	Journaling		
		Max. file size	Max. volume size		Block	Metadata only	Allocation techniques
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 Snapshot

**Table 29** Cluster filesystem features and limitations

Cluster filesystem	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
VxFS		2 <sup>63</sup> Byte (8 EiB)	2 <sup>77</sup> 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 gigabytes 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 physical disks as logical devices that are called volumes. A VxVM volume appears to applications and the operating system as a physical disk 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. By supporting the Redundant Array of Independent Disks (RAID) model, VxVM can be configured to protect against disk and hardware failure, and to increase I/O throughput. Additionally, VxVM provides features that enhance fault tolerance and fast recovery from disk failure.

For detailed documentation about Veritas VxVM and VxFS, refer to 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, refer to the EVMS website at <http://evms.sourceforge.net/>.

## LUN limits

The number of logical units seen by a host system is dependent on the SCSI scan algorithm 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 OS vendors for exact limitations.



## PATH limits

Device Mapper Multipathing enables an OS 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 softwares limits.

Reference: Supported Linux features and limitations

# CHAPTER 13

## Special Topics

This chapter provides the following information:

☒ Egenera .....	312
-----------------	-----

# Egenera

Dell EMC supports the Egenera BladeFrame platform from Egenera, Inc.

The Egenera BladeFrame is supported in both direct-attach and through FC-SAN. The Egenera c-Blades use Egenera native multipathing functionality packaged with the Egenera BladeFrame OS, which is supported with Symmetrix and VNX series or CLARiiON systems. Neither PowerPath or Linux DM-MPIO is needed or supported on Egenera platforms.

PowerPath is not supported on either the Egenera c-Blades or the Egenera p-Blades.

When attaching the Egenera BladeFrame to a VNX series or CLARiiON array the following settings need to be configured:

```
arraycommpath = enabled  
failovermode = 1
```

**Notes** Note the following:

- ☒ ALUA mode is currently not supported with the Egenera BladeFrame.
- ☒ When attached to a Symmetrix array the standard Linux director bits are used.
- ☒ Naviagent is supported running on the Egenera c-Blades only.

Refer to the [Dell EMC Simple Support Matrix](#) for supported Egenera BladeFrame OS versions, pBlade guest operating systems, and BladeFrame models.

Consult the Egenera website for more detailed SAN configuration documentation.