



# EMC XtremIO Storage Array

## Host Configuration Guide

P/N 302-001-521  
REV 21  
SEPTEMBER, 2016

Copyright © 2016 EMC Corporation. All rights reserved. Published in the USA.

Published September, 2016

EMC 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. EMC Corporation 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 EMC software described in this publication requires an applicable software license.

XtremIO, EMC<sup>2</sup>, EMC, and the EMC logo are registered trademarks or trademarks of EMC Corporation in the United States and other countries. All other trademarks used herein are the property of their respective owners.

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

# CONTENTS

## Preface

Pre-Requisites .....	12
----------------------	----

## Chapter 1

### Best Practices for Zoning and Subnetting

General Guidelines .....	16
Minimal Zoning/Subnetting Configuration .....	16
iSCSI SAN Guidelines .....	17
Fibre Channel SAN Guidelines .....	18
10TB Starter X-Brick (5TB) and Single X-Brick Cluster .....	18
Dual X-Brick Clusters .....	19
Clusters with Four X-Bricks (or More) .....	20

## Chapter 2

### Host Configuration for Linux®

Fibre Channel HBA Configuration .....	24
Pre-Requisites .....	24
Queue Depth and I/O Elevators .....	24
Setting the LUN Queue Depth and I/O Elevators, Using Linux Udev .....	26
Setting the LUN Queue Depth and I/O Elevators, Using CLI .....	28
Multipathing Software Configuration .....	33
Configuring Linux Native Multipathing .....	33
Configuring PowerPath Multipathing .....	36
Post-Configuration Steps - Using the XtremIO Cluster .....	38
Disk Formatting .....	38
Presenting XtremIO Volumes to the Linux Host .....	39
Partition Alignment in Linux .....	40
Logical Volume Manager (LVM) .....	40
Creating a File System .....	41
Using LUN 0 with XtremIO Storage .....	41
Space Reclamation .....	43
Configuring Boot from SAN with an XtremIO Volume .....	46

## Chapter 3

### Host Configuration for VMware® vSphere®

Fibre Channel HBA Configuration .....	48
Pre-Requisites .....	48
Queue Depth .....	48
Host Parameters Settings .....	53
vCenter Server Parameter Settings .....	55
vStorage API for Array Integration (VAAI) Settings .....	55
Enabling VAAI Features .....	56
Tuning VAAI XCOPY with XtremIO .....	58
Disabling VAAI in ESX .....	58
Configuring VMware vSphere with XtremIO Storage in a Multiple Storage Array Configuration .....	59
Multipathing Software Configuration .....	60
Configuring vSphere Native Multipathing .....	60
Configuring PowerPath Multipathing .....	64

	Post-Configuration Steps - Using the XtremIO Storage .....	65
	Disk Formatting.....	66
	Presenting XtremIO Volumes to the ESX Host .....	67
	Creating a File System .....	68
	Using LUN 0 with XtremIO Storage.....	68
	Virtual Machine Formatting .....	69
	Virtual Machine Guest OS Settings .....	71
	Space Reclamation .....	72
	Out of Space VM Suspend and Notification with Thin Provisioning (TPSTUN) .....	76
	Configuring Boot from SAN with an XtremIO Volume .....	76
<b>Chapter 4</b>	<b>Host Configuration for Microsoft® Windows®</b>	
	Fibre Channel HBA Configuration.....	78
	Pre-Requisites for FC HBA Configuration .....	78
	Queue Depth.....	78
	Windows Offloaded Data Transfer (ODX) Functionality .....	82
	Checking ODX Status .....	82
	Enabling ODX.....	82
	Disabling ODX.....	83
	Multipathing Software Configuration .....	84
	Configuring Native Multipathing Using Microsoft Multipath I/O (MPIO)..	84
	Configuring PowerPath Multipathing .....	86
	Post-Configuration Steps - Using the XtremIO Storage .....	88
	Disk Formatting.....	88
	Presenting XtremIO Volumes to the Windows Host .....	89
	Creating a File System .....	90
	Space Reclamation .....	90
	Disabling the UNMAP Feature.....	97
	Configuring the Volume Snapshot Service, Using XtremIO VSS Hardware Provider .....	97
	Configuring Boot from SAN with an XtremIO Volume.....	104
<b>Chapter 5</b>	<b>Host Configuration for Oracle® Solaris</b>	
	Fibre Channel HBA Configuration.....	106
	Pre-Requisites.....	106
	Queue Depth.....	106
	Host Parameters Settings.....	107
	Multipathing Software Configuration .....	113
	Configuring STMS to Manage MultiPath under Solaris .....	113
	Configuring PowerPath Multipathing .....	115
	Post-Configuration Steps - Using the XtremIO Cluster.....	116
	Disk Formatting.....	116
	Presenting XtremIO Volumes to the Host .....	117
	Overriding Disk Block Size Using ashift and Solaris 11.2 .....	118
	Logical Volume Manager (LVM) .....	121
	Creating a File System .....	121
	Configuring Boot from SAN with an XtremIO Volume.....	121

<b>Chapter 6</b>	<b>Host Configuration for AIX</b>	
	Fibre Channel HBA Configuration.....	124
	Queue Depth.....	124
	Fast I/O Failure for Fibre Channel Devices.....	126
	Fibre Channel Adapter Device Driver Maximum I/O Size .....	127
	EMC AIX ODM Installation Requirements and Procedure .....	128
	Multipathing Software Configuration .....	131
	Configuring AIX Native Multipathing .....	131
	Configuring PowerPath Multipathing .....	135
	Post-Configuration Steps - Using the XtremIO Cluster.....	139
	Disk Formatting.....	139
	Presenting XtremIO Volumes to the AIX Host .....	140
	Logical Volume Manager (LVM) .....	141
	Creating a File System .....	141
	Volume Configuration .....	142
	Using LUN 0 with XtremIO Storage .....	143
	Configuring Boot from SAN with an XtremIO Volume.....	144
 <b>Chapter 7</b>	 <b>Host Configuration for HP-UX</b>	
	Fibre Channel HBA Configuration.....	146
	Pre-Requisites.....	146
	Queue Depth.....	146
	Multipathing Software Configuration .....	147
	Configuring HP-UX Native Multipathing .....	147
	Configuring PowerPath Multipathing .....	148
	Post-Configuration Steps - Using the XtremIO Cluster .....	150
	Device Addressing .....	150
	Disk Formatting.....	150
	Presenting XtremIO Volumes to the HP-UX Host.....	151
	Logical Volume Manager (LVM) .....	152
	Creating a File System .....	152
	Space Reclamation .....	152
	Configuring Boot from SAN with an XtremIO Volume.....	153
 <b>Appendix A</b>	 <b>Configuring Boot from SAN with XtremIO</b>	
	Configuring the Qlogic HBA BIOS for SAN Boot .....	156
	Configuring the Emulex HBA BIOS for SAN Boot .....	159



# PREFACE

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

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

---

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

---

## NOTICE

Make sure to use the most recent revision of the XtremIO Host Configuration Guide. If this document is more than a month old, refer to EMC Online Support (<https://support.emc.com>) to check if a more recent document revision is published.

---

## Purpose

This document provides instructions for optimally configuring hosts for EMC XtremIO Storage Array.

## Audience

This document is intended for the host storage administrator, system administrator, or operator who will be involved in managing the XtremIO Storage Array.

## Related Documentation

The following EMC publications provide additional information:

- ◆ *XtremIO Storage Array User Guide*
- ◆ *XtremIO Storage Array Security Configuration Guide*
- ◆ *XtremIO Storage Array Release Notes*
- ◆ *XtremIO Simple Support Matrix*
- ◆ *EMC Host Connectivity with Qlogic Fibre Channel and iSCSI HBAs and Converged Network Adapters (CNAs) for the Linux Environment document*
- ◆ *EMC Host Connectivity with Emulex Fibre Channel and iSCSI HBAs and Converged Network Adapters (CNAs) for the Linux Environment*
- ◆ *EMC Host Connectivity with QLogic Fibre Channel and iSCSI Host Bus Adapters (HBAs) and Converged Network Adapters (CNAs) in the Windows Environment document*
- ◆ *EMC Host Connectivity with Emulex Fibre Channel and iSCSI Host Bus Adapters (HBAs) and Converged Network Adapters (CNAs) in the Windows Environment*

## Conventions Used in this Document

EMC uses the following conventions for special notices:

---

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

---

### Typographical conventions

EMC uses the following type style conventions in this document:

<b>Bold</b>	Use for names of interface elements, such as names of windows, dialog boxes, buttons, fields, tab names, key names, and menu paths (what the user specifically selects or clicks)
<i>Italic</i>	Use for full titles of publications referenced in text
Monospace	Use for: <ul style="list-style-type: none"> <li>• System output, such as an error message or script</li> <li>• System code</li> <li>• Pathnames, filenames, prompts, and syntax</li> <li>• Commands and options</li> </ul>
<i>Monospace italic</i>	Use for variables.
<b>Monospace bold</b>	Use for user input.
[ ]	Square brackets enclose optional values
	Vertical bar indicates alternate selections — the bar means “or”
{ }	Braces enclose content that the user must specify, such as x or y or z
...	Ellipses indicate nonessential information omitted from the example



## Where to Get Help

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

**Product information** — For documentation, release notes, software updates, or information about EMC products, go to EMC Online Support at:

<https://support.emc.com>

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

## Your Comments

Your suggestions will help us continue to improve the accuracy, organization, and overall quality of the user publications. Send your opinions of this document to:

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



# Introduction

This guide covers host configuration information for all released XtremIO versions up to the document's publication date (unless specified otherwise within this document).

To optimize performance, hosts accessing the XtremIO Storage Array may require configuring not only the XtremIO cluster, but also the host itself. This document describes the necessary procedures for optimally configuring the host for XtremIO storage. Topics include the following best practices:

- ◆ Host hardware and operating cluster settings
- ◆ FC and iSCSI connectivity and configuration
- ◆ Multipathing requirements and settings
- ◆ File system and application requirements

Chapters in this document provide information on the following topics:

- ◆ [Chapter 1, “Best Practices for Zoning and Subnetting”](#)  
Includes best practices for allowing a host and the XtremIO cluster to communicate.
- ◆ [Chapter 2, “Host Configuration for Linux®”](#)  
Includes instructions for configuring the Linux host platform.
- ◆ [Chapter 3, “Host Configuration for VMware® vSphere®”](#)  
Includes instructions for configuring the VMware vSphere host platform.
- ◆ [Chapter 4, “Host Configuration for Microsoft® Windows®”](#)  
Includes instructions for configuring the Microsoft Windows host platform.
- ◆ [Chapter 5, “Host Configuration for Oracle® Solaris”](#)  
Includes instructions for configuring the Oracle Solaris host platform.
- ◆ [Chapter 6, “Host Configuration for AIX”](#)  
Includes instructions for configuring the AIX host platform.
- ◆ [Chapter 7, “Host Configuration for HP-UX,”](#)  
Includes instructions for configuring the HP-UX host platform.
- ◆ [Appendix A, “Configuring Boot from SAN with XtremIO,”](#)  
Includes instructions for configuring boot from SAN with XtremIO.

---

**Note:**

Linux is a trademark of Linus Torvalds.

Oracle and Solaris are trademarks or registered trademarks of Oracle.

Microsoft, Windows, and the Windows logo are trademarks or registered trademarks of Microsoft Corporation.

VMware and vSphere are trademarks or registered trademarks of VMware.

---

# Pre-Requisites

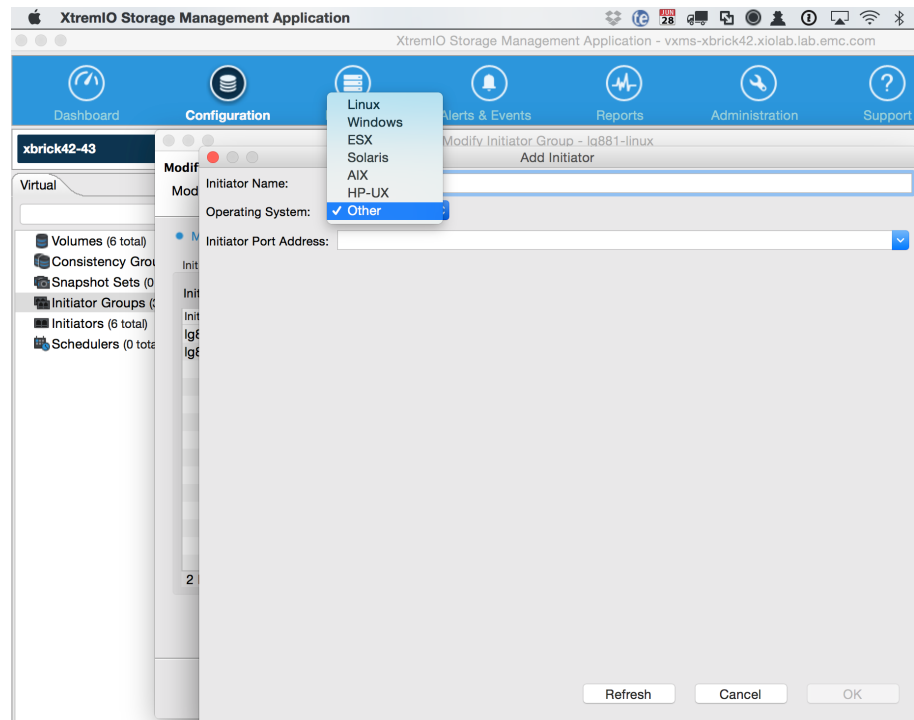
Before configuring the hosts accessing the XtremIO cluster, verify that the following prerequisites are met:

- ◆ The XtremIO cluster is fully installed and operational (refer to the *Release Notes* version, matching your cluster).
- ◆ The host is physically connected to the XtremIO cluster via either FC or iSCSI (refer to [“Best Practices for Zoning and Subnetting” on page 15](#)).
- ◆ The host is installed and configured with a supported operating system and hardware (refer to the *XtremIO Simple Support Matrix* version that matches your cluster).
- ◆ The appropriate driver and firmware are installed on the host, when using HBA for FC or NIC for iSCSI (refer to the *XtremIO Simple Support Matrix* version that matches your cluster).
- ◆ XtremIO volumes are created and presented to the host as required (refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster).
  - In XtremIO version 4.0.0 (or above), volumes are numbered by default starting from LUN ID 1 (and not 0 as was the case in previous XtremIO versions).
  - Although possible, it is not recommended to manually adjust the LUN ID to 0, as it may lead to issues with some operating systems. For further details on these issues, refer to the respective LUN0 sub-sections in the Linux, vSphere and AIX chapters.
  - When a cluster is updated from XtremIO version 3.0.x to 4.0.x, an XtremIO volume with a LUN ID 0 remains accessible following the upgrade.

## **IMPORTANT**

Make sure to use the most recent revision of the XtremIO Host Configuration Guide. If this document is more than a month old, refer to EMC Online Support (<https://support.emc.com>) to check if a more recent document revision is published.

- ◆ With XtremIO version 4.0 (or above), it is required to specify an operating system for each defined initiator, according to the host's corresponding operating system (refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster).



**Note:** It is recommended to avoid using host-level software for deduplication, compression and encryption of data stored in XtremIO. These operations are performed natively on the XtremIO cluster, thus releasing host computing resources. Furthermore, using these operations may severely impact data efficiency on the XtremIO cluster, as they can lead to a higher physical/logical capacity utilization.



# CHAPTER 1

## Best Practices for Zoning and Subnetting

This chapter includes the following topics:

◆ General Guidelines .....	16
◆ Minimal Zoning/Subnetting Configuration.....	16
◆ iSCSI SAN Guidelines .....	17
◆ Fibre Channel SAN Guidelines .....	18
◆ 10TB Starter X-Brick (5TB) and Single X-Brick Cluster.....	18
◆ Dual X-Brick Clusters .....	19
◆ Clusters with Four X-Bricks (or More) .....	20

FC Zones and IP subnetting are ways to control the access and traffic between initiators and targets.

This section explains the best practices for allowing a host and the XtremIO cluster to communicate using 4, 8 or 16 paths per device.

## General Guidelines

- ◆ The optimal number of paths depends on the operating system and server information. To avoid multipathing performance degradation, do not use more than 16 paths per device. It is recommended to use 8 paths.

---

**Note:** This recommendation is not applicable to Linux hosts connected to XtremIO. On such hosts, more than 16 paths per device can be used (if required).

---

- ◆ Balance the hosts between the Storage Controllers to provide a distributed load across all target ports.
- ◆ Host I/O latency can be severely affected by SAN congestion. Minimize the use of ISLs by placing the host and storage ports on the same physical switch. When this is not possible, ensure that there is sufficient ISL bandwidth and that both the Host and XtremIO interfaces are separated by no more than 2 ISL hops. For more information on proper SAN design, refer to the Networked Storage Concepts and Protocols techbook (<https://www.emc.com/collateral/hardware/technical-documentation/h4331-networked-storage-cncpts-prtcls-sol-gde.pdf>).
- ◆ Keep a consistent link speed and duplex across all paths between the host and the XtremIO cluster.
- ◆ To ensure continuous access to XtremIO storage during cluster software upgrade, verify that a minimum I/O timeout of 30 seconds is set on the HBAs of all hosts connected to the affected XtremIO cluster (see EMC KB article 167514 for references to EMC Host Connectivity Guides that include this guideline).

## Minimal Zoning/Subnetting Configuration

To prevent a host path from going down when two Storage Controllers (from separate X-Bricks) are down, while the XtremIO cluster remains active (because the failed Storage Controllers are not in the same X-Brick), follow these guidelines:

- ◆ 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. This zone can be expanded to additional Storage Controllers from other X-Bricks. Moreover, other host ports can be zoned to Storage Controllers from other X-Bricks.

---

**Note:** The diagrams throughout this chapter illustrate possible implementations of these guidelines. Other possible implementations exist, but are not illustrated.

---



## iSCSI SAN Guidelines

When setting up an iSCSI SAN infrastructure, follow these guidelines:

- ◆ If jumbo frames are required for iSCSI traffic, make sure that all ports (server, switch/es and storage) are configured with the correct MTU value.

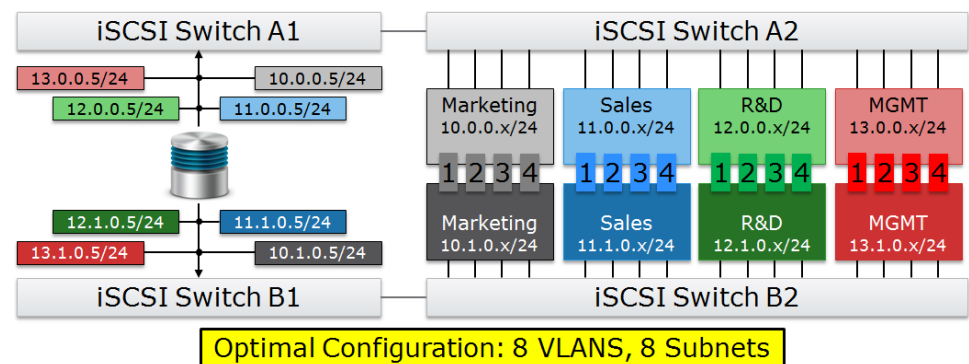
---

**Note:** When using Jumbo Frames with VMware ESX, the correct MTU size must be set on the virtual switch as well.

---

- ◆ Disable flow control features on the server, switch/es and array ports.
- ◆ Make sure that the different iSCSI IP subnets cannot transmit traffic between them.
- ◆ Use Virtual LAN (VLAN) technology to partition a single iSCSI network link into multiple distinct domains. If possible, use a dedicated VLAN for XtremIO iSCSI traffic and a dedicated IP subnet for each iSCSI fabric. Do not configure iSCSI routes between the different subnets.

The following example shows optimal VLAN and IP Subnetting configuration:



- ◆ Enable the TCP Offloading Engine (TOE) on the host iSCSI interfaces, to offload the TCP packet encapsulation from the host's CPU to the NIC or iSCSI HBA, and free up CPU cycles.
- ◆ It is recommended to use a dedicated NIC or iSCSI HBA for XtremIO iSCSI and not to partition the iSCSI interface (i.e. disable NIC Partitioning - NPAR).
- ◆ When using XtremIO iSCSI, it is recommended to use interfaces individually rather than using NIC Teaming (Link Aggregation), to combine multiple interfaces into a single virtual interface.

---

**Note:** Refer to your FC/iSCSI switch's user manual for instructions on the actual implementations.

---

## Fibre Channel SAN Guidelines

When setting up a Fibre Channel SAN infrastructure, follow these guidelines:

- ◆ Use a single-initiator per single-target (1:1) zoning scheme. If the FC switch zone count limitation has been reached, it is also possible to use single-initiator per multiple-target (1:many) zoning scheme.

---

**Note:** Refer to your FC/iSCSI switch's user manual for instructions on the actual implementations.

---

## 10TB Starter X-Brick (5TB) and Single X-Brick Cluster

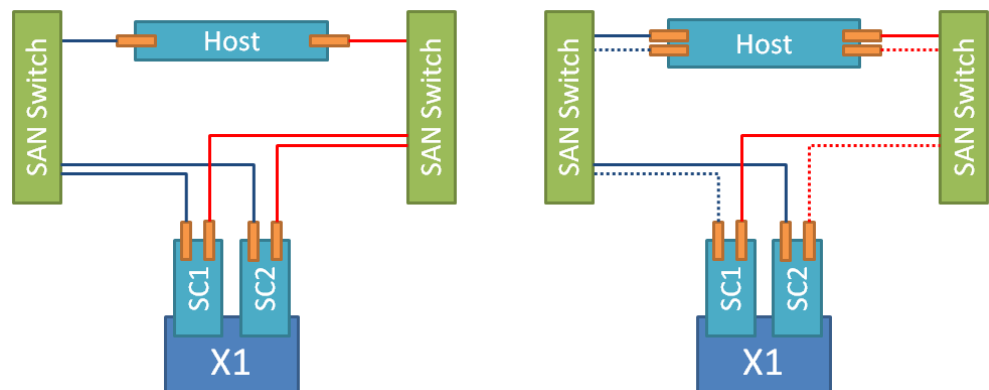
---

**Note:** When partially zoning/subnetting multiple servers to the XtremIO array, make sure that the servers' I/O load is distributed equally across all X-Bricks.

---

In a 10TB starter X-Brick (5TB) or a single X-Brick configuration, a host may have up to 4 paths per device.

The following diagram displays the logical connection topology for 4 paths. This topology applies to both dual and quad HBA/NIC host architecture.



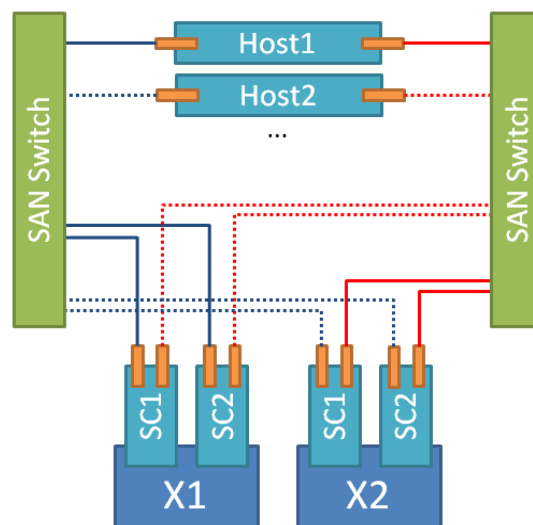
## Dual X-Brick Clusters

**Note:** When partially zoning/subnetting multiple servers to the XtremIO array, make sure that the servers' I/O load is distributed equally across all X-Bricks.

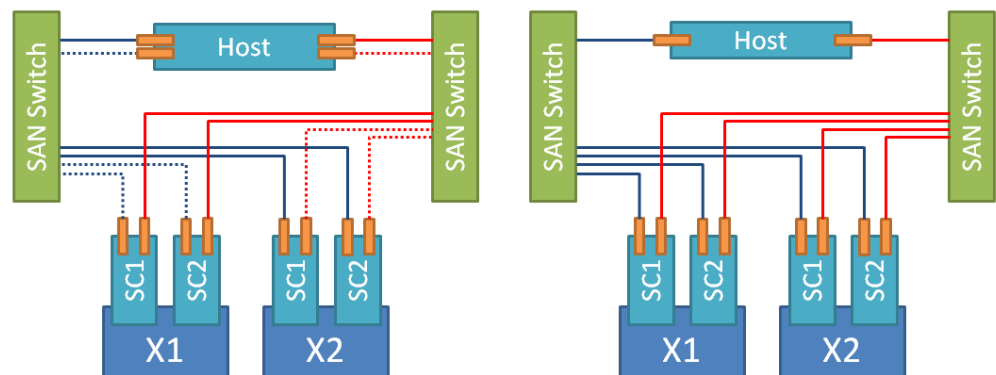
**Note:** For minimal zoning/subnetting configuration guidelines, refer to [“Minimal Zoning/Subnetting Configuration” on page 16](#).

In a dual X-Brick configuration, a host may have up to 8 paths per device.

The following diagram displays the logical connection topology for 4 paths. This topology applies to a dual HBA/NIC host architecture.



The following diagrams display the logical connection topology for 8 paths. This topology applies to both dual and quad HBA/NIC host architecture.



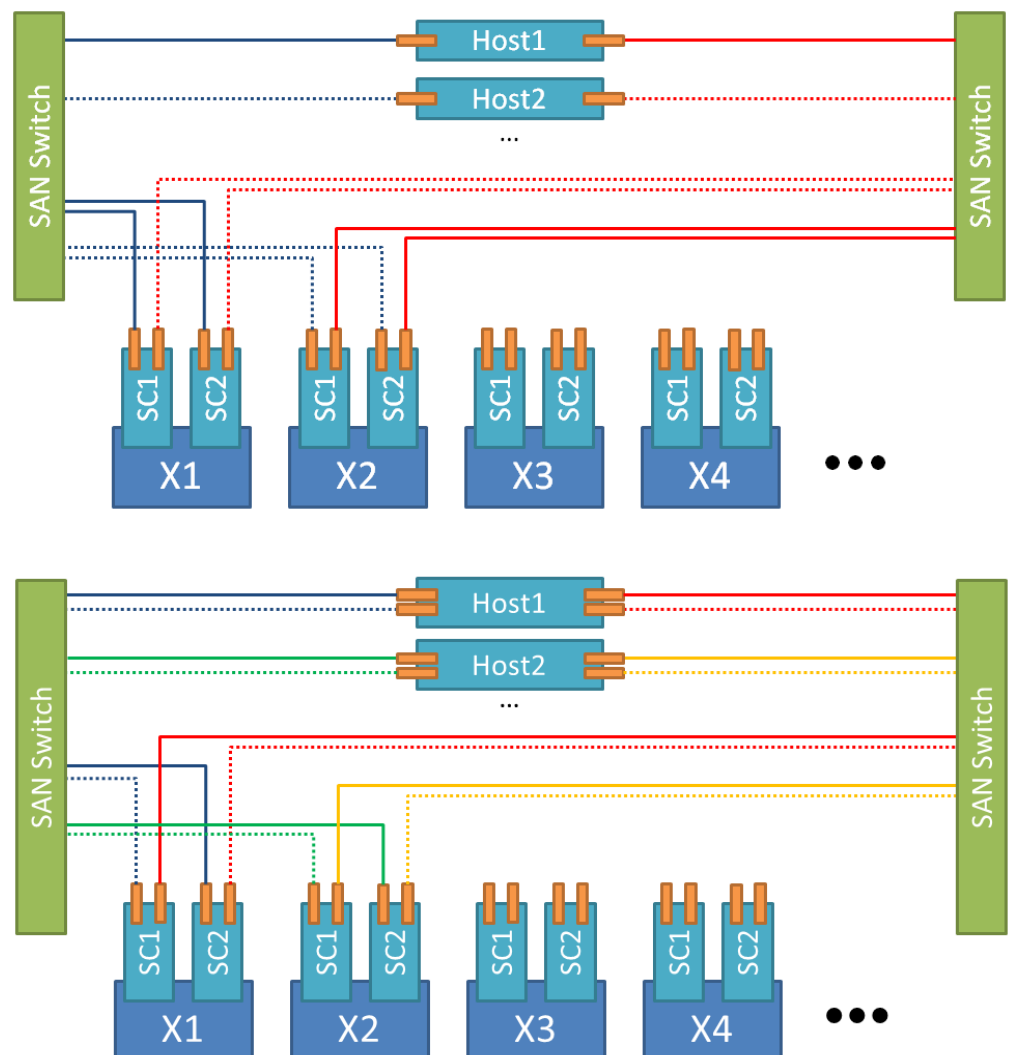
## Clusters with Four X-Bricks (or More)

**Note:** When partially zoning/subnetting multiple servers to the XtremIO array, make sure that the servers' I/O load is distributed equally across all X-Bricks.

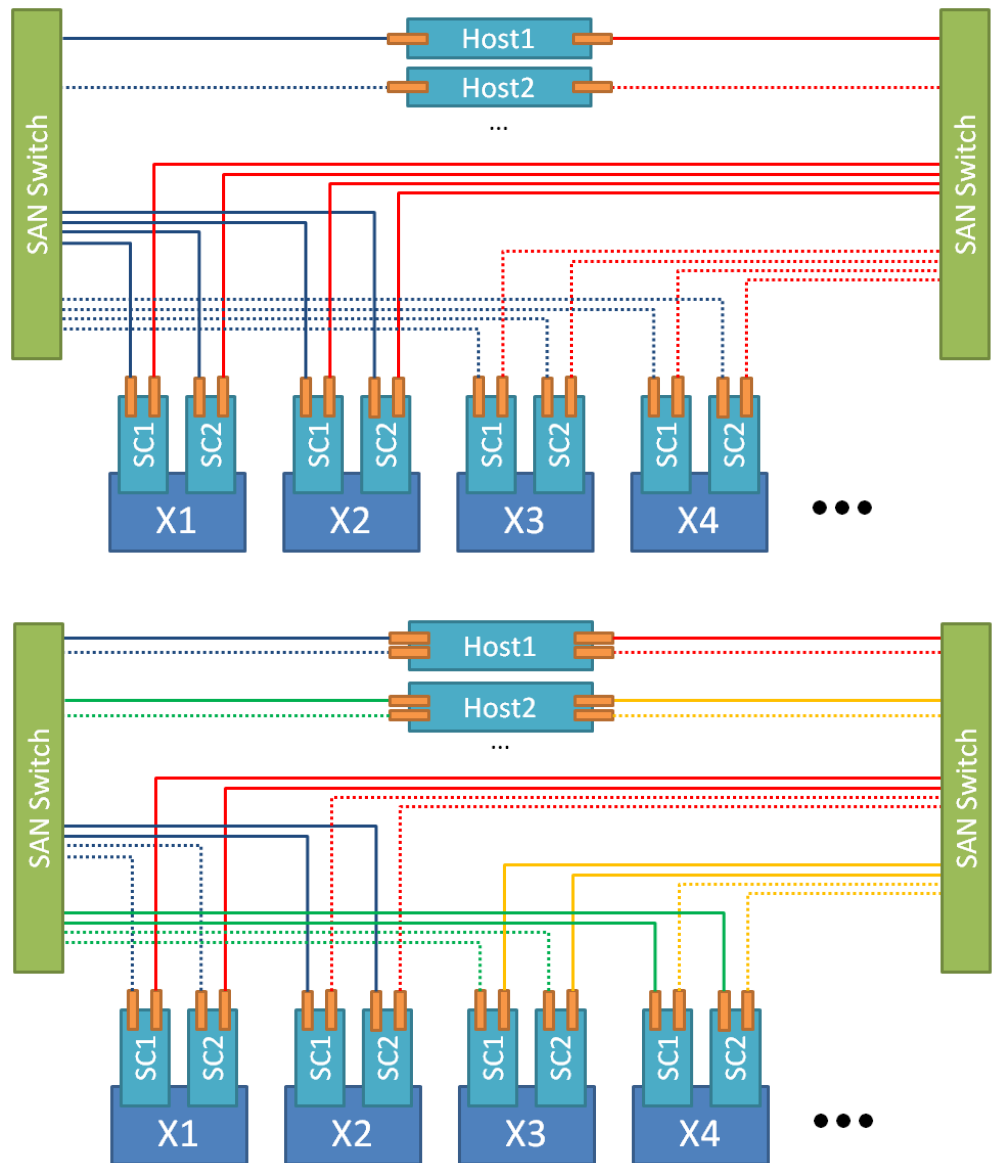
**Note:** For minimal zoning/subnetting configuration guidelines, refer to [“Minimal Zoning/Subnetting Configuration”](#) on page 16.

In four X-Brick configuration (or higher), a host may have up to 16 paths per device.

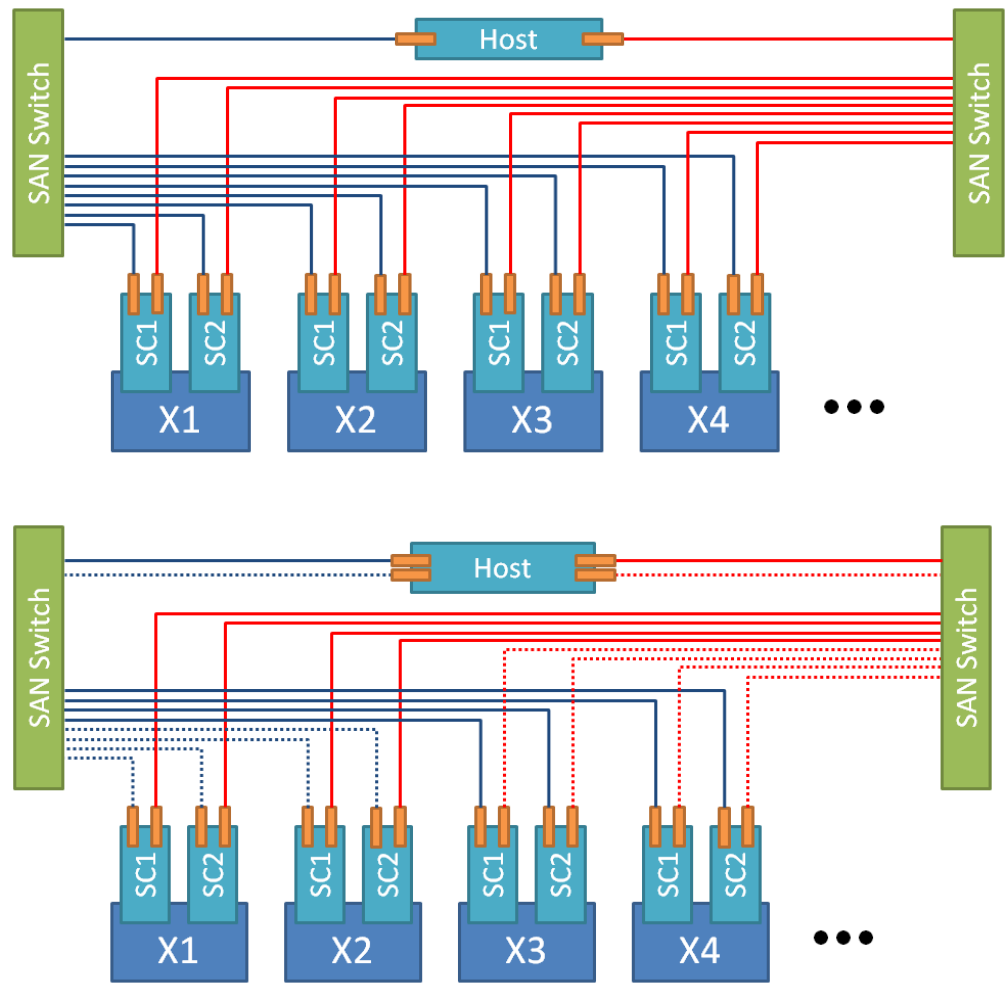
The following diagrams display the logical connection topology for 4 paths. This topology applies to both dual and quad HBA/NIC host architecture.



The following diagrams display the logical connection topology for 8 paths. This topology applies to both dual and quad HBA/NIC host architecture.



The following diagrams display the logical connection topology for 16 paths. This topology applies to both dual and quad HBA/NIC host architecture.



# CHAPTER 2

## Host Configuration for Linux®

This chapter includes the following topics:

- ◆ [Fibre Channel HBA Configuration.....](#) 24
- ◆ [Multipathing Software Configuration .....](#) 33
- ◆ [Post-Configuration Steps - Using the XtremIO Cluster.....](#) 38

---

**Note:** In hosts running a hypervisor, such as VMware ESX, Microsoft Hyper-V, or any clustering software, it is important to ensure that the logical unit numbers of XtremIO volumes are consistent across **all** hosts in the hypervisor cluster. Inconsistent LUNs may affect operations such as VM online migration or VM power-up.

---

## Fibre Channel HBA Configuration

When using Fibre Channel with XtremIO, the FC Host Bus Adapters (HBA) issues, described in this section, should be addressed for optimal performance.

### Pre-Requisites

To install one or more EMC-approved HBAs into a Linux host, follow the procedures in one of these documents according to the FC HBA type:

- ◆ **For Qlogic HBAs** - *EMC Host Connectivity with Qlogic Fibre Channel and iSCSI HBAs and Converged Network Adapters (CNAs) for the Linux Environment document*
- ◆ **For Emulex HBAs** - *EMC Host Connectivity with Emulex Fibre Channel and iSCSI HBAs and Converged Network Adapters (CNAs) for the Linux Environment document*

Refer to these documents to configure the host for connection to the EMC Storage Arrays over a Fibre Channel including any needed HBA BIOS settings.

The documents are available in the EMC OEM section of the QLogic site at <http://www.qlogic.com>, or the Emulex site at <http://www.emulex.com>. They can also be found at <http://support.EMC.com>.

### Queue Depth and I/O Elevators

This section describes queue depth and I/O elevators, and provides the recommended settings.

#### Queue Depth

---

**Note:** Changing the queue depth settings is designed for advanced users. Increasing the queue depth may cause the host to over-stress other arrays connected to the Linux host, resulting in performance degradation while communicating with them. Therefore, especially in mixed environments with multiple array types connected to the Linux host, compare the XtremIO recommendations for queue depth with those of other platforms before applying them.

---

Queue depth is the amount of SCSI commands (including I/O requests) that can be handled by a storage device at a given time. A queue depth can be set on either of the following:

- ◆ Initiator level - HBA queue depth
- ◆ LUN level - LUN queue depth

The HBA queue depth (also referred to as execution throttle) setting controls the amount of outstanding I/O requests per HBA port. The HBA queue depth should be set to the maximum value. This can be done on the HBA firmware level, using the HBA BIOS or CLI utility, provided by the HBA vendor:

- ◆ Qlogic - HBA queue depth - Change the default value (32) to 65535.
- ◆ Emulex - `lpfc_hba_queue_depth` - No need to change the default (and maximum) value (8192).



**Note:** HBA queue depth (execution throttle) does not apply to QLE2600 and QLE8300 Series Qlogic adapters, and is read only for 10GbE adapters.

For further information, refer to:

[http://qlogic.force.com/SupportCenter/articles/FAQ/What-are-the-HBA-Execution-Throttling-settings-best-practices?retURL=%2FSupportCenter%2Fapex%2FCustomer\\_Support\\_main&popup=true](http://qlogic.force.com/SupportCenter/articles/FAQ/What-are-the-HBA-Execution-Throttling-settings-best-practices?retURL=%2FSupportCenter%2Fapex%2FCustomer_Support_main&popup=true).

The LUN queue depth setting controls the amount of outstanding I/O requests per a single path. These settings are controlled in the driver module for the card at the OS level.

For optimal operation with XtremIO, it is recommended to adjust the LUN queue depth settings for Qlogic and Emulex HBAs, using Linux udev rules (see “[Setting the LUN Queue Depth, Using Linux Udev](#)” on page 26).

Alternatively, it is possible to adjust the LUN queue depth settings for Qlogic and Emulex HBAs, using the CLI. See “[Setting the LUN Queue Depth with Qlogic HBA, Using CLI](#)” on page 28 and “[Setting the LUN Queue Depth with Emulex HBA, Using CLI](#)” on page 30, respectively. Follow the instructions in either of these sections according to the HBA type.

**Note:** If the HBA queue depth is set to a value lower than the LUN queue depth, it may limit the LUN queue depth to a lower value than set.

The following table summarizes the default and recommended queue depth settings for Linux.

	LUN Queue Depth	HBA Queue Depth
Default Value	Qlogic: 32 Emulex: 30	Qlogic: 32 Emulex: 8192
Recommended Value	Qlogic: 256 Emulex: 128	Qlogic: 65536 (maximum) Emulex: 8192 (maximum)

## I/O Elevators

I/O scheduling controls how I/O operations are submitted to storage. Linux offers various I/O algorithms (also known as “Elevators”) to accommodate for different workloads. Elevators mainly help in reducing seek operations.

Linux sets a default I/O scheduler at boot-time that can be changed on the fly per block device.

When connecting a Linux host to XtremIO storage, set the I/O elevator to either `noop` or `deadline`. It is not recommended to use the `cfq` I/O elevator setting, as it is less optimal for XtremIO storage.

For optimal operation with XtremIO, it is recommended to set the I/O elevators, using Linux udev rules (see “[Setting the I/O Elevators, Using Linux Udev](#)” on page 27).

Alternatively, it is possible to set the I/O elevators via CLI (see “[Setting the I/O Elevators using CLI](#)” on page 31).

## Setting the LUN Queue Depth and I/O Elevators, Using Linux Udev

### Setting the LUN Queue Depth, Using Linux Udev

---

**Note:** The procedures for setting the LUN queue depth with Qlogic HBA and with Emulex HBA are identical, except for the set queue depth value.

---

#### To set the LUN queue depth, using Linux udev:

1. Create or edit the following file:

```
/etc/udev/rules.d/99-XtremIO.rules
```

2. Append the following rule to the file:

```
#increase queue depth on the volume

ACTION=="add|change", SUBSYSTEM=="scsi",
ATTR{vendor}=="XtremIO", ATTR{model}=="XtremApp      ",
ATTR{queue_depth}=="<selected LUN queue depth value>"
```

---

**Note:** There should be eight (8) spaces between 'XtremApp' and the closing quotation mark.

---



---

**Note:** Be sure to use the appropriate queue depth value according to the HBA vendor (Qlogic or Emulex). To determine the LUN queue depth value, refer to the corresponding guidance (see [“Queue Depth” on page 24](#)).

---

3. Save the changes made to the file.
4. To apply the changes, run the following command:

```
udevadm trigger
```

---

**Note:** Older kernels may use the `udevtrigger` command instead.

---

## Setting the I/O Elevators, Using Linux Udev

There are two options for disabling any I/O elevator:

- ◆ For Linux running with any multipath program other than Linux native multipath:

To disable any I/O elevator, follow one of these procedures according to the used GRUB version:

- GRUB (versions prior to GRUB2):
  - Modify the `/boot/grub/menu.lst` file and append `"elevator=noop"` as an option at the end of the kernel line.
- GRUB2:
  1. Modify the `/etc/default/grub` file and append `"elevator=noop"` to the list of `GRUB_CMDLINE_LINUX_DEFAULT` values.
  2. Rebuild GRUB2 configuration files, using the following shell command:
 

```
sudo update-grub
```

- ◆ For Linux running with native multipath:

1. Create or edit the following file:

```
/etc/udev/rules.d/99-XtremIO.rules
```

2. Copy and paste the following rules to the file:

```
# Use noop scheduler

ACTION=="add|change", SUBSYSTEM=="block",
ENV{ID_VENDOR}=="XtremIO", ENV{ID_MODEL}=="XtremApp",
ATTR{queue/scheduler}="noop"

ACTION=="add|change", SUBSYSTEM=="block", KERNEL=="dm*",
ENV{DM_NAME}=="??14f0c5*", ATTR{queue/Scheduler}="noop"
```

---

**Note:** These rules are general recommendations. Some Linux operating systems may benefit from using `deadline` elevator configuration.

---

3. Save the changes made to the file.
4. To apply the changes, run the following command:

```
udevadm trigger
```

---

**Note:** Older kernels may use the `udevtrigger` command instead.

---



---

**Note:** Following this procedure, if new devices are configured, udev discovers them and applies the rules automatically.

---

## Setting the LUN Queue Depth and I/O Elevators, Using CLI

### Setting the LUN Queue Depth with Qlogic HBA, Using CLI

**Note:** This procedure applies to other Linux environments, which are supported by XtremIO, as well. For details, refer to the XtremIO ESSM.

For Qlogic cards controlled by the qla2xxx module, set the **ql2xmaxqdepth** parameter (default value 32) to the maximum value (256).

#### To check the Qlogic LUN queue depth:

- ◆ For SUSE and RedHat versions below 6.6:

Run the following command to display the current Qlogic LUN queue depth:

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

Example:

```
# cat /sys/module/qla2xxx/parameters/ql2xmaxqdepth
32
```

If the value displayed is per the required queue depth, no further modifications are required for the Qlogic LUN queue depth.

- ◆ For SUSE and RedHat version 6.6 (or above):

1. Find the relevant module.

```
#modinfo qla2xxx | grep maxqdepth
parm:      ql2xmaxqdepth:Maximum queue depth to set for
each LUN. Default is 32. (int)
```

2. Create or update the qlx file in modprobe.d with this entry.

```
#vi /etc/modprobe.d/qlx.conf
options qla2xxx ql2xmaxqdepth=256
```

**To modify the LUN queue depth:**

1. Select one of the following options according to the version:

- For SuSE and Red Hat version 5.x:

Edit the file `/etc/modprobe.conf` file, and add or modify the following parameter:

```
options qla2xxx ql2xmaxqdepth=<new_queue_depth>
```

- For later SuSE and Red Hat releases, create or edit the `/etc/modprobe.d/qla2xxx.conf` file, and add or modify the following parameter:

```
options qla2xxx ql2xmaxqdepth=<new_queue_depth>
```

2. Run the following command to rebuild the RAMdisk:

```
mkinitrd -v --with=qla2xxx -f  
/boot/initramfs-w-qla2xxx-`uname -r`.img `uname -r`
```

---

**Note:** After copying and pasting this command, make sure that the back quote sign appears as ```. Otherwise, edit it manually before running this command.

---

3. Add a new entry to the GRUB file with the RAMdisk that was rebuilt in step 2.

---

**Note:** To find the name of the new RAMdisk, run the following command:

```
ls /boot/ | grep qla
```

---

**Note:** For details on the procedure for editing the GRUB file, refer to your Linux documentation.

---

4. After the parameter is set, reboot the host.
5. Run the following command to verify that the parameter is set:

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

## Setting the LUN Queue Depth with Emulex HBA, Using CLI

**Note:** This procedure applies to the other Linux environments, which are supported by XtremIO, as well. For details, refer to the XtremIO ESSM.

For Emulex cards there are two parameters:

- ◆ `lpfc_lun_queue_depth` (default value 30)
- ◆ `lpfc_hba_queue_depth` (default value 8192)

XtremIO recommends setting the LUN queue parameter to its maximum value (128).

### To check the Emulex LUN queue depth:

- ◆ Run the following command to display the current Emulex LUN queue depth:

```
cat /sys/class/scsi_host*/lpfc_lun_queue_depth
```

Example:

```
# cat /sys/class/scsi_host/host*/lpfc_lun_queue_depth
30
```

If the value displayed is per the required queue depth, no further modifications are required for the Emulex LUN queue depth.

### To modify the LUN queue depth value:

1. Select one of the following options according to the version:

- For SuSE and Red Hat 5.x and earlier:

Edit the `/etc/modprobe.conf` file, and add or modify the following parameter:

```
options lpfc lpfc_lun_queue_depth=<new_queue_depth>
```

- For later SuSE and Red Hat releases, create or edit the `/etc/modprobe.d/lpfc.conf` file, and add or modify the following parameter:

```
options lpfc lpfc_lun_queue_depth=<new_queue_depth>
```

2. Run the following command to rebuild the RAMdisk:

```
mkinitrd -v --with=lpfc820 -f
/boot/initramfs-w-lpfc820-`uname -r`.img `uname -r`
```

**Note:** After copying and pasting this command, make sure that the back quote sign appears as ```. Otherwise, edit it manually before running this command.

3. Add a new entry to the GRUB file with the RAMdisk that was rebuilt in step 2.

**Note:** To find the name of the new RAMdisk, run the following command:

```
ls /boot/ | grep lpfc
```

**Note:** For details on the procedure for editing the GRUB file, refer to your Linux documentation.

4. After the parameter is set, reboot the host.
5. Run the following command to verify that the parameter is set:

```
cat /sys/class/scsi_host/host*/lpfc_lun_queue_depth
```

## Setting the I/O Elevators using CLI

Follow one of these procedures according to the GRUB version and RHEL version:

### ◆ GRUB (versions prior to GRUB2):

1. Check if I/O elevator is already disabled by searching for the ``elevator=noop`` or ``elevator=deadline`` string in the `/etc/grub.conf` file. If it is disabled, skip the rest of this procedure.
2. Modify the `/boot/grub/menu.lst` file, and append "elevator=noop" or "elevator=deadline" as an option to the end of the kernel line.

Example:

```
default=0
timeout=5
title CentOS (2.6.32-358.el6.x86_64)
root (hd0,0)
kernel /vmlinuz-2.6.32-358.el6.x86_64 ro ...
elevator=noop
initrd /initramfs-2.6.32-358.el6.x86_64.img
```

### ◆ GRUB2 (for RHEL versions prior to 7):

1. Check if I/O elevator is already disabled by searching for the ``elevator=noop`` or ``elevator=deadline`` string in the `/etc/default/grub` file. If it is disabled, skip the rest of this procedure.
2. Modify the `/etc/default/grub` file, and append/write "elevator=noop" or "elevator=deadline" to the list of `GRUB_CMDLINE_LINUX_DEFAULT` values.

Example:

```
GRUB_CMDLINE_LINUX_DEFAULT=" ... elevator=noop"
```

3. Run the following shell command to rebuild the GRUB2 configuration files:

```
sudo update-grub
```

◆ GRUB2 (for RHEL versions 7 and above):

1. Check if I/O elevator is already disabled by searching for the ``elevator=noop`` or ``elevator=deadline`` string in the `/etc/default/grub` file. If it is disabled, skip the rest of this procedure.
2. Modify the `/etc/default/grub` file, and append/write `"elevator=noop"` or `"elevator=deadline"` to the list of `GRUB_CMDLINE_LINUX_DEFAULT` values.

Example:

```
GRUB_CMDLINE_LINUX_DEFAULT="elevator=noop"
```

3. Run the following shell command to rebuild the GRUB2 configuration files:

```
grub2-mkconfig -o /boot/grub2/grub.cfg
```

4. Run the following command to verify that the grub file is rebuilt:

```
grep elevator /etc/grub2.cfg
```

The output should include lines containing `"elevator=noop"` or `"elevator=deadline"`.

Example:

```
#grep elevator /etc/grub2.cfg

linux16 /vmlinuz-3.10.0-229.el7.x86_64
root=/dev/mapper/centos-root ro
rd.lvm.lv=centos/swap crashkernel=auto
ipv6.disable=1 rd.lvm.lv=centos/root
net.ifnames=0 rhgb quiet elevator=noop
linux16 /vmlinuz-0-rescue-f6c987899a1d456ea1fa9b97fe3
5bcf7 root=/dev/mapper/centos-root ro
rd.lvm.lv=centos/swap crashkernel=auto
ipv6.disable=1 rd.lvm.lv=centos/root
net.ifnames=0 rhgb quiet elevator=noop
```



# Multipathing Software Configuration

## Configuring Linux Native Multipathing

Even when the default multipath configuration appears to be functioning and provides path failover, key values must be set to enable the cluster to survive controller failover. It is recommended to configure a volume as a multipath device whenever possible. Even in situations in which a single path is initially used, configuring the volume as a multipath device is advantageous.

The Linux native multipathing requires installing the device-mapper-multipath version 0.4.9-72 (or later) RPM.

The configuration file for the multipath daemon is `multipath.conf`. It is used to overwrite the built-in configuration table of the multipath daemon. When iSCSI is used with XtremIO, the `iscsi.conf` file is used to overwrite iSCSI specific multipathing related settings.

To configure the XtremIO disk device, modify the `/etc/multipath.conf` file with the following parameters:

```
devices {
    device {
        vendor                XtremIO
        product               XtremApp
        path_selector          "queue-length 0" (FOR RHEL>=6)
        path_selector          "round-robin 0" (FOR RHEL<6)
        rr_min_io              1 (for kernels older than 2.6.31)
        rr_min_io_rq           1 (for kernels 2.6.31 and above)
        path_grouping_policy   multibus
        path_checker            tur
        failback               immediate
        fast_io_fail_tmo       15
    }
}
```

Parameter	Value	Description
<code>path_selector</code>	<code>round-robin 0</code>	Loops through every path in the path group, sending the same amount of I/O to each.
<code>path_selector</code>	<code>queue-length 0</code>	Sends the next bunch of I/O down the path with the least number of outstanding I/O requests.
<code>rr_min_io</code>	1	Specifies the number of I/O requests to route to a path before switching to the next path in the current path group (for kernels older than 2.6.31).
<code>rr_min_io_rq</code>	1	Specifies the number of I/O requests to route to a path before switching the next path in the current path group (for kernels 2.6.31 and above).
<code>path_grouping_policy</code>	<code>multibus</code>	All paths are in a single group (all paths have the same priority).
<code>path_checker</code>	<code>tur</code>	Specifies TEST UNIT READY as the default method used to determine the state of the paths.
<code>failback</code>	<code>immediate</code>	Manages the path group failback. <code>immediate</code> refers to immediate failback to the highest priority path group that contains active paths.
<code>fast_io_fail_tmo</code>	15	Specifies the number of seconds between detection of a problem on an FC remote port and failing I/O to devices on the remote port by the SCSI layer.

To configure the XtremIO disk device, modify the `/etc/iscsi/iscsi.conf` file, using the following parameters:

Parameter	Value	Description
<code>node.session.timeo.replacement_timeout</code>	5	Specifies the number of seconds the iSCSI layer waits for a timed-out path/session to re-establish before failing any commands on that path/session. The default value is 120.
<code>FirstBurstLength/MaxBurstLength</code>	<at least block size used> (for both parameters)	<p><b>FirstBurstLength:</b> Specifies the maximum amount of unsolicited data (in bytes) which an iSCSI initiator can send to the target during the execution of a single SCSI command.</p> <p><b>MaxBurstLength:</b> Specifies the maximum SCSI data payload per a single SCSI command.</p> <p>Adjust these parameters to the same value when the used block size is larger than the default setting for this parameter (256KB).</p>

---

**Note:** With Veritas Storage Foundation, keep the `replacement_timeout` parameter default value of 120.

---



---

**Note:** With Red Hat Enterprise Linux 4 series, instead of `node.session.timeo.replacement_timeout`, set the value of `ConnFailTimeout` to 5.

---



---

**Note:** With Citrix XenServer 6 series, for existing Storage Repositories (SR), after editing the `iscsi.conf` file per the above table, detach and then reattach SRs from XenCenter or by using XenServer xe CLI. This will reflect the new iSCSI timeout for the existing SRs.

---



---

**Note:** With Red Hat Enterprise Virtualization Hypervisor 6.2 and above, after editing the `iscsi.conf` file per the above table, execute the following command to ensure that your `iscsid.conf` file is persistent across reboots:

```
persist /etc/iscsi/iscsid.conf.
```

---

Using these settings prevents commands from being split by the iSCSI initiator and enables instantaneous mapping from the host to the volume.

To apply the adjusted `iscsi.conf` settings, run the following command on the Linux host:

```
service iscsi restart
```

## Configuring PowerPath Multipathing

**Note:** For the most updated information on PowerPath support with XtremIO storage, refer to the *XtremIO Simple Support Matrix*.

XtremIO supports multipathing using EMC PowerPath on Linux. PowerPath versions 5.7 SP3 and above provide Native Loadable Array Module (LAM) for XtremIO Array devices. With this support, XtremIO devices running versions 2.2 SP1 and above are managed under the XtremIO class.

PowerPath provides enhanced path management capabilities for up to 32 paths per logical device, as well as intelligent dynamic I/O load-balancing functionalities. Having multiple paths enables the host to access a storage device even if a specific path is unavailable. Multiple paths share the I/O traffic to a storage device, using intelligent load-balancing policies that enhance I/O performance and increase application availability. EMC PowerPath is the recommended multipathing choice.

PowerPath features include:

- ◆ Multiple paths - enables higher availability and I/O performance.
- ◆ Path management insight capabilities - PowerPath characterizes I/O patterns and aides in diagnosing I/O problems due to flaky paths or unexpected latency values. Metrics are provided on:
  - Read and write - in MB/seconds per LUN
  - Latency distribution - the high and low watermarks per path
  - Retries - the number of failed I/Os on a specific path
- ◆ Autostandby - automatically detects intermittent I/O failures and places paths into autostandby (also known as flaky paths).
- ◆ PowerPath Migration Enabler - a host-based migration tool that allows migrating data between storage systems. PowerPath Migration Enabler works in conjunction with the host operating system (also referred to as Host Copy) and other underlying technologies, such as Open Replicator (OR). Starting with PowerPath for Linux version 6.1, PowerPath Migration Enabler is supported from VMAX2 to XtremIO storage.
- ◆ Remote monitoring:
  - SNMP management daemon
  - PowerPath Virtual Appliance 2.0 (vApp 2.0)
- ◆ Automated kernel upgrade support in RHEL and Oracle Linux kernels (available in PowerPath 6.0 and later).

Further PowerPath related information:

- ◆ For details on the PowerPath releases supported for your Linux host, refer to the *XtremIO Simple Support Matrix*.
- ◆ For details on native class support with XtremIO for your host, refer to the *EMC PowerPath release notes* document for the PowerPath version you are installing
- ◆ For details on installing and configuring PowerPath with XtremIO native class support on your host, refer to the *EMC PowerPath on Linux Installation and Administration Guide* for the PowerPath version you are installing. This guide provides the required information for placing XtremIO volumes under PowerPath control.

## Post-Configuration Steps - Using the XtremIO Cluster

After the host configuration is completed, you can access the XtremIO cluster from the host. For details on creating, presenting and managing volumes accessed from the host via either the GUI or the CLI, refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

### Disk Formatting

Generally, a 512B logical block size should be used for a new XtremIO volume.

Name	Size	Logical Block Size	Small IO Al...	Unaligned ...	VAAI TP Alerts
VOL 746	8GB	Normal (512 LBs)	Disabled	Disabled	Disabled

1 volumes (8GB)

Current Physical Storage  
7.587TB free

< Back   Next >   Finish   Cancel

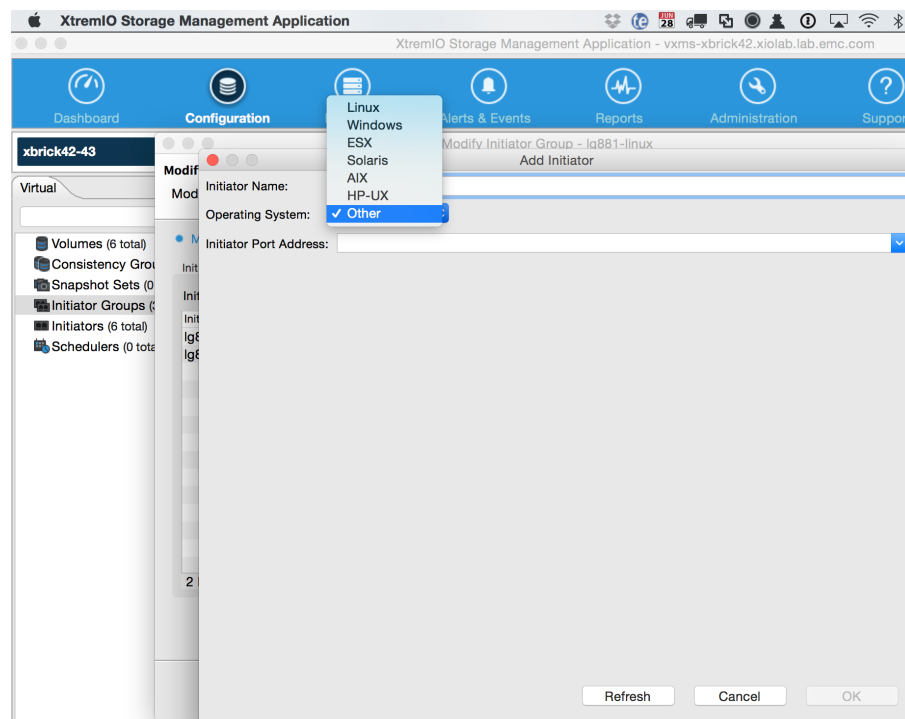
**Note:** In XtremIO version 4.0.0 (and above), the Legacy Windows option is not supported.

## Presenting XtremIO Volumes to the Linux Host

**Note:** It is recommended to avoid using host-level software for deduplication, compression and encryption of data stored in XtremIO. These operations are performed natively on the XtremIO cluster, thus releasing host computing resources. Furthermore, using these operations may severely impact data efficiency on the XtremIO cluster as they can lead to a higher physical or logical capacity utilization.

**Note:** The following information in this section applies only to XtremIO version 4.0 and above.

When adding Initiator Groups and Initiators to allow Linux hosts to access XtremIO volumes, specify **Linux** as the operating system for the newly-created Initiators, as shown in the figure below.



**Note:** Setting the Initiator's Operating System is required for optimal interoperability and stability of the host with XtremIO storage. You can adjust the setting while the host is online and connected to the XtremIO cluster with no I/O impact.

**Note:** Refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster for details.

Following a cluster upgrade from XtremIO version 3.0.x to version 4.0 (or above), make sure to modify the operating system for each initiator that is connected to a Linux host.

## Partition Alignment in Linux

To align partitions on XtremIO volumes 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.

The following example demonstrates using the `fdisk` command to create an aligned partition on an XtremIO volume.

```
#fdisk -cu /dev/sdX
Partition number (1-4): 1
First sector (2048-524287999, default 2048):
Using default value 2048
Last sector, +sectors or +size{K,M,G} (2048-524287999, default
524287999):
Using default value 524287999

Command (m for help): p

Disk /dev/sdd: 268.4 GB, 268435456000 bytes
255 heads, 63 sectors/track, 32635 cylinders, total 524288000 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal):8192 bytes /65536 bytes
Diskidentifier: 0x5fbcbcf4

Device Boot      Start         End      Blocks   Id  System
/dev/sdd1        2048       524287999    262142976   83   Linux
```

In this mode, rather than using “cylinders” for creating partitions, the `fdisk` command uses sectors, which are a direct mapping to the array's LBA space. Thus, verifying that the partition is aligned is simply a case of verifying that the starting sector number is a multiple of 16 (16 sectors, at 512 bytes each, is 8KB). The `fdisk` command will default to a starting sector for the first partition of 2048, which is divisible by 16, and thus is correctly aligned.

## Logical Volume Manager (LVM)

---

**Note:** Logical Volume Manager (LVM) configuration and management are out of the scope of this document.

---

When configuring LVM with a Linux host connected to XtremIO storage, follow these recommendations:

- ◆ LVM stripe size - The LVM stripe size should be set to a value greater than that of the application block size. XtremIO storage supports physical extent size of 8KB and above. the recommended physical extent size is 1MB or above.



## Creating a File System

---

**Note:** File system configuration and management are out of the scope of this document.

---

It is recommended to create the file system using its default block size (using a non-default block size may lead to unexpected behavior). Refer to your operating system and file system documentation.

For a more efficient data utilization and better performance, use Ext4 file system with XtremIO storage instead of Ext3. For details on converting to Ext4 file system (from either Ext3 or Ext2), refer to the following link:

<https://ext4.wiki.kernel.org/index.php/UpgradeToExt4>

## Using LUN 0 with XtremIO Storage

This section details considerations and steps that should be performed when using LUN 0 with Linux.

Notes on the use of LUN numbering:

- ◆ In XtremIO version 4.0.0 (or above), volumes are numbered by default starting from LUN id 1 (and not 0 as was the case in previous XtremIO versions).
- ◆ Although possible, it is not recommended to manually adjust the LUN id to 0, as it may lead to issues with some operating systems.

---

**Note:** In case LUN 0 is required for an XtremIO volume (e.g. boot from SAN, PowerPath), restart the Linux host if a rescan fails to locate this volume.

---

- ◆ When a cluster is updated from XtremIO version 3.0.x to 4.0.x, an XtremIO volume with a LUN id 0 remains accessible following the upgrade.
- ◆ With XtremIO version 4.0.0 (or above), no further action is required if volumes are numbered starting from LUN id 1.

The following output to the `ls SCSI` command indicates that the device at 1:0:0:0 is the XtremIO cluster controller:

```
[1:0:0:0]      storage XtremIO  XtremApp          3000  -
```

In that case, an XtremIO volume with LUN 0 will be inaccessible to the host.

**To access a volume with a LUN 0 on a Linux host:**

1. Perform one of the following commands to remove the controller device:

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

---

**Note:** `-r` enables the device removal.

---

- `# echo 1 > /sys/class/scsi_device/1:0:0:0/device/delete`

2. Run the following command:

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

---

**Note:** In older Linux versions a host reboot is required instead of a rescan.

---

## Space Reclamation

This section provides a comprehensive list of capacity management steps for achieving optimal capacity utilization on the XtremIO array, when connected to a Linux host.

Data space reclamation helps to achieve optimal XtremIO capacity utilization. Space reclamation is a Linux operating system function, enabling to reclaim used space by sending zeros to a specific address of the volume after being notified by the file system that the address space was deleted. While some file systems can perform this action automatically, others require a user-initiated operation.

The following sections present steps for performing space reclamation with:

- ◆ EXT3 file system
- ◆ EXT4 file system
- ◆ XFS file system

A suggested script for initiating space reclamation is also provided.

### EXT3 File System

**Manual Space Reclamation - EXT3** space reclamation can be performed, using the `fstrim` utility. Using `fstrim` to reclaim unused space does not create a balloon file. Therefore, there is no need to review the free space.

The following example shows running `fstrim` on `/mnt/ext3` mount point.

```
# fstrim /mnt/ext3
```

In some cases, `fstrim` may fail due to T10 commands blocking (VPLEX / VMDK).

In such scenarios it is required to apply a manual copy of zeroes to the relevant free space.

The following example shows running a manual script on `/mnt/ext3` mount point (refer to [“Linux Space Reclamation Script” on page 45](#)).

```
# ./reclaim_space.sh /mnt/ext3
```

### EXT4 File System

- ◆ **Automatic space reclamation - EXT4** supports automatic space reclamation, using the `discard` option. This option should be enabled when mounting the file system (either manually or using `fstab`).

The following example shows `discard` usage when mounting a file system.

```
# mount -o discard /dev/vg_ext4_test/lv_ext4_test /mnt/ext4
```

The following example shows `fstab` usage.

```
/dev/vg_ext4_test/lv_ext4_test /mnt/ext4 ext4 defaults,discard 0 1
```

If automatic space reclamation is not enabled immediately after file system creation, it is recommended to run a manual initial space reclamation prior to automatic mounting.

- ◆ Manual space reclamation - EXT4 space reclamation can be performed, using the `fstrim` utility.

Using the `fstrim` utility does not create a balloon file. Therefore, it is not necessary to review the free space.

The following example shows running `fstrim` on `/mnt/ext4` mount point.

```
# fstrim /mnt/ext4
```

The `fstrim` utility may fail due to T10 commands blocking (VPLEX / VMDK). In such cases, it is required to apply a manual copy of zeroes to the relevant free space.

The following example shows running a manual script on `/mnt/ext4` mount point (refer to [“Linux Space Reclamation Script” on page 45](#)).

```
# ./reclaim_space.sh /mnt/ext4
```

## XFS File System

- ◆ Automatic space reclamation - XFS supports automatic space reclamation, using the `discard` option. This option should be enabled when mounting the file system (either manually or using `fstab`).

The following example shows `discard` usage when mounting a file system.

```
# mount -o discard /dev/vg_xfs_test/lv_xfs_test /mnt/xfs
```

The following example shows `fstab` usage.

```
/dev/vg_xfs_test/lv_xfs_test /mnt/xfs xfs defaults,discard 0 1
```

When automatic space reclamation is not enabled immediately after the file system creation, it is recommended to run a manual initial space reclamation prior to automatic mounting.

- ◆ Manual space reclamation - XFS space reclamation can be performed, using the `fstrim` utility.

Using the `fstrim` utility does not create a balloon file. Therefore, it is not necessary to review the free space.

The following example shows running `fstrim` on `/mnt/xfs` mount point.

```
# fstrim /mnt/xfs
```

The `fstrim` utility may fail due to T10 commands blocking (VPLEX / VMDK). In such cases, it is required to apply a manual copy of zeroes to the relevant free space.

The following example shows running a manual script on `/mnt/xfs` mount point (refer to [“Linux Space Reclamation Script” on page 45](#)).

```
# ./reclaim_space.sh /mnt/xfs
```

## Linux Space Reclamation Script

The following example shows the space reclamation script usage, followed by the space reclamation script for the Linux shell.

```
# python linux-reclaim.py --help
usage: linux-reclaim.py [-c <cluster_name>]

optional arguments:
  -h, --help            show this help message and exit
  -v, --version          show program's version number and exit
  -t, --thread_number    Threads Amount
  -m, --mount_point      Mount Point Name

# python linux-reclaim.py --thread_number 4 --mount_point
/mnt/reclaim1

#!/usr/bin/env python
from __future__ import print_function
import time
import threading
import subprocess
import argparse
import re

class myThread (threading.Thread):
    def __init__(self, name, count, running):
        threading.Thread.__init__(self)
        self.name = name
        self.count = count
        self.running = running

    def run(self):
        dd_something(self.name, self.count, self.running)

def dd_something(name, count, running):
    execute('/bin/dd count={1} bs=131072 if=/dev/zero
of={0}/xtremio_file-{2}.zf oflag=direct'.format(name, count,
running))

def execute(command):
    return subprocess.Popen(command, shell=True,
stdout=subprocess.PIPE,
stderr=subprocess.PIPE).communicate()[0].splitlines()

def get_mount_points():
    return execute('df -m')

threadLock = threading.Lock()
```

```

def main():
    parser = argparse.ArgumentParser(usage='% (prog) s [-c
<cluster_name>]')
    parser.add_argument('-v', '--version', action='version',
version='% (prog) s 1.02')
    parser.add_argument('-t', '--thread_number', action='store',
dest='thread_number', default='1', required=False, help='Threads
Amount', metavar='')
    parser.add_argument('-m', '--mount_point', action='store',
dest='mount_point', default=None, required=True, help='Mount Point
Name', metavar='')
    args = parser.parse_args()
    thread_number = int(args.thread_number)
    mount_point = args.mount_point
    print('Starting Main Thread {0}'.format(time.ctime(time.time())))
    threads = []

    for entry in get_mount_points():
        if mount_point in entry:
            filesystem, mblocks, used, available, usep, mounted =
map(str.strip, re.split(" +", entry))

            for i in xrange(thread_number):
                i = myThread(mounted, int(int(available) * 0.95 /
thread_number) * 8, str(i))
                i.start()
                threads.append(i)

            for t in threads:
                t.join()

            execute('/bin/rm -rf {0}/xtremio_file-*.zf'.format(mounted))
            print('Exiting Main Thread {0}'.format(time.ctime(time.time())))

if __name__ == "__main__":
    exit(main())

```

---

**Note:** While Increasing the percentage causes elevated precision, it may increase the probability of receiving a 'no free space' SCSI error during the reclamation process.

---

## Configuring Boot from SAN with an XtremIO Volume

For guidelines on configuring the host to boot from an XtremIO volume, refer to [“Configuring Boot from SAN with XtremIO” on page 155](#).

# CHAPTER 3

## Host Configuration for VMware® vSphere®

This chapter includes the following topics:

◆ Fibre Channel HBA Configuration.....	48
◆ Host Parameters Settings .....	53
◆ vCenter Server Parameter Settings .....	55
◆ vStorage API for Array Integration (VAAI) Settings .....	55
◆ Configuring VMware vSphere with XtremIO Storage in a Multiple Storage Array Configuration .....	59
◆ Multipathing Software Configuration .....	60
◆ Post-Configuration Steps - Using the XtremIO Storage .....	65

---

**Note:** XtremIO Storage Array supports both ESX and ESXi. For simplification, all references to ESX server/host apply to both ESX and ESXi, unless stated otherwise.

---



---

**Note:** In hosts running a hypervisor, such as VMware ESX or Microsoft Hyper-V, it is important to ensure that the logical unit numbers of XtremIO volumes are consistent across **all** hosts in the hypervisor cluster. Inconsistent LUNs may affect operations such as VM online migration or VM power-up.

---



---

**Note:** When using Jumbo Frames with VMware ESX, the correct MTU size must be set on the virtual switch as well.

---

## Fibre Channel HBA Configuration

When using Fibre Channel with XtremIO, the following FC Host Bus Adapters (HBA) issues should be addressed for optimal performance.

### Pre-Requisites

To install one or more EMC-approved HBAs on an ESX host, follow the procedures in one of these documents, according to the FC HBA type:

- ◆ For Qlogic and Emulex HBAs - Typically the driver for these HBAs is preloaded with ESX. Therefore, no further action is required. For details, refer to the vSphere and HBA documentation.
- ◆ For Cisco UCS fNIC HBAs (vsphere 5.x and above) - Refer to the *Virtual Interface Card Drivers* section in the *Cisco UCS Manager Install and Upgrade Guides* for complete driver installation instructions ([http://www.cisco.com/en/US/partner/products/ps10281/prod\\_installation\\_guides\\_list.html](http://www.cisco.com/en/US/partner/products/ps10281/prod_installation_guides_list.html)).

### Queue Depth

---

**Note:** Changing queue depth settings is designed for advanced users. Increasing queue depth may cause hosts to over-stress other arrays connected to the ESX host, resulting in performance degradation while communicating with them. To avoid this, in mixed environments with multiple array types connected to the ESX host, compare the XtremIO recommendations with those of other platforms before applying them.

---

Queue depth is the amount of SCSI commands (including I/O requests) that can be handled by a storage device at a given time. A queue depth can be set on either of the following:

- ◆ Initiator level - HBA queue depth
- ◆ LUN level - LUN queue depth

The LUN queue depth setting controls the amount of outstanding I/O requests per a single path. On vSphere, the LUN queue depth can be adjusted through the ESX CLI.

The HBA queue depth (also referred to as execution throttle) settings control the amount of outstanding I/O requests per HBA port.



The HBA queue depth should be set to the maximum value. This can be done on the HBA firmware level, using the HBA BIOS or CLI utility provided by the HBA vendor:

- ◆ Qlogic - Execution Throttle - This setting is no longer read by vSphere and is therefore not relevant when configuring a vSphere host with Qlogic HBAs.
- ◆ Emulex - lpfc\_hba\_queue\_depth - No need to change the default (and maximum) value (8192).
- ◆ Cisco UCS fNIC - The I/O Throttle setting determines the total number of outstanding I/O requests per virtual HBA. With Cisco UCS fNIC, it is recommended to set the I/O throttle to 1024.

For optimal operation with XtremIO storage, it is also recommended to adjust the LUN queue depth of the FC HBA.

**Note:** For further information on adjusting LUN queue depth with ESX, refer to VMware KB article 1267 on the VMware website ([http://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=1267](http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1267)).

**Note:** If the HBA queue depth is set to a value lower than the LUN queue depth, it may limit the LUN queue depth to a lower value than set.

**Note:** The setting adjustments for Cisco UCS fNIC HBA detailed in this section, apply to VMware vSphere only. Since these settings are global to the UCS chassis, they may impact other blades in the UCS chassis running a different OS (e.g. Windows).

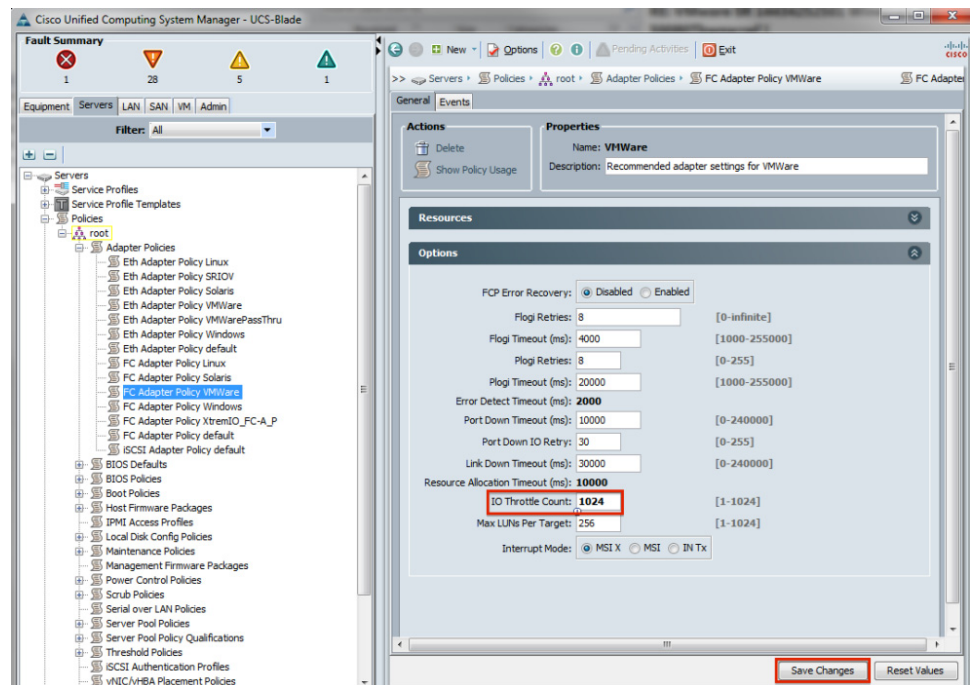
The following table summarizes the default and recommended queue depth settings for VMware vSphere.

	LUN Queue Depth	HBA Queue Depth
Default Value	Qlogic: 32 Emulex: 30 UCS: 32	Qlogic: 32 Emulex: 8192 UCS: 16
Recommended Value	Qlogic: 256 Emulex: 128 UCS: 128	Qlogic: 65536 (maximum) Emulex: 8192 (maximum) UCS: 1024

## Setting the HBA I/O Throttle of the Cisco UCS fNIC HBA

### To set HBA I/O throttle of the Cisco UCS fNIC HBA:

1. In the UCSM navigation tree, click the **Servers** tab.
2. In the navigation tree, expand the **Policies and Adapter Policies**.
3. Click the **FC Adapter Policy Linux** or **FC Adapter Policy VMWare**.
4. In the main window, expand the **Options** drop-down.
5. Configure the **I/O Throttle Count** field to 1024.
6. Click **Save Changes**.



**Note:** For more details on Cisco UCS fNIC FC adapter configuration, refer to:

<https://www.cisco.com/c/dam/en/us/solutions/collateral/data-center-virtualization/unified-computing/guide-c07-730811.pdf>

## Setting the LUN Queue Depth on a Host Running vSphere 5.x or Above

### To set the LUN queue depth on a host running vSphere 5.x or above:

1. Open an SSH session to the host as `root`.
2. Run one of the following commands to verify which HBA module is currently loaded:

HBA Vendor	Command
Qlogic	<code>esxcli system module list   egrep "ql Loaded"</code>
Emulex	<code>esxcli system module list   egrep "lpfc Loaded"</code>
Cisco UCS fNIC	<code>esxcli system module list   egrep "fnic Loaded"</code>

Example (for a host with Emulex HBA):

```
# esxcli system module list | egrep "lpfc|Loaded"
Name                               Is Loaded   Is Enabled
lpfc                               true        true
lpfc820                           false       true
```

In this example the native `lpfc` module for the Emulex HBA is currently loaded on ESX.

3. Run one of the following commands on the currently loaded HBA module, to set the LUN queue depth:

**Note:** The commands displayed in the table refer to the Qlogic `qla2xxx/qlnativefc`, Emulex `lpfc` and Cisco UCS `fnic` modules. Use an appropriate module name based on the output of the previous step.

HBA Vendor	Command
Qlogic	<b>vSphere 5.0/5.1:</b> <code>esxcli system module parameters set -p ql2xmaxqdepth=256 -m qla2xxx</code>
	<b>vSphere 5.5 (with Qlogic native drivers):</b> <code>esxcli system module parameters set -p ql2xmaxqdepth=256 -m qlnativefc</code>
	<b>vSphere 6.x:</b> <code>esxcli system module parameters set -p qlfxmaxqdepth=256 -m qlnativefc</code>
Emulex	<code>esxcli system module parameters set -p lpfc0_lun_queue_depth=128 -m lpfc</code>
Cisco UCS fNIC	<code>esxcli system module parameters set -p fnic_max_qdepth=128 -m fnic</code>

---

**Note:** The command for Emulex HBA sets the LUN queue depth for the lpfc0 Emulex HBA. If another Emulex HBA is connected to the XtremIO storage, change `lpfc0_lun_queue_depth` accordingly. For example, if lpfc1 Emulex HBA is connected to XtremIO, replace `lpfc0_lun_queue_depth` with `lpfc1_lun_queue_depth`.

---



---

**Note:** If all Emulex HBAs on the host are connected to the XtremIO storage, replace `lpfc0_lun_queue_depth` with `lpfc_lun_queue_depth`.

---

4. Reboot the ESX host.
5. Open an SSH session to the host as `root`.
6. Run the following command to confirm that queue depth adjustment is applied:

```
esxcli system module parameters list -m <driver>
```

---

**Note:** When using the command, replace `<driver>` with the module name, as received in the output of [step 2](#) (for example, `lpfc`, `qla2xxx` and `qlnativefc`).

---

Examples:

- For a vSphere 5.x host with Qlogic HBA and LUN queue depth set to 256:

```
# esxcli system module parameters list -m qla2xxx | grep ql2xmaxqdepth
ql2xmaxqdepth int 256 Max queue depth to report for target devices.
```

- For a vSphere 6.x host with Qlogic HBA and LUN queue depth set to 256:

```
# esxcli system module parameters list -m qlnativefc | grep qlfxmaxqdepth
qlfxmaxqdepth int 256 Maximum queue depth to report for target devices.
```

- For a host with Emulex HBA and LUN queue depth set to 128:

```
# esxcli system module parameters list -m lpfc | grep lpfc0_lun_queue_depth
lpfc0_lun_queue_depth int 128 Max number of FCP commands we can queue to a specific LUN
```

If LUN queue depth is set for all Emulex HBAs on the host, run the following command instead:

```
# esxcli system module parameters list -m lpfc | grep
lun_queue_depth
```

# Host Parameters Settings

This section details the ESX host parameters settings necessary for optimal configuration when using XtremIO storage.

---

**Note:** The following setting adjustments may cause hosts to over-stress other arrays connected to the ESX host, resulting in performance degradation while communicating with them. To avoid this, in mixed environments with multiple array types connected to the ESX host, compare these XtremIO recommendations with those of other platforms before applying them.

---

When using XtremIO storage with VMware vSphere, it is recommended to set the following parameters to their maximum values:

- ◆ `Disk.SchedNumReqOutstanding` – Determines the maximum number of active storage commands (I/Os) allowed at any given time at the VMkernel. The maximum value is 256.

---

**Note:** When using vSphere 5.5 or above, the `Disk.SchedNumReqOutstanding` parameter can be set on a specific volume rather than on all volumes presented to the host. Therefore, it should be set only after XtremIO volumes are presented to the ESX host using ESX command line.

---

- ◆ `Disk.SchedQuantum` – Determines the maximum number of consecutive “sequential” I/Os allowed from one VM before switching to another VM (unless this is the only VM on the LUN). The maximum value is 64.

In addition, the following parameter setting is required:

- ◆ `Disk.DiskMaxIOSize` – Determines the maximum I/O request size passed to storage devices. With XtremIO, it is required to change it from 32767 (default setting of 32MB) to 4096 (4MB). This adjustment allows a Windows VM to EFI boot from XtremIO storage with a supported I/O size of 4MB.

---

**Note:** For details on the possible Windows EFI boot issue with XtremIO storage (in case the above maximum I/O block size setting adjustment cannot be done), refer to VMware KB article 2137402 on the VMware website (<http://kb.vmware.com/kb/2137402>).

---



---

**Note:** For details on adjusting the maximum I/O block size in ESX, refer to VMware KB article 1003469 on the VMware website ([http://kb.vmware.com/selfservice/microsites/search.do?cmd=displayKC&docType=k&docTypeID=DT\\_KB\\_1\\_1&externalId=1003469](http://kb.vmware.com/selfservice/microsites/search.do?cmd=displayKC&docType=k&docTypeID=DT_KB_1_1&externalId=1003469)).

---

These setting adjustments should be carried out on each ESX host connected to XtremIO cluster via either the vSphere Client or the ESX command line.

## Adjusting the ESX Host Parameters for XtremIO Storage

To adjust ESX host parameters for XtremIO storage, follow one of these procedures:

◆ Using the vSphere WebUI client:

1. Launch the vSphere Web client and navigate to **Home > Hosts and Clusters**.
2. In the left menu section, locate the ESX host and click it.
3. In the right pane, click **Manage > Settings**.
4. From the **System** section, click **Advanced System Settings**.
5. Locate the `Disk.SchedNumReqOutstanding` parameter. Click the **Edit** icon and set the parameter to its maximum value (256).

---

**Note:** Do not apply [step 5](#) in a vSphere 5.5 (or above) host, where the parameter is set on a specific volume using ESX command line.

---

6. Locate the `Disk.SchedQuantum` parameter. Click the **Edit** icon and set it to its maximum value (64).
7. Locate the `Disk.DiskMaxIOSize` parameter. Click the **Edit** icon and set it to 4096.
8. Click **OK** to apply the changes.

◆ Using the ESX host command line (for vSphere 5.0 and 5.1):

1. Open an SSH session to the host as `root`.
2. Run the following commands to set the `SchedQuantum`, `SchedNumReqOutstanding`, and `DiskMaxIOSize` parameters, respectively:
  - `esxcfg-advcfg -s 64 /Disk/SchedQuantum`
  - `esxcfg-advcfg -s 256 /Disk/SchedNumReqOutstanding`
  - `esxcfg-advcfg -s 4096 /Disk/DiskMaxIOSize`

◆ Using the ESX host command line (for vSphere 5.5 or above):

1. Open an SSH session to the host as `root`.
2. Run the following commands to set the `SchedQuantum` and `DiskMaxIOSize` parameters, respectively:
  - `esxcfg-advcfg -s 64 /Disk/SchedQuantum`
  - `esxcfg-advcfg -s 4096 /Disk/DiskMaxIOSize`
3. Run the following command to obtain the NAA for XtremIO LUNs presented to the ESX host and locate the NAA of the XtremIO volume:
  - `esxcli storage nmp path list | grep XtremIO -B1`
4. Run the following command to set `SchedNumReqOutstanding` for the device to its maximum value (256):
  - `esxcli storage core device set -d naa.xxx -O 256`

## vCenter Server Parameter Settings

The maximum number of concurrent full cloning operations should be adjusted, based on the XtremIO cluster size. The vCenter Server parameter `config.vpxd.ResourceManager.maxCostPerHost` determines the maximum number of concurrent full clone operations allowed (the default value is 8). Adjusting the parameter should be based on the XtremIO cluster size as follows:

- ◆ 10TB Starter X-Brick (5TB) and a single X-Brick - 8 concurrent full clone operations
- ◆ Two X-Bricks - 16 concurrent full clone operations
- ◆ Four X-Bricks - 32 concurrent full clone operations
- ◆ Six X-Bricks - 48 concurrent full clone operations

**To adjust the maximum number of concurrent full cloning operations:**

1. Launch vSphere WebUI client to log in to the vCenter Server.
2. From the top menu, select **vCenter Inventory List**.
3. From the left menu, under Resources, Click **vCenter Servers**.
4. Select **vCenter > Manage Tab > Settings > Advanced Settings**.
5. Click **Edit**.
6. Locate the `config.vpxd.ResourceManager.maxCostPerHost` parameter and set it according to the XtremIO cluster size. If you cannot find the parameter, type its name in the **Key** field and the corresponding value in the **Value** field.
7. Click **Add**.
8. Click **OK** to apply the changes.

## vStorage API for Array Integration (VAAI) Settings

VAAI is a vSphere API that offloads vSphere operations such as virtual machine provisioning, storage cloning and space reclamation to storage arrays that supports VAAI. XtremIO Storage Array fully supports VAAI.

To ensure optimal performance of XtremIO storage from vSphere, VAAI must be enabled on the ESX host before using XtremIO storage from vSphere. Failing to do so may expose the XtremIO cluster to the risk of datastores becoming inaccessible to the host.

This section describes the necessary settings for configuring VAAI for XtremIO storage.

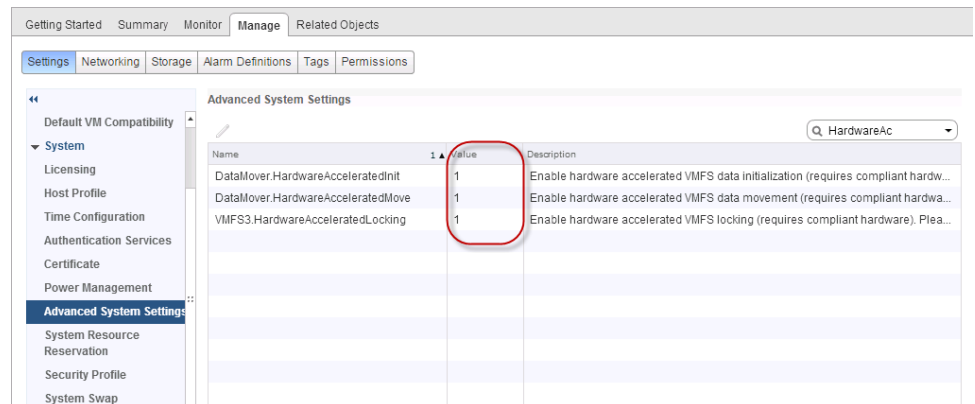
## Enabling VAAI Features

### Confirming that VAAI is Enabled on the ESX Host

When using vSphere version 5.x and above, VAAI is enabled by default. Before using the XtremIO storage, confirm that VAAI features are enabled on the ESX host.

#### To confirm that VAAI is enabled on the ESX host:

1. Launch the vSphere Web Client and navigate to **Home > Hosts and Clusters**.
2. In the left menu section, locate the ESX host. and click it.
3. In the right pane, click **Manage > Settings**.
4. From the **System** section, click **Advanced System Settings**.
5. Verify that the following parameters are enabled (i.e. both are set to "1"):
  - `DataMover.HardwareAcceleratedMove`
  - `DataMover.HardwareAcceleratedInit`
  - `VMFS3.HardwareAcceleratedLocking`



If any of the above parameters are not enabled, adjust them by clicking the **Edit** icon and click **OK**.



## Manually Setting VAAI on Datastore

### NOTICE

If VAAI setting is enabled after a datastore was created on XtremIO storage, the setting does not automatically propagate to the corresponding XtremIO Volumes. The setting must be manually configured to avoid data unavailability to the datastore.

Perform the following procedure on all datastores created on XtremIO storage before VAAI is enabled on the ESX host.

**To manually set VAAI setting on a VMFS-5 datastore created on XtremIO storage with VAAI disabled on the host:**

1. Confirm that the VAAI Hardware Accelerator Locking is enabled on this host. For details, refer to [“Confirming that VAAI is Enabled on the ESX Host” on page 56](#).
2. Using the following `vmkfstools` command, confirm that the datastore is configured as "public ATS-only":

```
# vmkfstools -Ph -v1 <path to datastore> | grep public
```

- In the following example, a datastore volume is configured as "public":

```
# vmkfstools -Ph -v1 /vmfs/volumes/datastore1 | grep public
Mode: public
```

- In the following example, a datastore volume is configured as "public ATS-only":

```
# vmkfstools -Ph -v1 /vmfs/volumes/datastore2 | grep public
Mode: public ATS-only
```

3. If the datastore was found with mode "public", change it to "public ATS-only" by executing the following steps:
  - a. Unmount the datastore from all ESX hosts on which it is mounted (except one ESX host).
  - b. Access the ESX host on which the datastore is still mounted.
  - c. Run the following `vmkfstools` command to enable ATS on the datastore:
 

```
# vmkfstools --configATSONly 1 <path to datastore>
```
  - d. Click **0** to continue with ATS capability.
  - e. Repeat step 2 to confirm that ATS is set on the datastore.
  - f. Unmount datastore from the last ESX host.
  - g. Mount datastore on all ESX hosts.

## Tuning VAAI XCOPY with XtremIO

By default, vSphere instructs the storage array to copy data in 4MB chunks. To optimize VAAI XCOPY operation with XtremIO, it is recommended to adjust the chunk size to 256KB. The VAAI XCOPY chunk size is set using the `MaxHWTransferSize` parameter.

To adjust the VAAI XCOPY chunk size, run the following CLI commands according to the vSphere version running on your ESX host:

- ◆ For vSphere version earlier than 5.5:

```
esxcli system settings advanced list -o
/DataMover/MaxHWTransferSize

esxcli system settings advanced set --int-value 0256
--option /DataMover/MaxHWTransferSize
```

- ◆ For vSphere version 5.5 and above:

```
esxcfg-advcfg -s 0256 /DataMover/MaxHWTransferSize
```

## Disabling VAAI in ESX

In some cases (mainly for testing purposes) it is necessary to temporarily disable VAAI.

As a rule, VAAI should be enabled on an ESX host connected to XtremIO. Therefore avoid disabling VAAI if possible. If not, disable it temporarily.

---

**Note:** For further information about disabling VAAI, refer to VMware KB article 1033665 on the VMware website

([http://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=1033665](http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1033665)).

---

### NOTICE

As noted in the Impact/Risk section of VMware KB 1033665, disabling the ATS (Atomic Test and Set) parameter can cause data unavailability in ESXi 5.5 for volumes created natively as VMFS5 datastore.

---

### To disable VAAI on the ESX host:

1. Browse to the host in the vSphere Web Client navigator.
2. Select the **Manage** tab and click **Settings**.
3. In the System section, click **Advanced System Settings**.
4. Click **Edit** and modify the following parameters to disabled (i.e. set to zero):
  - `DataMover.HardwareAcceleratedMove`
  - `DataMover.HardwareAcceleratedInit`
  - `VMFS3.HardwareAcceleratedLocking`
5. Click **OK** to apply the changes.

# Configuring VMware vSphere with XtremIO Storage in a Multiple Storage Array Configuration

The following table shows the recommended vSphere settings when multiple storage arrays are connected to vSphere (in addition to XtremIO). These recommendations should be followed instead of the other recommendations included in this chapter.

For reference, this table also includes the corresponding recommendations for settings when vSphere is connected to XtremIO storage only.

Parameter Name	Scope/Granularity	Multi-Storage Setting	XtremIO Only Setting
FC Adapter Policy IO Throttle Count	Per vHBA	256	1024
fnic_max_qdepth	Global	32	128
Disk.SchedNumReqOutstanding	LUN <sup>1</sup>	32	256
Disk.SchedQuantum	Global	8	64
Disk.DiskMaxIOSize	Global	4MB	4MB
XCOPY (/DataMover/MaxHWTransferSize)	Global	4MB	256KB
config.vpxd.ResourceManager.maxCostPerHost	vCenter	8	8 per X-Brick (up to 48)

1: When using vSphere 5.5 or above, the parameter can be set on a specific volume, as noted here. With earlier vSphere versions, this is an ESX global setting.

- ◆ FC Adapter Policy IO Throttle Count - The total number of I/O requests that can be outstanding on a per-virtual host bus adapter (vHBA) in UCS. For details, refer to [“Setting the HBA I/O Throttle of the Cisco UCS fNIC HBA” on page 50](#).
- ◆ fnic\_max\_qdepth - A UCS FC HBA driver level setting that manages the total number of I/O requests that can be outstanding on a per-LUN basis. For details, refer to [“Setting the LUN Queue Depth on a Host Running vSphere 5.x or Above” on page 51](#) (the steps relevant to the Cisco UCS fNIC HBA).
- ◆ Disk.SchedNumReqOutstanding - The total number of outstanding commands permitted from all virtual machines collectively on the host to a LUN. For details, refer to [“Adjusting the ESX Host Parameters for XtremIO Storage” on page 54](#).
- ◆ Disk.SchedQuantum - The maximum number of consecutive “sequential” I/Os allowed from one VM before forcing a switch to another VM. For details, refer to [“Adjusting the ESX Host Parameters for XtremIO Storage” on page 54](#).
- ◆ Disk.DiskMaxIOSize - The maximum I/O size ESX allows before splitting I/O requests. For details, refer to [“Adjusting the ESX Host Parameters for XtremIO Storage” on page 54](#).
- ◆ XCOPY (/DataMover/MaxHWTransferSize) - The maximum number of blocks used for XCOPY operations. For details, refer to [“Tuning VAAI XCOPY with XtremIO” on page 58](#).
- ◆ config.vpxd.ResourceManager.maxCostPerHost - The maximum number of concurrent full clone operations allowed (the default value is 8). For details, refer to [“vCenter Server Parameter Settings” on page 55](#).

# Multipathing Software Configuration

---

**Note:** You can use EMC Virtual Storage Integrator (VSI) Path Management to configure path management across EMC platforms, including XtremIO. For information on using this vSphere Client plug-in, refer to the *EMC VSI Path Management Product Guide*.

---

## Configuring vSphere Native Multipathing

XtremIO supports the VMware vSphere Native Multipathing (NMP) technology. This section describes the procedure required for configuring native vSphere multipathing for XtremIO volumes.

For best performance, it is recommended to do the following:

- ◆ Set the native `round robin` path selection policy on XtremIO volumes presented to the ESX host.

---

**Note:** With NMP in vSphere versions below 5.5, clustering is not supported when the path policy is set to Round Robin. For details, see *vSphere MSCS Setup Limitations* in the Setup for Failover Clustering and Microsoft Cluster Service guide for [ESXi 5.0](#) or [ESXi/ESX 4.x](#). In vSphere 5.5, Round Robin PSP (PSP\_RR) support is introduced. For details, see [MSCS support enhancements in vSphere 5.5 \(VMware KB 2052238\)](#).

---

- ◆ Set the vSphere NMP Round Robin path switching frequency to XtremIO volumes from the default value (1000 I/O packets) to 1.

These settings ensure optimal distribution and availability of load between I/O paths to the XtremIO storage.

---

**Note:** Use the ESX command line to adjust the path switching frequency of vSphere NMP Round Robin.

---

To set vSphere NMP Round-Robin configuration, it is recommended to use the ESX command line for all the XtremIO volumes presented to the host. Alternatively, for an XtremIO volume that was already presented to the host, use one of the following methods:

- ◆ Per volume, using vSphere Client (for each host where the volume is presented)
- ◆ Per volume, using ESX command line (for each host where the volume is presented)

The following procedures detail each of these three methods.

**To configure vSphere NMP Round Robin as the default pathing policy for all XtremIO volumes, using the ESX command line:**

---

**Note:** Use this method when no XtremIO volume is presented to the host. XtremIO volumes already presented to the host are not affected by this procedure (unless they are unmapped from the host).

---

1. Open an SSH session to the host as `root`.
2. Run the following command to configure the default pathing policy for newly defined XtremIO volumes to **Round Robin** with path switching after each I/O packet:

```
esxcli storage nmp satp rule add -c tpgs_off -e "XtremIO
Active/Active" -M XtremApp -P VMW_PSP_RR -O iops=1 -s
VMW_SATP_DEFAULT_AA -t vendor -V XtremIO
```

This command also sets the vSphere NMP Round Robin path switching frequency for newly defined XtremIO volumes to one (1).

---

**Note:** Using this method does not impact any non-XtremIO volume presented to the ESX host.

---

**To configure vSphere NMP Round Robin on an XtremIO volume in an ESX host, using vSphere WebUI Client:**

1. Browse to the host in the vSphere Web Client navigator.
2. Select the **Manage** tab and click **Storage**.
3. Click **Storage Devices**.
4. Locate the XtremIO volume and select the **Properties** tab.
5. Under Multipathing Properties, click **Edit Multipathing**.
6. From the Path Selection policy drop-down list, select **Round Robin (VMware)** policy.
7. Click **OK** to apply the changes.
8. Click **Edit Multipathing** and verify that all listed paths to the XtremIO Volume are set to Active (I/O) status.

Path selection policy:

Round Robin (VMware)

Select the preferred path for this policy:

Filter

Runtime Name	Status	Target	LUN	Preferred
vmhba3:C0:T0:L19	Active (I/O)	51:4f:0c:50:23:80:46:00 51:4f:0c:50:2...	19	
vmhba3:C0:T1:L19	Active (I/O)	51:4f:0c:50:23:80:46:00 51:4f:0c:50:2...	19	
vmhba1:C0:T0:L19	Active (I/O)	51:4f:0c:50:23:80:46:00 51:4f:0c:50:2...	19	
vmhba1:C0:T1:L19	Active (I/O)	51:4f:0c:50:23:80:46:00 51:4f:0c:50:2...	19	

OK Cancel

**To configure vSphere NMP Round Robin on an XtremIO volume in an ESX host, using ESX command line:**

1. Open an SSH session to the host as `root`.
2. Run the following command to obtain the NAA of XtremIO LUNs presented to the ESX host:

```
#esxcli storage nmp path list | grep XtremIO -B1
```

3. Run the following command to modify the path selection policy on the XtremIO volume to Round Robin:

```
esxcli storage nmp device set --device <naa_id> --psp  
VMW_PSP_RR
```

Example:

```
#esxcli storage nmp device set --device naa.514f0c5e3ca0000e  
--psp VMW_PSP_RR
```

---

**Note:** When using this method, it is not possible to adjust the vSphere NMP Round Robin path switching frequency. Adjusting the frequency changes the NMP PSP policy for the volume from `round robin` to `iops`, which is not recommended with XtremIO. As an alternative, use the first method described in this section.

---

For details, refer to VMware KB article 1017760 on the VMware website ([http://kb.vmware.com/selfservice/microsites/search.do?cmd=displayKC&docType=kc&docTypeID=DT\\_KB\\_1\\_1&externalId=1017760](http://kb.vmware.com/selfservice/microsites/search.do?cmd=displayKC&docType=kc&docTypeID=DT_KB_1_1&externalId=1017760)).

## Configuring PowerPath Multipathing

**Note:** For the most updated information on PowerPath support with XtremIO storage, refer to the *XtremIO Simple Support Matrix*.

XtremIO supports multipathing using EMC PowerPath/VE on VMware vSphere. PowerPath/VE versions 5.9 SP1 and above provide Native Loadable Array Module (LAM) for XtremIO Array devices. With this support, XtremIO devices running versions 2.2 and above are managed under the XtremIO class.

PowerPath/VE provides enhanced path management capabilities for up to 32 paths per logical device and intelligent dynamic I/O load-balancing functionalities that are transparent to VMware vSphere, as well as to Guest operating systems. Having multiple paths enables the host to access a storage device even if a specific path is unavailable. Multiple paths share the I/O traffic to a storage device, using intelligent load-balancing policies which enhance I/O performance and increase application availability. EMC PowerPath is the recommended multipathing choice.

PowerPath/VE features include:

- ◆ Multiple paths - enables higher availability and I/O performance.
- ◆ Path management insight capabilities - PowerPath characterizes I/O patterns and aids in diagnosing I/O problems due to flaky paths or unexpected latency values. The following capabilities are provided:
  - Read and write - in MB/seconds per LUN
  - Latency distribution - the high and low watermarks per path
  - Retries - the number of failed I/Os on a specific path
  - Virtual machine performance monitoring support - available from EMC PowerPath 6.0 SP1 and above
  - Support performance monitoring - for a specific virtual machine (VM) or for all VMs in the environment
- ◆ Autostandby - automatically detects intermittent I/O failures and places paths in autostandby (also known as flaky paths).
- ◆ Remote monitoring and management:
  - PowerPath Virtual Appliance (vApp)
  - Remote PowerPath CLI (rpowermt)
  - VSI for VMware vSphere Path Management
  - VMware vCenter Update Manager
  - VMware Auto Deploy



Further PowerPath related information:

- ◆ For details on the PowerPath/VE releases supported for your VMware vSphere host, refer to the *XtremIO Simple Support Matrix*.
- ◆ For details on native class support with XtremIO for your host, refer to the *EMC PowerPath/VE release notes* document for the PowerPath/VE version you are installing.
- ◆ For details on installing and configuring PowerPath/VE with XtremIO native class on your host, refer to the *EMC PowerPath on VMware vSphere Installation and Administration Guide* for the PowerPath/VE version you are installing. This guide provides the required information for placing XtremIO volumes under PowerPath/VE control.

## Post-Configuration Steps - Using the XtremIO Storage

When host configuration is completed, you can use the XtremIO storage from the host. For details on creating, presenting and managing volumes that can be accessed from the host via either GUI or CLI, refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

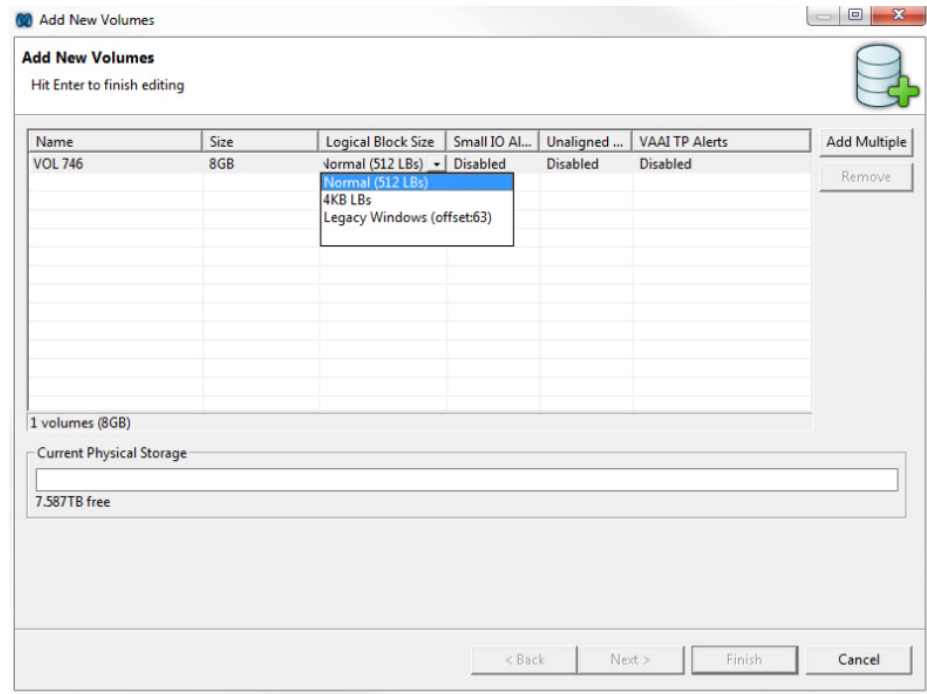
*EMC Virtual Storage Integrator (VSI) Unified Storage Management* version 6.2 and above can be used to provision from within vSphere Client Virtual Machine File System (VMFS) datastores and Raw Device Mapping volumes on XtremIO. Furthermore, *EMC VSI Storage Viewer* version 6.2 (and above) extends the vSphere Client to facilitate the discovery and identification of XtremIO storage devices allocated to VMware ESX/ESXi hosts and virtual machines.

For further information on using these two vSphere Client plug-ins, refer to the VSI Unified Storage Management product guide and the VSI Storage Viewer product guide.

## Disk Formatting

When creating volumes in XtremIO for a vSphere host, the following considerations should be made:

- ◆ Disk logical block size - The only logical block (LB) size supported by vSphere for presenting to ESX volumes is 512 bytes.



**Note:** In XtremIO version 4.0.0 (and above), the Legacy Windows option is not supported.

- ◆ Disk alignment - Unaligned disk partitions may substantially impact I/O to the disk. With vSphere, data stores and virtual disks are aligned by default as they are created. Therefore, no further action is required to align these in ESX.

With virtual machine disk partitions within the virtual disk, alignment is determined by the guest OS. For virtual machines that are not aligned, consider using tools such as UBERalign to realign the disk partitions as required.

## Presenting XtremIO Volumes to the ESX Host

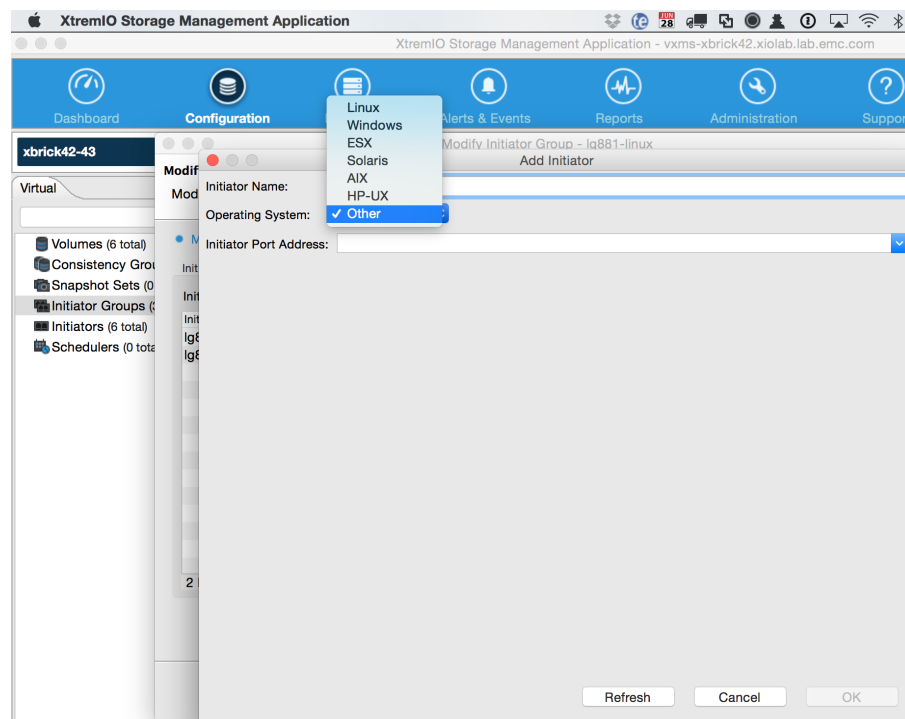
**Note:** It is recommended to avoid using host-level software for deduplication, compression and encryption of data stored in XtremIO. These operations are performed natively on the XtremIO cluster, thus releasing host computing resources. Furthermore, using these operations may severely impact data efficiency on the XtremIO cluster, as they can lead to a higher physical or logical capacity utilization.

**Note:** The following information in this section applies only to XtremIO version 4.0 and above.

**Note:** When using iSCSI software initiator with ESX and XtremIO storage, it is recommended to use only lower case characters in the IQN to correctly present the XtremIO volumes to ESX. For more details, refer to VMware KB article 2017582 on the VMware website.

[http://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=2017582](http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2017582)

When adding Initiator Groups and Initiators to allow ESX hosts to access XtremIO volumes, specify **ESX** as the operating system for newly-created Initiators, as shown in the figure below.



---

**Note:** Setting the Initiator's Operating System is required for optimal interoperability and stability of the host with XtremIO storage. You can adjust the setting while the host is online and connected to the XtremIO cluster with no I/O impact.

---

---

**Note:** Refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

---

Following a cluster upgrade from XtremIO version 3.0.x to version 4.0 (or above), make sure to modify the operating system for each initiator that is connected to an ESX host.

## Creating a File System

---

**Note:** File system configuration and management are out of the scope of this document.

---

It is recommended to create the file system using its default block size (using a non-default block size may lead to unexpected behavior). Refer to your operating system and file system documentation.

## Using LUN 0 with XtremIO Storage

This section details the considerations and steps that should be performed when using LUN 0 with vSphere.

Notes on the use of LUN numbering:

- ◆ In XtremIO version 4.0.0 (or above), volumes are numbered by default starting from LUN id 1 (and not 0 as was the case in previous XtremIO versions).
- ◆ Although possible, it is not recommended to manually adjust the LUN id to 0, as it may lead to issues with some operating systems.

---

**Note:** In case LUN 0 is needed for an XtremIO volume (e.g. boot from SAN or PowerPath/VE), restart the ESX host if a rescan fails to find this volume.

---

- ◆ When a cluster is updated from XtremIO version 3.0.x to 4.0.x, an XtremIO volume with a LUN id 0 remains accessible following the upgrade.
- ◆ With XtremIO version 4.0.0 (or above), no further action is required if volumes are numbered starting from LUN id 1.

## Virtual Machine Formatting

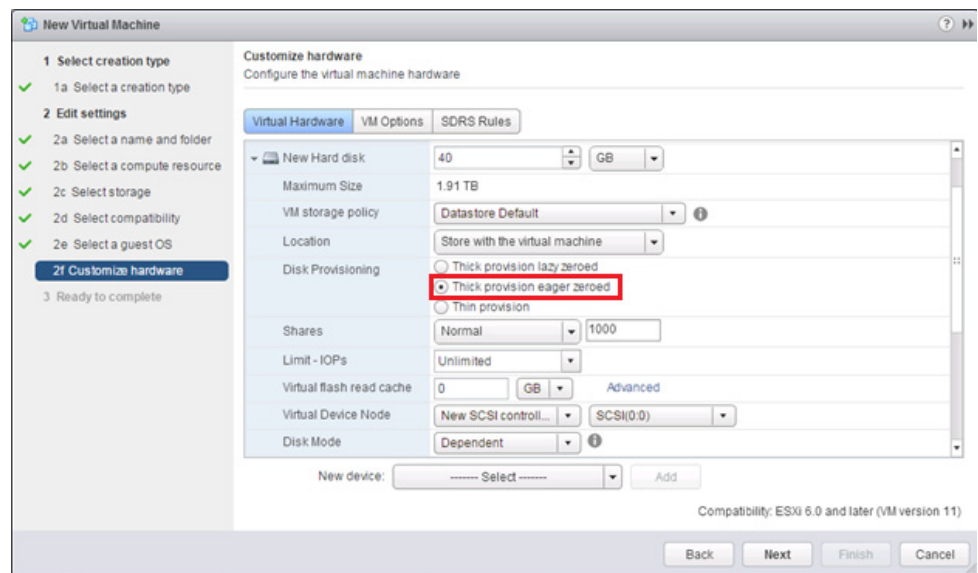
For optimal performance, it is recommended to format virtual machines on XtremIO storage, using Thick Provision Eager Zeroed. Using this format, the required space for the virtual machine is allocated and zeroed on creation time. However, with native XtremIO data reduction, thin provisioning, and VAAI support, no actual physical capacity allocation occurs.

Thick Provision Eager Zeroed format advantages are:

- ◆ Logical space is allocated and zeroed on virtual machine provisioning time, rather than scattered, with each I/O sent by the virtual machine to the disk (when Thick Provision Lazy Zeroed format is used).
- ◆ Thin provisioning is managed in the XtremIO Storage Array rather than in the ESX host (when Thin Provision format is used).

**To format a virtual machine using Thick Provision Eager Zeroed:**

1. From vSphere Web Client launch the **Create New Virtual Machine** wizard.
2. Proceed using the wizard up to the **2f Customize hardware** screen.
3. In the Customize hardware screen, click **Virtual Hardware**.
4. Toggle the **New Hard Disk** option.
5. Select the **Thick Provision Eager Zeroed** option to format the virtual machine's virtual disk.



6. Proceed using the wizard to complete creating the virtual machine.

## In-Guest Space Reclamation with Thin-Provision VMDK

For optimal space utilization with vSphere 6.x, it is recommended to format virtual machines on XtremIO storage, using Thin Provision. Using this format, in-guest space reclamation is available, provided the following requirements are fulfilled:

- ◆ Thin virtual disks
- ◆ VM hardware version 11
- ◆ ESXi 6.0
- ◆ EnableBlockDelete set to 1

Name	Value	Description
VMFS3.EnableBlockDelete	1	Enable VMFS block delete when UNMAP is issued from guest OS

- ◆ Guest OS support of UNMAP

**Note:** Some guest OSs that support unmapping of blocks, such as Linux-based systems, do not generate UNMAP commands on virtual disks in vSphere 6.0. This occurs because the level of SCSI support for ESXi 6.0 virtual disks is SCSI-2, while Linux expects 5 or higher for SPC-4 standard. This limitation prevents the generation of UNMAP commands until the virtual disks are able to claim support for at least SPC-4 SCSI commands.

For details on virtual volumes and UNMAP, refer to VMware KB# 2112333 ([http://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=2112333](http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2112333)).

### To format a virtual machine using the Thin Provision option:

1. From vSphere Web Client, launch the Create New Virtual Machine wizard.
2. Proceed, using the wizard, up to the 2f Customize Hardware screen.
3. In the Customize Hardware screen, click **Virtual Hardware**.
4. Toggle to the **New Hard Disk** option.
5. Select the **Thin Provision** option to format the virtual machine's virtual disk.

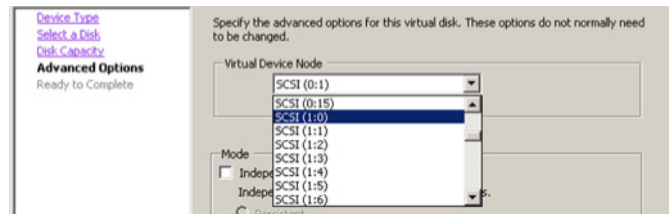
New Hard disk	40	GB
Maximum Size	73.62 GB	
VM storage policy	Datastore Default	
Location	Store with the virtual machine	
Disk Provisioning	<input type="radio"/> Thick provision lazy zeroed <input type="radio"/> Thick provision eager zeroed <input checked="" type="radio"/> Thin provision	
Shares	Normal	1000
Limit - IOPs	Unlimited	
Virtual flash read cache	0	GB
Virtual Device Node	SCSI controller 0	SCSI(0:5)
Disk Mode	Dependent	

For details on migrating a virtual machine from Thick provision to Thin provision, refer to VMware KB# 2014832 ([https://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=2014832](https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2014832)).

## Virtual Machine Guest OS Settings

This section details the recommended settings and considerations for virtual machines guest OS.

- ◆ LUN Queue Depth - For optimal virtual machine operation, configure the virtual machine guest OS to use the maximum queue depth of the virtual SCSI controller. For details on adjusting the guest OS LUN queue depth, refer to VMware KB 2053145 on the VMware website (<http://kb.vmware.com/kb/2053145>).
- ◆ RDM volumes in Guest OS - Span RDM volumes, used by the virtual machine, across SCSI controllers to prevent a bottleneck on a single SCSI controller.



- ◆ RDM volumes in guest OS used for Microsoft Cluster (MSCS) - ESX hosts with visibility to RDM volumes, used by Microsoft Cluster (MSCS), may take a long time to start or to perform LUN rescan. For the required settings on the RDM volumes, refer to VMware KB 1016106 ([http://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=1016106](http://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=1016106)).

## Space Reclamation

This section provides a comprehensive list of capacity management steps for achieving optimal capacity utilization on the XtremIO array, when connected to an ESX host.

Data space reclamation helps to achieve optimal XtremIO capacity utilization. Space reclamation is a vSphere function, enabling to reclaim used space by sending zeros to a specific address of the volume, after being notified by the file system that the address space was deleted.

Unlike traditional operating systems, ESX is a hypervisor, running guest operating systems on its file-system (VMFS). As a result, space reclamation is divided to guest OS and ESX levels.

ESX level space reclamation should be run only when deleting multiple VMs, and space is reclaimed from the ESX datastore. Guest level space reclamation should be run as a periodic maintenance procedure to achieve optimal capacity savings.

The following figure displays a scenario in which VM2 is deleted while VM1 and VM3 remain.



### Space Reclamation at Guest Level

**Note:** Refer to the relevant OS chapter to run space reclamation in the guest OS level.

On VSI environments, every virtual server should be treated as a unique object. When using VMDK devices, T10 trim commands are blocked. Therefore, it is required to run space reclamation manually. RDM devices pass through T10 trim commands.

There are two types of VDI provisioning that differ by their space reclamation guidelines:

- ◆ Temporary desktop (Linked Clones) - Normally, temporary desktops are deleted once the end users log off. Therefore, running space reclamation on the guest OS is not relevant, and only ESX level space reclamation should be used.
- ◆ Persistent desktop (Full Clones) - Persistent desktop contains long-term user data. Therefore, space reclamation should be run on guest OS level first, and only then on ESX level.

On large-scale VSI/VDI environments, it is recommended to divide the VMs to groups to avoid overloading the SAN fabric.



## Space Reclamation at ESX Level

### ◆ ESX 5.1 and below

In versions prior to ESX 5.5, the `vmkfstools` command is used for space-reclamation. This command supports datastores up to 2TB.

The following example describes running `vmkfstool` on a datastore XtremIO\_DS\_1 with 1% free space to allow user writes.

```
# cd /vmfs/volumes/XtremIO_DS_1
# vmkfstools -y 99
```

Vmfs reclamation may fail due to T10 commands blocking (VPLEX). In such cases, it is required to apply a manual copy of zeroes to the relevant free space.

The following example describes running a manual script on X41-VMFS-3 datastore (refer to [“ESX Space Reclamation Script” on page 74](#)).

```
# ./reclaim_space.sh X41-VMFS-3
```

---

**Note:** The datastore name cannot include spaces.

---

### ◆ ESX 5.5 and above

ESX 5.5 introduces a new command for space reclamation and supports datastores larger than 2TB.

The following example describes running space reclamation on a datastore XtremIO\_DS\_1:

```
# esxcli storage vmfs unmap --volume-label=XtremIO_DS_1
--reclaim-unit=20000
```

The `reclaim-unit` argument is an optional argument, indicating the number of vmfs blocks to UNMAP per iteration.

Vmfs reclamation may fail due to T10 commands blocking (VPLEX). In such cases, it is required to apply a manual copy of zeroes to the relevant free space.

The following example describes running a manual script on X41-VMFS-3 datastore (refer to [“ESX Space Reclamation Script” on page 74](#)):

```
# ./reclaim_space.sh X41-VMFS-3
```

---

**Note:** The datastore name cannot include spaces.

---

## ESX Space Reclamation Script

The following example describes an ESX space reclamation script usage, followed by the ESX space reclamation script.

```
# python linux-reclaim.py --help
usage: esx-reclamation.py [-c <cluster_name>]

optional arguments:
  -h, --help            show this help message and exit
  -v, --version          show program's version number and exit
  -t, --thread_number    Threads Amount
  -m, --mount_point      Mount Point Name

# python esx-reclamation.py --thread_number 4 --mount_point
XBR143-VMFS-01

#!/usr/bin/env python
from __future__ import print_function
import time
import threading
import subprocess
import argparse
import re

class myThread (threading.Thread):
    def __init__(self, name, count, running):
        threading.Thread.__init__(self)
        self.name = name
        self.count = count
        self.running = running

    def run(self):
        dd_something(self.name, self.count, self.running)

def dd_something(name, count, running):
    execute('/bin/dd count={1} bs=131072 if=/dev/zero
of={0}/xtremio_file-{2}.zf conv=fxync'.format(name, count, running))

def execute(command):
    return subprocess.Popen(command, shell=True,
stdout=subprocess.PIPE,
stderr=subprocess.PIPE).communicate()[0].splitlines()

def get_mount_points():
    return execute('df -m')

threadLock = threading.Lock()
```

```

def main():
    parser = argparse.ArgumentParser(usage='% (prog) s [-c
<cluster_name>]')
    parser.add_argument('-v', '--version', action='version',
version='% (prog) s 1.02')
    parser.add_argument('-t', '--thread_number', action='store',
dest='thread_number', default='1', required=False, help='Threads
Amount', metavar='')
    parser.add_argument('-m', '--mount_point', action='store',
dest='mount_point', default=None, required=True, help='Mount Point
Name', metavar='')
    args = parser.parse_args()
    thread_number = int(args.thread_number)
    mount_point = args.mount_point
    print('Starting Main Thread {0}'.format(time.ctime(time.time())))
    threads = []

    for entry in get_mount_points():
        if mount_point in entry:
            filesystem, mblocks, used, available, usep, mounted =
map(str.strip, re.split(" +", entry))

            for i in xrange(thread_number):
                i = myThread(mounted, int(int(available) * 0.95 /
thread_number) * 8, str(i))
                i.start()
                threads.append(i)
            for t in threads:
                t.join()

            execute('/bin/rm -rf {0}/xtremio_file-*.zf'.format(mounted))
            print('Exiting Main Thread {0}'.format(time.ctime(time.time())))

if __name__ == "__main__":
    exit(main())

```

---

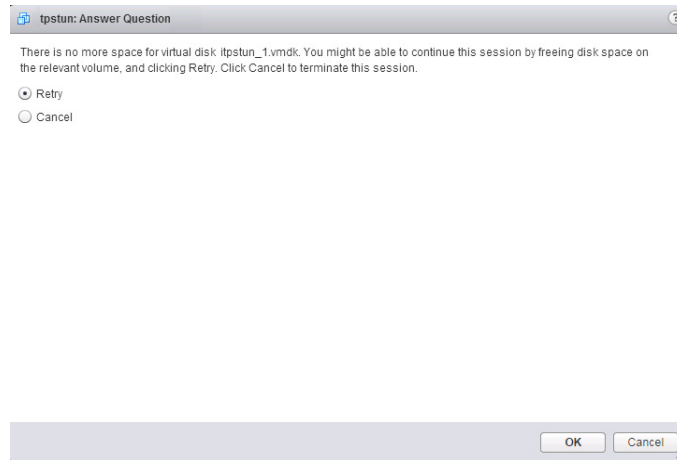
**Note:** While Increasing percentage leads to elevated precision, it may increase the probability of receiving a 'no free space' SCSI error during the reclamation.

---

## Out of Space VM Suspend and Notification with Thin Provisioning (TPSTUN)

TPSTUN is a VAAI primitive that enables the array to notify vSphere when a LUN is running out of space due to thin provisioning over-commit. The command causes suspending all virtual machines on that LUN. XtremIO supports this VAAI primitive.

A virtual machine provisioned on a LUN that is approaching full capacity usage becomes suspended, and the following message appears:



At this point, the VMware administrator can resolve the out-of-space situation on the XtremIO cluster, and prevent the guest OS in the VMs from crashing.

## Configuring Boot from SAN with an XtremIO Volume

For guidelines on configuring the host to boot from an XtremIO volume, refer to [“Configuring Boot from SAN with XtremIO” on page 155](#).

# CHAPTER 4

## Host Configuration for Microsoft® Windows®

This chapter includes the following topics:

- ◆ [Fibre Channel HBA Configuration.....](#) 78
- ◆ [Windows Offloaded Data Transfer \(ODX\) Functionality .....](#) 82
- ◆ [Multipathing Software Configuration .....](#) 84
- ◆ [Post-Configuration Steps - Using the XtremIO Storage .....](#) 88

---

**Note:** In hosts running a hypervisor, such as VMware ESX or Microsoft Hyper-V, it is important to ensure that the logical unit numbers of XtremIO volumes are consistent across **all** hosts in the hypervisor cluster. Inconsistent LUNs may affect operations such as VM online migration or VM power-up.

---

## Fibre Channel HBA Configuration

This section details FC HBA related issues that must be addressed when using Fibre Channel with XtremIO.

### Pre-Requisites for FC HBA Configuration

To install one or more EMC approved Host Bus Adapters (HBAs) into a Windows host, follow one of these documents according to the FC HBA type:

**For Qlogic HBAs** - *EMC Host Connectivity with QLogic Fibre Channel and iSCSI Host Bus Adapters (HBAs) and Converged Network Adapters (CNAs) in the Windows Environment document*

**For Emulex HBAs** - *EMC Host Connectivity with Emulex Fibre Channel and iSCSI Host Bus Adapters (HBAs) and Converged Network Adapters (CNAs) in the Windows Environment document*

These documents provide guidance on configuring the host for connection to the EMC Storage Arrays over a Fibre Channel including any needed HBA BIOS settings. Both documents are available in the EMC OEM section of the Qlogic site at <http://www.qlogic.com>, or Emulex site at <http://www.emulex.com>. They can also be found on <http://support.EMC.com>.

### Queue Depth

---

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

To avoid this, in mixed environments with multiple array types connected to the Windows host, compare XtremIO recommendations for queue depth with those of other platform before applying them.

---

For optimal operation with XtremIO storage, consider adjusting the queue depth of the FC HBA.

Queue depth is the amount of SCSI commands (including I/O requests) that can be handled by storage device at a given time. Queue depth can be set on either of the following:

- ◆ Initiator level - HBA queue depth
- ◆ LUN level - LUN queue depth

The HBA queue depth (also referred to as execution throttle) setting controls the amount of outstanding I/O requests per HBA port. The HBA queue depth should be set to the maximum value. This can be done on the HBA firmware level using the HBA BIOS or CLI utility provided by the HBA vendor as follows:

- ◆ Qlogic - Execution Throttle - Change the default value (32) to 65536.
- ◆ Emulex - `lpfc_hba_queue_depth` - No need to change the default (and maximum) value (8192).

**Note:** HBA queue depth (execution throttle) does not apply to QLE2600 and QLE8300 Series Qlogic adapters, and is read only for 10GbE adapters.

For further information, refer to:

[http://qlogic.force.com/SupportCenter/articles/FAQ/What-are-the-HBA-Execution-Throttle-settings-best-practices?retURL=%2FSupportCenter%2Fapex%2FCustomer\\_Support\\_main&popup=true](http://qlogic.force.com/SupportCenter/articles/FAQ/What-are-the-HBA-Execution-Throttle-settings-best-practices?retURL=%2FSupportCenter%2Fapex%2FCustomer_Support_main&popup=true)

The LUN queue depth setting controls the amount of outstanding I/O requests per a single path. On Windows, the LUN queue depth can be adjusted via the Windows Registry.

**Note:** When the HBA queue depth is set to a value lower than the LUN queue depth, it may limit the LUN queue depth to a lower value than set.

The following table summarizes the default and recommended queue depth settings for Windows.

	LUN Queue Depth	HBA Queue Depth
Default Value	Qlogic: 32 Emulex: 30	Qlogic: 32 Emulex: 8192
Recommended Value	Qlogic: 256 Emulex: 128	Qlogic: 65536 (maximum) Emulex: 8192 (maximum)

The following procedures detail setting the LUN queue depth for Qlogic and Emulex HBAs as follows:

- ◆ Qlogic - Set the Qlogic HBA adapter LUN queue depth in Windows StorPort driver to 256 (maximum value).
- ◆ Emulex - Set the Emulex HBA adapter LUN queue depth in Windows to 128 (maximum value).

Follow the appropriate procedure according to the HBA type.

## Setting the LUN Queue Depth for Qlogic HBA

### To set the LUN queue depth for Qlogic HBA:

1. On the desktop, click **Start**, select **Run** and open the **REGEDIT** (Registry Editor).

---

**Note:** Some driver versions do not create the registry by default. In these cases the user needs to manually create the registry.

---

2. Select **HKEY\_LOCAL\_MACHINE** and follow the tree structure down to the Qlogic driver as follows:

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\services\ql2300\Parameters\Device
```

---

**Note:** In some cases the Windows host detects the Qlogic HBA as ql2300i (instead of ql2300). In such cases, the following registry tree structure should be used instead:

```
HKEY_LOCAL_MACHINE\SYSTEM\CurrentControlSet\services\ql2300i\Parameters\Device
```

---

3. Double-click `DriverParameter:REG_SZ:qd=32`.
4. Change the value of `qd` to 256.
5. If the string "`qd=`" does not exist, append the following text to the end of the string using a semicolon (";"):

```
qd=256
```

6. Click **OK**.
7. Exit the Registry Editor and reboot the Windows host.

---

**Note:** Setting the queue-depth per this procedure is not disruptive.

---

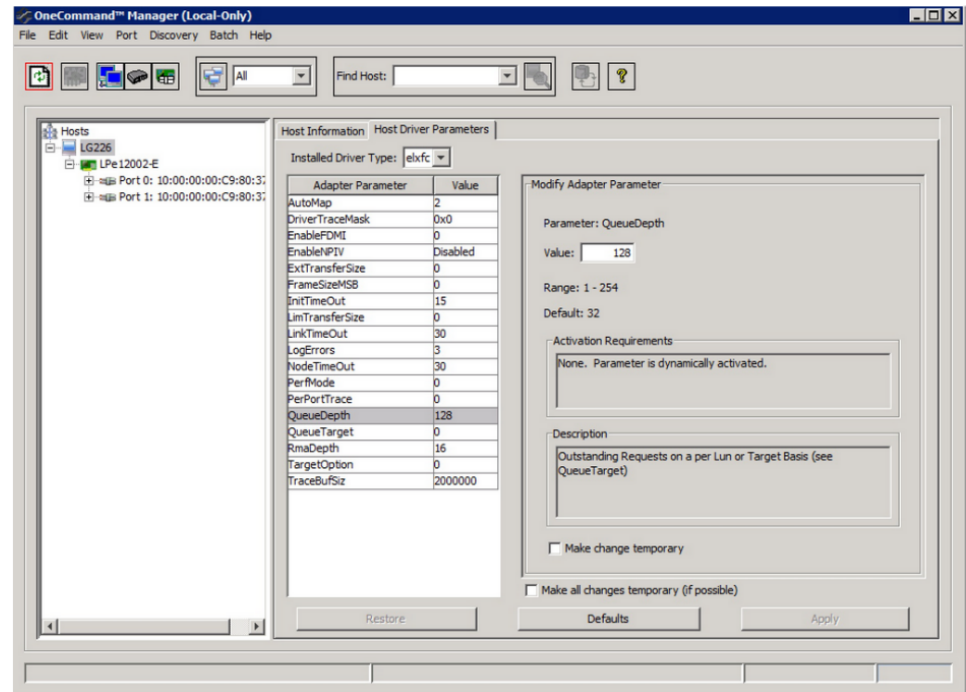


## Setting the LUN Queue Depth for Emulex FC HBA

Setting the LUN queue depth on the Emulex FC HBA includes using Emulex UI (OneCommand) to set the LUN queue depth of the Emulex driver HBA. OneCommand detects the active Emulex driver and enables changing the relevant driver's parameters, specifically LUN queue depth.

### To set the LUN queue depth on the Emulex FC HBA:

1. Install OneCommand.
2. Launch the OneCommand UI.



3. Select the relevant hostname from the Hosts list.
4. Select the **Host Driver Parameters** tab.
5. On the Adapter Parameter list, locate the QueueDepth parameter and set its value to 128.
6. Click **Apply**.

## Windows Offloaded Data Transfer (ODX) Functionality

Windows Offloaded Data Transfer (ODX) functionality in Windows maximizes the enterprise's investment by enabling a storage array to directly transfer data within or between compatible storage devices, bypassing the Windows host. By offloading the file transfer to the storage array, ODX minimizes latencies, maximizes array throughput and reduces resource usage such as CPU and network consumption on the host computer.

XtremIO version 4.0 (and higher) supports the Windows ODX functionality.

### Checking ODX Status

Although ODX in Windows is enabled by default, it can be disabled. Therefore, before using Windows with XtremIO, confirm that ODX is enabled.

**To verify that ODX is enabled:**

1. Run the following command:

```
Get-ItemProperty  
hklm:\system\currentcontrolset\control\filesystem -Name  
"FilterSupportedFeaturesMode"
```

2. Verify that the `FilterSupportedFeaturesMode` value in the Windows registry is 0.

### Enabling ODX

If ODX is disabled on Windows, perform the following steps to enable it.

**To enable ODX:**

1. Open a Windows PowerShell session as an administrator.
2. Run the following command:

```
Set-ItemProperty  
hklm:\system\currentcontrolset\control\filesystem -Name  
"FilterSupportedFeaturesMode" -Value 0
```

## Disabling ODX

In general, it is recommended to enable ODX functionality, when using Windows with XtremIO version 4.0 (or above), for optimal performance. However, in some cases (mainly for testing purposes), it is necessary to disable ODX.

Before disabling ODX, on each Windows host that uses ODX, list the file system filter drivers attached to the volume on which you want to disable ODX. Make sure that this list is empty.

### To list the file system drivers attached to a volume for ODX:

1. Open a Windows PowerShell session as an administrator.
2. Run the following command:

```
Fltmc instances -v <volume>
```

---

**Note:** <volume> refers to the drive letter of the volume.

---

The example below shows the expected `Fltmc` command output prior to disabling ODX:

```
PS C:\Users\Administrator> fltmc instances -v d:\
No instances for this volume
```

### To disable ODX:

1. Open a Windows PowerShell session as an administrator.
2. Check whether ODX is currently enabled. Refer to [“Checking ODX Status” on page 82](#).
3. If ODX is enabled on Windows, run the following command to disable ODX:

```
Set-ItemProperty
hklm:\system\currentcontrolset\control\filesystem -Name
"FilterSupportedFeaturesMode" -Value 1
```

## Multipathing Software Configuration

XtremIO supports native multipathing using Microsoft Native Microsoft Multipath I/O (MPIO) with Windows Server 2008 and above, or multipathing using PowerPath.

### Configuring Native Multipathing Using Microsoft Multipath I/O (MPIO)

For optimal operation with XtremIO storage, configure the **Least Queue Depth** policy for MPIO for devices presented from XtremIO. Using this policy, I/O is sent down the path with the fewest currently outstanding I/O requests.

Before configuring the native multipathing, you should enable MPIO on the server by adding the MPIO feature to Windows.

#### To enable MPIO on the server:

1. On the desktop, click **Start**, right-click **Computer**, and select the **Manage** option to display the Server Manager window.
2. On the Server Manager window, click **Features** to switch to the Features section, and then click **Add Feature** to launch the Add Feature Wizard.
3. From the list of possible features select **Multipath I/O** and click **Next**.
4. Click **Install** to install the MPIO on Windows.
5. If requested, reboot the host upon conclusion of the MPIO feature installation.

When MPIO is enabled on the server, perform the following procedure (using the command prompt) to configure the MPIO to handle XtremIO volumes and to set the load-balancing algorithm to Least Queue Depth as the default policy to all LUNs presented by the XtremIO Array.

**To configure native multipathing on XtremIO volumes presented to the Windows host:**

1. Run the following MPCLAIM command to install MPIO optional component and add multipath support for XtremIO device:

```
mpclaim -r -i -d "XtremIO XtremApp      "
```

---

**Note:** There should be eight (8) spaces between 'XtremApp' and the closing quotation mark.

---



---

**Note:** The host reboots immediately after running this command.

---

After reboot, use the `mpclaim -h` command to display the hardware IDs that are already managed by MPIO (the list should include XtremIO).

2. Run the following `mpclaim` command, to set the load-balancing algorithm to Least Queue Depth for all XtremIO volumes:

```
mpclaim -l -t "XtremIO XtremApp      " 4
```

---

**Note:** There should be eight (8) spaces between 'XtremApp' and the closing quotation mark.

---



---

**Note:** This command does not affect non-XtremIO volumes presented to the Windows host.

---



---

**Note:** Use the `mpclaim -s -t` command to check the default load-balancing settings for XtremIO devices. Use the `mpclaim -s -d` command to list all disks currently claimed by MPIO and their load-balancing settings.

---

## Configuring PowerPath Multipathing

**Note:** For the most updated information on PowerPath support with XtremIO storage, refer to the *XtremIO Simple Support Matrix*.

XtremIO supports multipathing using EMC PowerPath on Windows. PowerPath versions 5.7 SP2 and above provide Native Loadable Array Module (LAM) for XtremIO Array devices. With this support, XtremIO devices running versions 2.2 and above are managed under the XtremIO class.

PowerPath provides enhanced path management capabilities for up to 32 paths per logical device, as well as intelligent dynamic I/O load-balancing functionalities specifically designed to work within the Microsoft Multipathing I/O (MPIO) framework. Having multiple paths enables the host to access a storage device even if a specific path is unavailable. Multiple paths share the I/O traffic to a storage device, using intelligent load-balancing policies which enhance I/O performance and increase application availability. EMC PowerPath is the recommended multipathing choice.

PowerPath features include:

- ◆ Multiple paths - provides higher availability and I/O performance.
  - Includes the support on Server Core and Hyper-V (available in Windows Server 2008 and later).
- ◆ Running PowerPath in Hyper-V VMs (guest operating systems), PowerPath supports:
  - iSCSI through software initiator
  - Virtual Fibre Channel for Hyper-V (available in Windows Server 2012 and above) that provides the guest operating system with unmediated access to a SAN through vHBA
- ◆ Path management insight capabilities - PowerPath characterizes I/O patterns and aides in diagnosing I/O problems due to flaky paths or unexpected latency values. Metrics are provided on:
  - Read and write - in MB/seconds per LUN
  - Latency distribution - the high and low watermarks per path
  - Retries - the number of failed I/Os on a specific path
- ◆ Autostandby - automatically detects intermittent I/O failures and places paths into autostandby (also known as flaky paths).

- ◆ PowerPath Migration Enabler - is a host-based migration tool that allows migrating data between storage systems and supports migration in an MSCS environment (for Windows 2008 and later). PowerPath Migration Enabler works in conjunction with the host operating system (also called Host Copy) and other underlying technologies such as Open Replicator (OR).
- ◆ Remote monitoring and management
  - PowerPath Virtual Appliance 2.0 (vApp 2.0)
  - Systems Management Server (SMS)
  - Microsoft Operations Manager
  - SNMP management daemon

Further PowerPath related information:

- ◆ For details on the PowerPath releases supported for your Windows host, refer to the *XtremIO Simple Support Matrix*.
- ◆ For details on native class support with XtremIO for your host, refer to the *EMC PowerPath release notes* document for the PowerPath version you are installing
- ◆ For details on installing and configuring PowerPath with XtremIO native class support on your host, refer to the *EMC PowerPath on Windows Installation and Administration Guide* for the PowerPath version you are installing. This guide provides the required information for placing XtremIO volumes under PowerPath control.

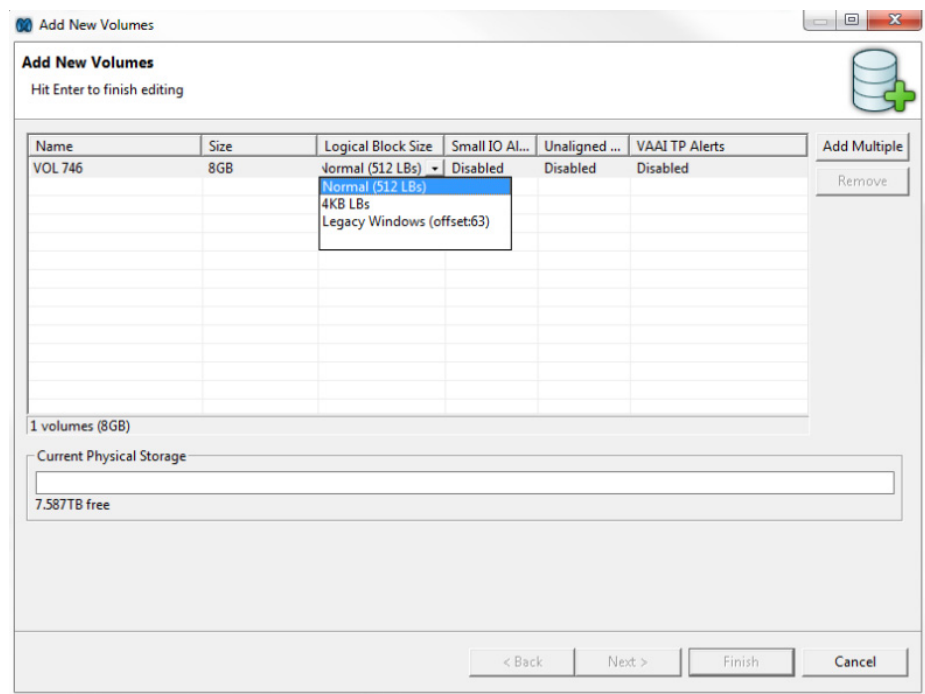
## Post-Configuration Steps - Using the XtremIO Storage

After the host configuration is completed, you can use the XtremIO storage from the host. For details on creating, presenting and managing volumes accessed from the host via either the GUI or the CLI, refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

### Disk Formatting

When creating volumes in XtremIO for a Windows host, the following considerations should be made:

- ◆ **Disk Logical Block Size** – Generally, a 512B logical block size should be used for a new XtremIO volume.



**Note:** In XtremIO version 4.0.0 (and above), the Legacy Windows option is not supported.

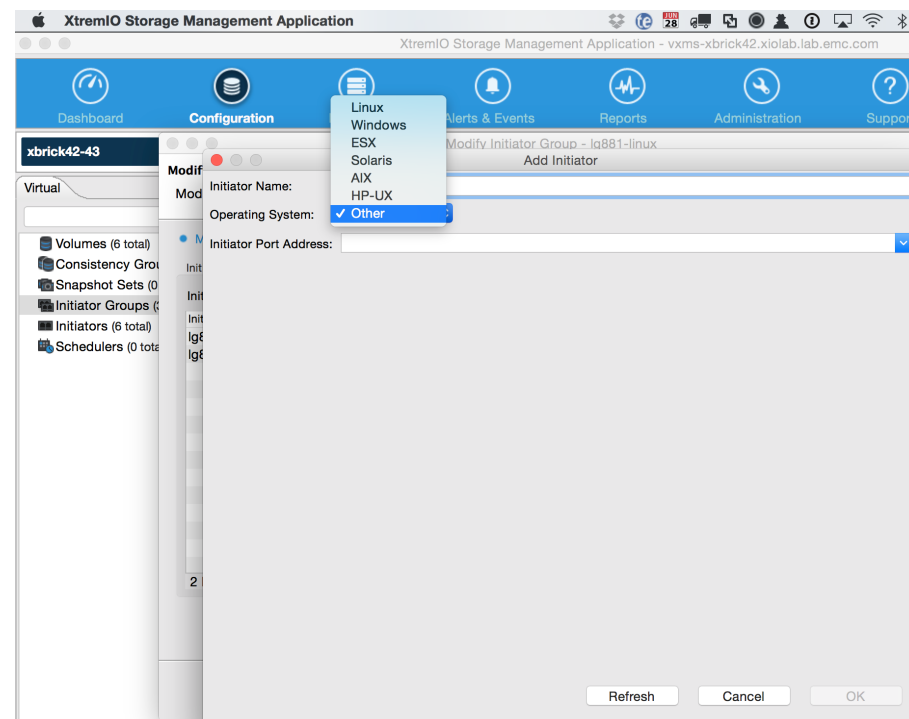


## Presenting XtremIO Volumes to the Windows Host

**Note:** It is recommended to avoid using host-level software for deduplication, compression and encryption of data stored in XtremIO. These operations are performed natively on the XtremIO cluster, thus releasing host computing resources. Furthermore, using these operations may severely impact data efficiency on the XtremIO cluster, as they can lead to a higher physical or logical capacity utilization.

**Note:** The following information in this section applies only to XtremIO version 4.0 and above.

When adding Initiator Groups and Initiators to allow Windows hosts to access XtremIO volumes, specify **Windows** as the operating system for the newly-created Initiators, as shown in the figure below.



**Note:** Setting the Initiator's Operating System is required for optimal interoperability and stability of the host with XtremIO storage. You can adjust the setting while the host is online and connected to the XtremIO cluster with no I/O impact.

**Note:** Refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

Following a cluster upgrade from XtremIO version 3.0.x to version 4.0 (or above), make sure to modify the operating system for each initiator that is connected to a Windows host.

## Creating a File System

---

**Note:** File system configuration and management are out of the scope of this document.

---

It is recommended to create the file system using an 8KB NTFS Cluster size (or a multiple of 8KB) to achieve optimal I/O from the Windows host with XtremIO storage. Refer to your operating system and file system documentation.

---

**Note:** Some file systems may require you to properly align the file system on the XtremIO volume. It is recommended to use specified tools to optimally match your host with application requirements.

---

## Space Reclamation

This section provides a comprehensive list of capacity management steps for achieving optimal capacity utilization on the XtremIO array, when connected to a Windows host.

Data space reclamation helps to achieve optimal XtremIO capacity utilization. Space reclamation is a Windows operating system function that enables to reclaim used space by sending zeros to a specific address of the volume after being notified by the file system that the address space was deleted. While some Windows operating systems can perform this action automatically, others require a user-initiated operation.

The following sections detail steps for performing space reclamation with:

- ◆ NTFS file systems
- ◆ Hyper-V

### NTFS File Systems

- ◆ Automatic space reclamation - On Windows server 2012 and above, NTFS supports automatic space reclamation when a file is deleted from the file system.

Delete notification (also referred to as TRIM or UNMAP) is a feature that notifies the underlying storage device of clusters that have been freed, following a file delete operation. The `DisableDeleteNotify` parameter is used to disable (using the value 1) or enable (using the value 0) delete notifications for all volumes. This parameter was introduced in Windows Server 2008 R2 and Windows 7.

The following example shows querying the delete notification status.

```
PS C:\> fsutil behavior query disableddeletenotify
DisableDeleteNotify = 0
```

The following example shows enabling delete notification (TRIM).

```
PS C:\> fsutil behavior set disableddeletenotify 0
```

The following example shows disabling delete notification (TRIM).

```
PS C:\> fsutil behavior set disableddeletenotify 1
```

---

**Note:** Disabling delete notifications disables all TRIM commands from the Windows Host.

---

- Hard disk drives and SANs that do not support TRIM, will not receive TRIM notifications.
- TRIM is enabled by default and can be disabled by the administrator.
- Enabling or disabling delete notifications does not require a restart.
- TRIM is effective when the next unmap command is issued.
- Existing inflight I/Os are not impacted by the registry change.

---

**Note:** In Windows Server 2012, when a large file is deleted, the file system performs space reclamation according to the storage array setting. Large file deletions can potentially affect the performance of the regular I/O. To avoid this, delete notification should be disabled. For further details, refer to [https://technet.microsoft.com/en-us/library/jj674351\(v=ws.11\).aspx](https://technet.microsoft.com/en-us/library/jj674351(v=ws.11).aspx).

---



---

**Note:** All files must be permanently deleted (i.e. removed from the Recycle Bin) for the automatic deletion to be performed.

---

- ◆ Manual space reclamation - You can perform manual space reclamation, using the following options:
  - Windows Optimizer (Optimize-Volume cmdlet) - Starting from Windows 2012, Windows introduced the option to reclaim space on a TRIM-enabled array (the delete notifications option must be enabled).

---

**Note:** In versions prior to Windows 2012R2, the entire file system is locked during every file deletion. For further details, contact Microsoft.

---

To run TRIM on a volume, use the following command:

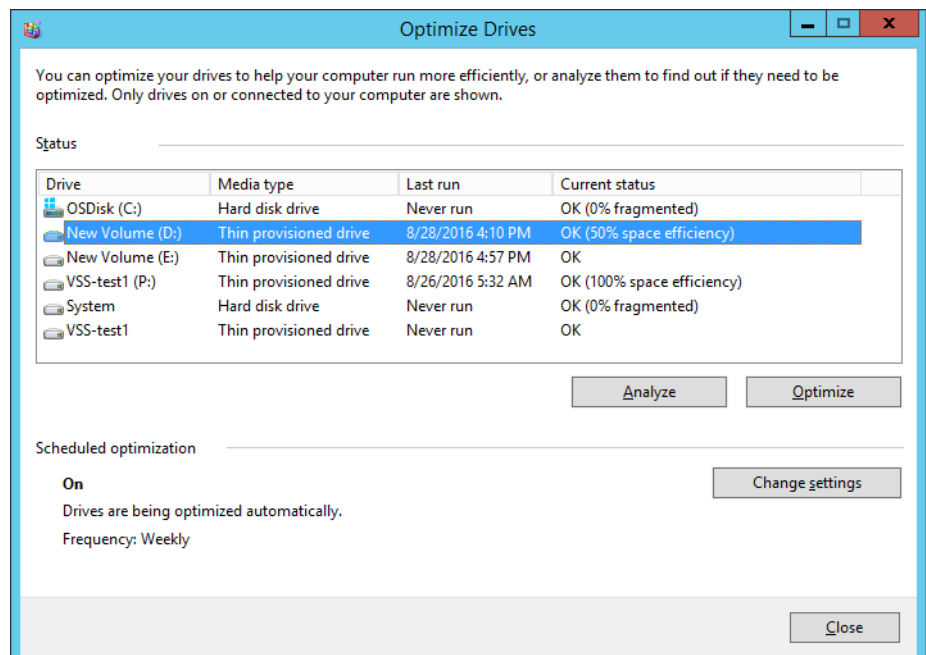
```
Optimize-Volume
```

The following example show running TRIM on volume E.

```
PS C:\> Optimize-Volume -DriveLetter E -ReTrim -Verbose
VERBOSE: Invoking retrim on New Volume (E:)...
VERBOSE: Performing pass 1:
VERBOSE: Retrim: 0% complete...
VERBOSE: Retrim: 10% complete...
VERBOSE: Retrim: 20% complete...
VERBOSE: Retrim: 30% complete...
VERBOSE: Retrim: 40% complete...
VERBOSE: Retrim: 50% complete...
VERBOSE: Retrim: 60% complete...
VERBOSE: Retrim: 70% complete...
VERBOSE: Retrim: 80% complete...
VERBOSE: Retrim: 90% complete...
VERBOSE: Retrim: 100% complete.
VERBOSE:
Post Defragmentation Report:
VERBOSE:
Volume Information:
VERBOSE: Volume size = 1023.99 GB
VERBOSE: Cluster size = 4 KB
VERBOSE: Used space = 207.73 MB
VERBOSE: Free space = 1023.79 GB
VERBOSE:
Retrim:
VERBOSE: Backed allocations = 1023
VERBOSE: Allocations trimmed = 1022
VERBOSE: Total space trimmed = 1020.87 GB
```

When using Windows Optimizer, update Windows, using the following link:  
<https://support.microsoft.com/en-us/kb/2964429>

Another option is running the tool from Windows GUI.



- **SDELETE** utility - Supported starting from Windows 2008. This utility was originally designed to provide an option for securely deleting data on magnetic disks by overwriting on-disk data, using various techniques to ensure that disk data is unrecoverable.

Download the SDELETE utility from the following link:

<https://technet.microsoft.com/en-us/sysinternals/bb897443.aspx>

SDELETE allocates the largest file possible, using non-cached file I/O, to prevent the contents of the NT file system cache from being replaced with useless data, associated with SDELETE's space-hogging file. Since non-cached file I/O must be sector (512-byte) aligned, there may be non-allocated space left even when SDELETE cannot increase the file size. To solve this, SDELETE allocates the largest cache file possible.

The SDELETE option has the following drawbacks:

- Slow performance.
- No control on the balloon file. This may lead to a 100% full file system during the reclamation process, generating an I/O error.

The following example shows running SDELETE on s:\ mount point.

```
PS C:\Users\Administrator> sdelete.exe -z s:
SDelete - Secure Delete v1.61
Copyright (C) 1999-2012 Mark Russinovich
Sysinternals - www.sysinternals.com
SDelete is set for 1 pass.
Free space cleaned on S:\
1 drives zapped
```

---

**Note:** During space reclamation, the drive capacity is reduced to minimum because of the created balloon file.

---

- PowerShell script - Supported on Windows 2008 and above.

The following example shows running PowerShell script on m:\ mount point with 10% free space to allow user writes (refer to [“Windows Powershell Reclaim Script” on page 95](#)).

```
PS C:\Users\Administrator> Write-ZeroesToFreeSpace.ps1 -Root
"m:\" -PercentFree .1
```

If the server is running custom made PS scripts for the first time, refer to [“Example 2”](#) in [“Windows Powershell Reclaim Script” on page 95](#).

## HYPER-V Space Reclamation

- ◆ UNMAP requests from the Hyper-V guest operating system - During the virtual machine (VM) creation, a Hyper-V host inquires whether the storage device, holding the virtual hard disk (VHD) supports UNMAP or TRIM commands. When a large file is deleted from the file system of a VM guest operating system, the guest operating system sends a file delete request to the virtual machine's virtual hard disk (VHD) or VHD file. The VM's VHD or VHD file tunnels the SCSI UNMAP request to the class driver stack of the Windows Hyper-V host.

Microsoft hypervisor passes T10 commands. Therefore, if the guest OS file system supports online space reclamation, no additional task is required.

To run manual space reclamation in the guest OS level, refer to the relevant OS chapter.

---

**Note:** If deletenotification is disabled on the hypervisor level, all guest VMs cannot utilize space reclamation features until deletenotification is re-enabled.

---

## Windows Powershell Reclaim Script

The following example describes a Windows Powershell reclaim script.

```
param(
    [Parameter(Mandatory=$true,ValueFromPipelineByPropertyName=$true)]
    [ValidateNotNullOrEmpty()]
    [Alias("name")]
    $Root,
    [Parameter(Mandatory=$false)]
    [ValidateRange(0,1)]
    $PercentFree = .05
)
process{
    #Convert the #Root value to a valid WMI filter string
    $FixedRoot = ($Root.Trim("\") -replace "\\","\\") + "\\"
    $FileName = "ThinSAN.tmp"
    $FilePath = Join-Path $Root $FileName

    if( (Test-Path $FilePath) ) {
        Write-Error -Message "The file $FilePath already exists, please
delete the file and try again"
    } else {
        $Volume = gwmi win32_volume -filter "name='$FixedRoot'"
        if($Volume) {
            #Change block size to control performance impact
            $ArraySize = 64kb
            $SpaceToLeave = $Volume.Capacity * $PercentFree
            $FileSize = $Volume.FreeSpace - $SpaceToLeave
            $ZeroArray = new-object byte[] ($ArraySize)

            $Stream = [io.File]::OpenWrite($FilePath)
            try {
                $CurFileSize = 0
                while($CurFileSize -lt $FileSize) {
                    $Stream.Write($ZeroArray,0, $ZeroArray.Length)
                    $CurFileSize += $ZeroArray.Length
                }
            } finally {
                #always close our file stream, even if an exception occurred
                if($Stream) {
                    $Stream.Close()
                }
                #always delete the file if we created it, even if an exception
occurred
                if( (Test-Path $filePath) ) {
                    del $FilePath
                }
            }
        } else {
            Write-Error "Unable to locate a volume mounted at $Root"
        }
    }
}
```

In case the error message, as displayed in [“Example 1”](#), appears while running the script for the first time, update the execution policy as displayed in [“Example 2”](#).

◆ Example 1

```
PS C:\Users\Administrator> Write-ZeroesToFreeSpace.ps1 -Root "m:\"
-PercentFree .1
File C:\Windows\system32\Write-ZeroesToFreeSpace.ps1 cannot be
loaded because the execution of scripts is disabled on this system.
Please see "get-help about_signing" for more details.
At line:1 char:28
+ Write-ZeroesToFreeSpace.ps1 <<<< -Root "m:\" -PercentFree .1
+ ~~~~~ CategoryInfo          : NotSpecified: (:) [],
 PSSecurityException
+ ~~~~~ FullyQualifiedErrorId : RuntimeException
```

◆ Example 2

```
PS C:\Users\Administrator> Set-ExecutionPolicy unrestricted
Execution Policy Change
The execution policy helps protect you from scripts that you do not
trust. Changing the execution policy might expose
you to the security risks described in the about_Execution_Policies
help topic. Do you want to change the execution
policy?
[Y] Yes [N] No [S] Suspend [?] Help (default is "Y"): Y
```



## Disabling the UNMAP Feature

When using space reclamation with Windows 2012 or Windows 2012 R2, two problems may arise:

- ◆ Quick formatting of a volume can take a long time to complete since a large number of TRIM commands is sent to the array.
- ◆ Due to a limitation in Windows 2012 (solved in Windows 2012 R2), asynchronous TRIM command locks the entire disk (MS PhysicalDisk) for an I/O until all of its required space is reclaimed.

It is possible to avoid these situations by disabling the UNMAP feature, and then re-enabling it when the relevant operation (i.e. format or delete) is completed.

### To disable the UNMAP feature:

- ◆ Issue the following command to disable the sending of TRIM commands to the array:

```
fsutil behavior set disabledeletemotify 1
```

---

**Note:** This setting applies to all volumes in the host and therefore needs to be re-issued for each disk to format.

---

While UNMAP is disabled, automatic TRIM commands are not sent to XtremIO. Space reclamation can only be carried out manually using the Windows SDelete utility (for details on using the SDelete utility, refer to

<http://technet.microsoft.com/en-us/sysinternals/bb897443.aspx>).

## Configuring the Volume Snapshot Service, Using XtremIO VSS Hardware Provider

VSS (Volume snapshot service) is a set of Component Object Model (COM) application programming interfaces (APIs), that provides standardized interfaces. VSS enables third-party backup and restoration software to centrally manage the backup and restore operations on a variety of applications. VSS also implements a framework that enables volume backups to be performed while applications on the system continue to write to the volumes.

VSS includes the following components:

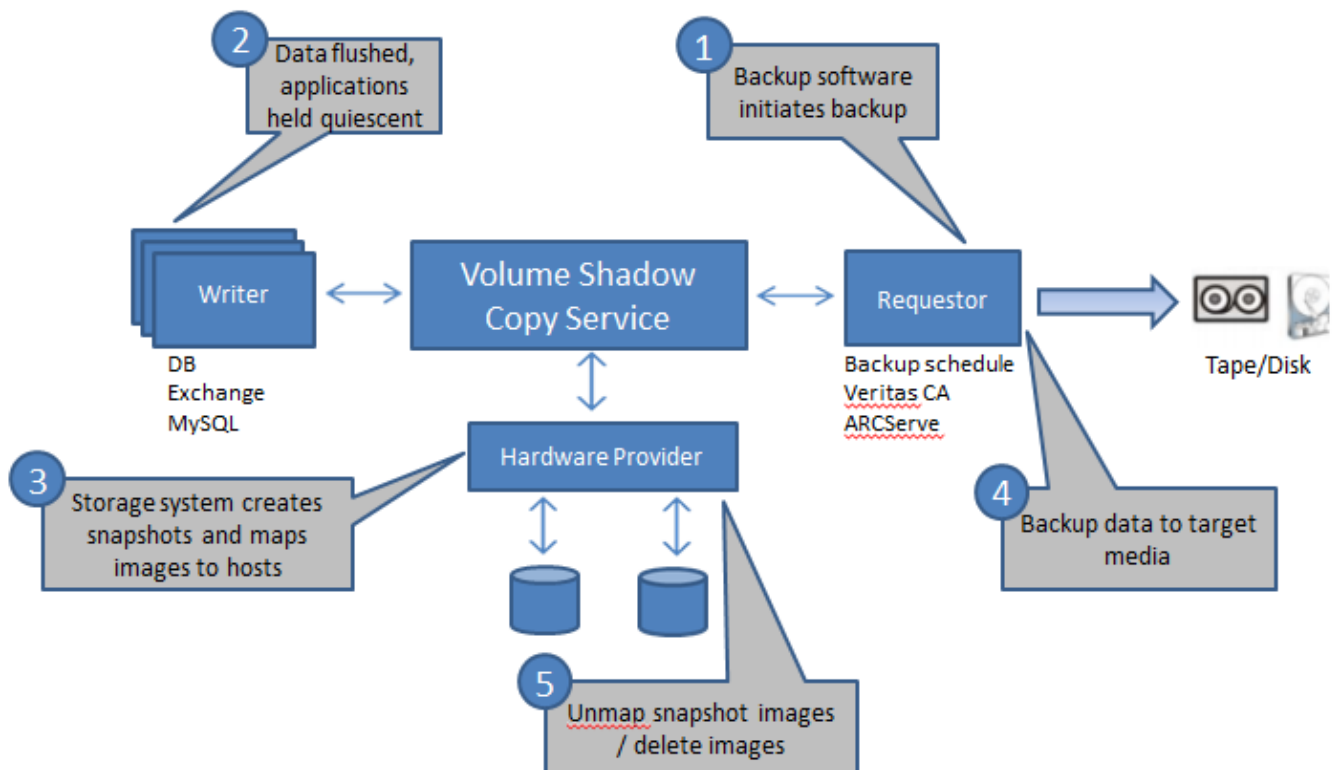
- ◆ **Requestor** - the application that requests the creation of a shadow copy
- ◆ **Provider** - the interface that provides the functionality to perform the shadow copy
- ◆ **Writer** - an application-specific software, that ensures application data is ready for shadow copy creation

The Requestor initiates the backup and restore processes, and the Provider controls the processes and instructs the Writer to prepare a dataset for backup. When these steps are completed, the Requestor instructs the Provider to create a shadow copy.

## Basic VSS Scenario

The following steps describe a basic VSS scenario, in which a backup application (Requestor) requests to back up a volume while the Exchange server (Writer) is up and running:

1. The Requestor interacts with the VSS service to start the backup.
  2. VSS service sends a Freeze request to the Writer. The Writer freezes a stable state of the volume (after closing relevant files, locking additional requests etc.).
- Note:** each Writer provides its own implementation of the freeze procedure.
3. The Requestor queries the Writer Metadata Document xml document to acquire information about files that require backup.
  4. The Provider creates a snapshot of the file-system, maps a new volume with the snapshot content (mapping is performed on-demand when the snapshot is exposed as a volume), and returns a response to the VSS Service.
  5. The VSS service requests the Writer to unfreeze the volume, by sending a Thaw event.
  6. The Requestor uses the snapshot to back up the content while having a read-only/complete state of that volume.



## Requirements for Configuring VSS with XtremIO VSS Hardware Provider

- ◆ The Windows host should be installed with Windows Server 2008 or above.
- ◆ The Windows host should have access to the XMS RESTful API.
- ◆ XtremIO volume(s) should be mapped to the Windows host.
- ◆ XMS access level should be Configuration (or above).
- ◆ XtremIO VSS provider should be installed.
- ◆ VSS service should be up and running.

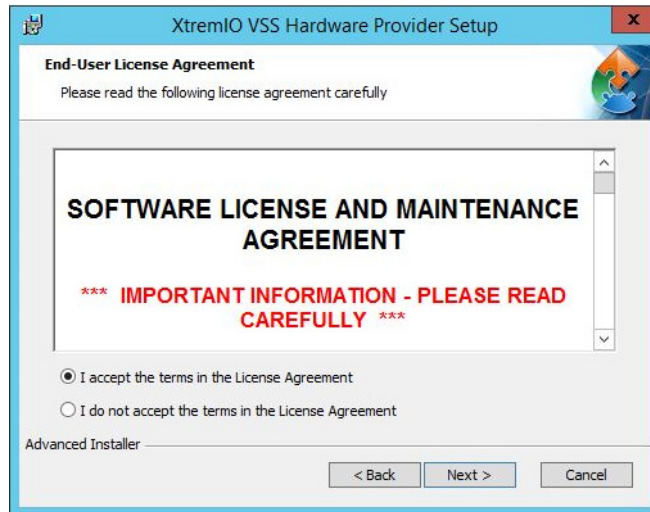
## Installing the VSS Provider on the Windows Host

Perform the following steps to locate the VSS Provider package on the support page and to install the package on the Windows host:

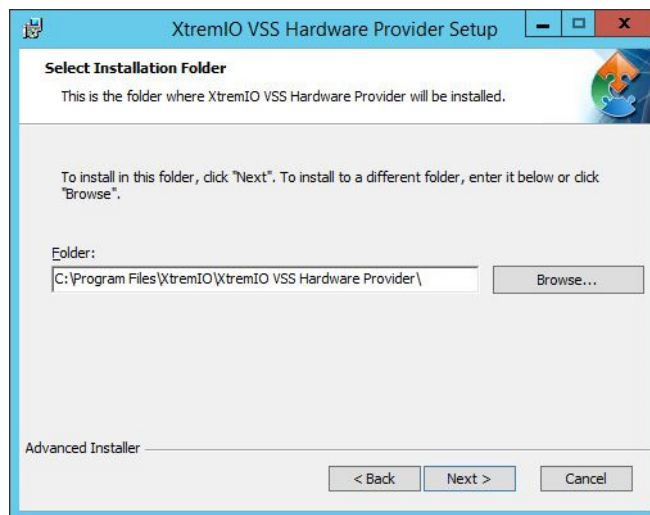
1. Access the EMC Support page for XtremIO in [support.emc.com](http://support.emc.com).
2. Download the XtremIO VSS Provider package to the Windows host. For details on which XtremIO VSS Provider package to download, refer to the Release Notes of the version you are installing.
3. Run the *XtremIOVSSProvider.msi* from an elevated command prompt or PowerShell prompt (i.e. the shell should be run as an administrator).



4. Accept the software license and maintenance agreement and click **Next**.

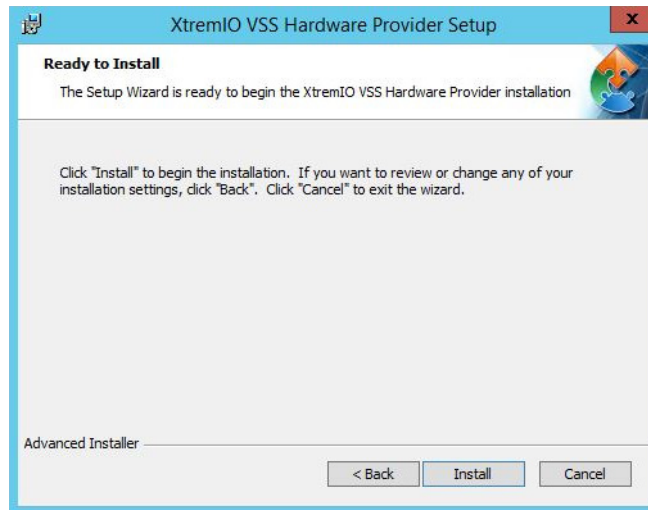


5. Select the installation folder.

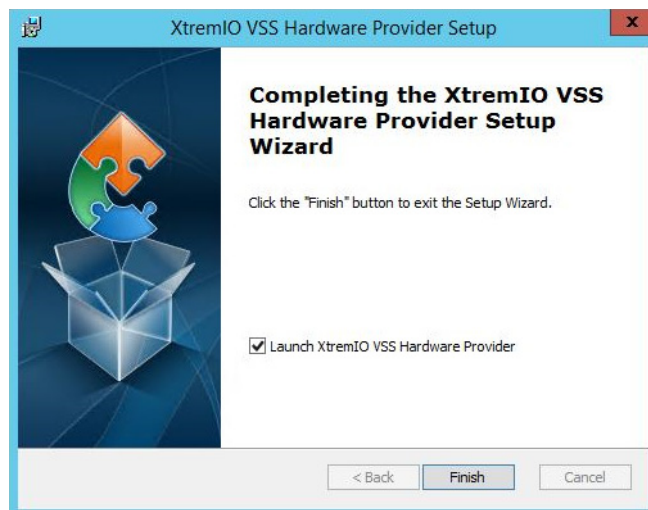


**Note:** The XtremIO Hardware Provider should be installed on the local disk.

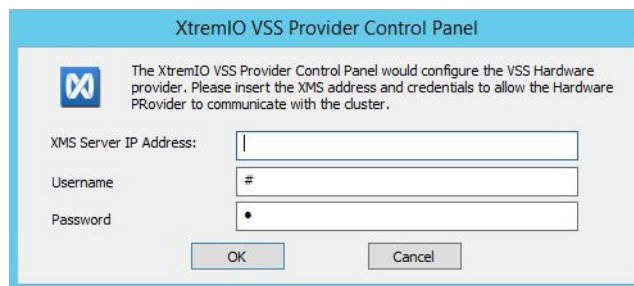
6. Click **Install**.



7. When installation is complete, verify that the **Launch XtremIO VSS Hardware Provider** option is selected, and click **Finish**.



8. In the opened XtremIO VSS Provider Control Panel window, provide the required details.



---

**Note:** Verify that the typed IP address and the credentials are correct before clicking OK.

---

9. Click **OK**; a confirmation message is displayed.



10. In the PowerShell, run the `vssadmin list providers` command and verify that the XtremIO VSS provider appears in the providers list and is properly installed.

**Note:** If XtremIO VSS Hardware Provider fails to successfully register as a COM+ application on the machine it is installed on, it may not function properly.

## XtremIO VSS Provider Components

- ◆ Windows Installer (MSI)
  - Installs and uninstalls the provider service and the control panel Applet.
  - Installs VC2012 Runtime environment if required.
  - Supports silent mode.
- ◆ Control Panel GUI Applet
  - The control panel displays an XtremIO provider icon.
  - Running the applet enables configuring the provider's server address and username/password.
- ◆ VSS Hardware Provider service
  - VSS COM service is registered as a hardware provider and communicates with the VSS service.
  - The XMS communicates with the VSS on the Windows host via RESTful API.

## XtremIO VSS Features

### ◆ VSS Functionality

- [AreLunsSupported](#) -  
Used to identify LUNs managed by XtremIO Storage system (using SCSI inquiry Vendor-id and Product-Id) and request VSS to flag these volumes, to be managed by XtremIO provider.
- [GetTargetLuns](#) -  
Used to provide VSS with the LUN(s) information for the snapshot volumes.
- [LocateLuns](#) -  
Used to map LUNs requested by the VSS service.
- [OnLunEmpty](#) -  
A VSS request used to delete and unmap a LUN.
- [BeginPrepare Snapshot](#) -  
A VSS request used to prepare for snapshotting.
- [FillInLunInfo](#) -  
A VSS request used to confirm and fix LUN(s) structure.
- [CommitSnapshots](#) -  
Used to commit the snapshots, using a naming convention of "[Original volume name]-Year-Month-Day-Hour-Minute-Sec".
- [AbortSnapshots](#) -  
Used to abort snapshot creation.
- [RegisterProvider](#) -  
Used to register the hardware provider.
- [UnregisterProvider](#) -  
Used to unregister the hardware provider.
- [OnLunStateChange](#) -  
A VSS request used to unmap a LUN.

### ◆ Extra VSS Provider Features:

- Multi-clusters support
- Reconfiguration (changing configuration via the control panel)
- Fibre Channel and iSCSI interfaces support
- On Fibre Channel, map to all ports with different Initiator Groups

## Tools

### ◆ Vshadow

- A command-line tool used for creating and managing volume shadow copies.
- Acts as a requestor.
- Provided with Microsoft SDK 7+.

---

**Note:** Verify that you are running the x64 version, by checking the path C:\Program Files (x86)\Microsoft SDKs\Windows\v7.1A\Bin\x64.

---

- Vshadow Syntax examples:
  - Create a persistent shadow copy set from the d: volume (d: must be an XtremIO volume).  
`Vshadow.exe -p d:`
  - Create a persistent shadow copy set from multiple volumes x: y: z: (must be XtremIO volumes).  
`Vshadow.exe -p x: y: z:`
  - Delete a specific shadow copy.  
`Vshadow.exe -ds=ShadowCopyId`
  - Delete all shadow copies.  
`Vshadow.exe -da`
  - Expose a shadow copy to the X: drive (read-only).  
`Vshadow.exe -er=ShadowCopyId,x:`

For more syntax examples, use the following link:

[https://msdn.microsoft.com/en-us/library/windows/desktop/bb530725\(v=vs.85\).aspx](https://msdn.microsoft.com/en-us/library/windows/desktop/bb530725(v=vs.85).aspx)

## Configuring Boot from SAN with an XtremIO Volume

For guidelines on configuring the host to boot from an XtremIO volume, refer to “Configuring Boot from SAN with XtremIO” on page 155.



# CHAPTER 5

## Host Configuration for Oracle® Solaris

This chapter includes the following topics:

- ◆ [Fibre Channel HBA Configuration.....](#) 106
- ◆ [Host Parameters Settings .....](#) 107
- ◆ [Multipathing Software Configuration .....](#) 113
- ◆ [Post-Configuration Steps - Using the XtremIO Cluster.....](#) 116

---

**Note:** In hosts running a hypervisor, such as VMware ESX, Microsoft Hyper-V, or any clustering software, it is important to ensure that the logical unit numbers of XtremIO volumes are consistent across **all** hosts in the hypervisor cluster. Inconsistent LUNs may affect operations such as VM online migration or VM power-up.

---

## Fibre Channel HBA Configuration

When using Fibre Channel with XtremIO, address the following FC Host Bus Adapters (HBA) issues for optimal performance.

### Pre-Requisites

To install one or more EMC-approved HBAs into a Solaris host, follow the procedures in one of these documents, according to the FC HBA type:

- ◆ For Qlogic HBAs -Fibre Channel with QLogic Host Bus Adapters in the Solaris Environment document
- ◆ For Emulex HBAs - *Fibre Channel with Emulex Host Bus Adapters in the Solaris Environment document*

Refer to these documents to configure the host for connection to the EMC Storage Arrays over a Fibre Channel, including any needed HBA BIOS settings.

The documents are available on <http://support.EMC.com>.

### Queue Depth

---

**Note:** Changing the queue depth settings is designed for advanced users. Increasing the queue depth may cause the host to over-stress other arrays connected to the Solaris host, resulting in performance degradation while communicating with them. Therefore, especially in mixed environments with multiple array types connected to the Solaris host, compare the XtremIO recommendation for queue depth with those of other platforms before applying them.

---

Queue depth is the amount of SCSI commands (including I/O requests) that can be handled by a storage device at a given time. A queue depth can be set on either of the following:

- ◆ Initiator level - HBA queue depth
- ◆ LUN level - LUN queue depth

The LUN queue depth setting controls the amount of outstanding I/O requests per a single path.

The HBA queue depth (also referred to as execution throttle) setting controls the amount of outstanding I/O requests per HBA port.

With XtremIO and Solaris, both HBA and LUN queue depth settings should retain their default values.

# Host Parameters Settings

This section describes the Solaris host parameters settings required for optimal configuration when using XtremIO storage.

---

**Note:** The following setting adjustments may cause hosts to over-stress other arrays connected to the Solaris host, resulting in performance degradation while communicating with them. To avoid this, in mixed environments with multiple array types connected to the Solaris host, compare these XtremIO recommendations with those of other platforms before applying them.

---

When using XtremIO storage with Solaris, it is recommended to increase the `Maximum I/O Size` parameter. By default, Solaris breaks large I/Os into 1MB chunks due to disk and kernel limitation on the Maximum I/O size. XtremIO can handle I/Os as large as 4MB. Therefore, increasing the maximum I/O size to the same value will render sequential I/O more efficient.

The SCSI timeout setting of `sd_io_time` (for Solaris x86) or `ssd_io_time` (for Solaris SPARC), defines how long the host waits for a response from the target, before a SCSI command is timed-out. The parameter's default value is 60 seconds. However, it is recommended to hard-set this value with XtremIO, to prevent any change due to Solaris operating system default change.

---

**Note:** It is recommended to backup the configuration file before making any changes.

---

## To adjust the maximum I/O size with XtremIO storage:

1. Add the following Solaris kernel parameter to `/etc/system` file, to increase the I/O size sent from the Solaris host to XtremIO:

```
set maxphys = 0x400000
```

2. Increase the maximum I/O size for the disk driver. The corresponding entry and file for this change are different for different Solaris versions, as described below.

### For Solaris 10 (SPARC):

FILE: `/kernel/drv/ssd.conf`

Setting: `ssd_max_xfer_size=0x400000;`

### For Solaris 10 (x86):

FILE: `/kernel/drv/sd.conf`

Setting: `sd_max_xfer_size=0x400000;`

### For Solaris 11 (SPARC):

FILE: `/etc/driver/drv/ssd.conf`

Setting: `ssd_max_xfer_size=0x400000;`

### For Solaris 11 (x86):

FILE: `/etc/driver/drv/sd.conf`

Setting: `sd_max_xfer_size=0x400000;`

**To configure timeout in /etc/system:**

1. Run the following command to check the current SCSI timeout setting:
  - For Solaris x86:
 

```
#grep sd:sd_io_time /etc/system
```
  - For Solaris Sparc:
 

```
#grep ssd:ssd_io_time /etc/system
```
2. If timeout is not defined, run the following command to create a backup of the current /etc/system file:
 

```
#cp -p /etc/system /etc/system_ORIG
```
3. Edit the /etc/system file by appending or modifying the SCSI timeout settings as follows:
  - For Solaris x86:
 

```
set sd:sd_io_time=0x3C
```
  - For Solaris Sparc host:
 

```
set ssd:ssd_io_time=0x3C
```

**Configuring scsi\_vhci.conf for Solaris 11**

The `scsi_vhci.conf` file is used to configure third-party symmetric storage on the Solaris 11 host. The host sends SCSI inquiry commands and needs to know what strings to expect in return.

**To configure /etc/driver/drv/scsi\_vhci.conf file for a Solaris 11 host:**

1. Run the following command to verify the `scsi_vhci.conf` file location:
 

```
#ls /etc/driver/drv/
```
2. If the file is not in the expected location, run the following command to copy it from /kernel/drv:
 

```
#cp /kernel/drv/scsi_vhci.conf /etc/driver/drv
```
3. Run the following commands to create a backup and modify:
 

```
#cp -p /etc/driver/drv/scsi_vhci.conf  
/etc/driver/drv/scsi_vhci.conf_ORIG  
#vi /etc/driver/drv/scsi_vhci.conf
```

4. Append the following line to the file:

```
scsi-vhci-failover-override = "XtremIO XtremApp",
"f_sym_emc";
```

The following list explains the different fields in the appended line:

- `scsi-vhci-failover-override` - The override field needs to be set since the default 'probe' output is not used.
- `XtremIO XtremApp` - The VID PID (Vendor ID/Product ID) returned by the storage to a SCSI inquiry command.
- `f_sym_emc` - Indicates EMC symmetric (Non Asymmetric) multipathing/arrays.

5. Save the `scsi_vhci.conf` file.

6. Proceed as follows:

- If native MPxIO is not used on the Solaris host, no further action is required.
- If native MPxIO is used on the Solaris host, proceed to the next step.

7. Review the `fp.conf` file and set the value of the `mpxio-disable` parameter as follows:

- File location:

```
/etc/driver/drv/fp.conf
```

- `mpxio-disable="no";`

8. Run the following command to boot the system and enable STMS:

```
stmsboot -e
```

9. Run the following command to verify that STMS is properly configured:

```
luxadm probe
```

## Configuring scsi\_vhci.conf for Solaris 10

**To configure `/kernel/drv/scsi_vhci.conf` file for a Solaris 10 host:**

1. Run the following commands to create a backup and modify:

```
cp /kernel/drv/scsi_vhci.conf
  /kernel/drv/scsi_vhci.conf_ORIG
vi /kernel/drv/scsi_vhci.conf
```

2. Append the following lines to the file:

```
device-type-scsi-options-list = "XtremIO XtremApp",
"symmetric-option";

symmetric-option = 0x1000000;
```

3. Save the `scsi_vhci.conf` file.

4. Proceed as follows:

- If native MPxIO is not used on the Solaris host, no further action is required.
- If native MPxIO is used on the Solaris host, proceed to the next step.

5. Review the `fp.conf` file and set the value of the `mpxio-disable` parameter as follows:

- File location:  
`/kernel/drv/fp.conf`
- `mpxio-disable="no";`

6. Run the following command to boot the system and enable STMS:

```
stmsboot -e
```

7. Run the following command to verify that STMS is properly configured:

```
stmsboot -L
```

## Configuring Initiator-Target SCSI Communication

**Note:** When using some layered clustering products, the default of 512 bytes physical-block-size, as defined in the `sd.conf` or `ssd.conf` file, is required.

The `sd.conf` or `ssd.conf` file is used for controlling the SCSI communication between a target and an initiator. To configure the initiator-target SCSI communication, the following single-line configuration entry is used:

```
sd-config-list = "XtremIO XtremApp", "throttle-max:64,
delay-busy:30000000000, retries-busy:90, retries-timeout:30,
physical-block-size:8192, retries-notready:30";
```

The following list explains the different fields in the configuration entry:

- ◆ `XtremIO XtremApp` - The VID PID (Vendor ID/Product ID) returned by the storage to a SCSI inquiry command.
- ◆ `throttle-max:64` - The maximum number of retries to issue before giving up on the command.
- ◆ `delay-busy:30000000000` - The amount of time to wait upon receiving a 'Busy' response from the target.
- ◆ `retries-busy:90` - The number of retries to issue before giving up on the command.
- ◆ `retries-timeout:30` - The amount of time to wait for a retry to complete.
- ◆ `physical-block-size:8192` - The block size used on the storage system. For XtremIO versions 3.x (or above), use 8192. Otherwise, use 4096.

**Note:** This setting is only relevant for Solaris 11.

- ◆ `retries-notready:30` - Upon receiving a 'not ready' response from the target, the host delays the next retry to avoid a rapid exhaustion of retries.

**Note:** Make sure that the command is written as a single line.

**Note:** XtremIO does not recommend using 4KB XtremIO volumes with Solaris. Further testing is required to support this configuration.

**Note:** Using a physical block size other than the default 512 bytes can result in incompatibilities when using some layered clustering products. Using the default 512 byte physical block size, as defined in the `sd/ssd.conf` file, alleviates this problem.

**To configure `sd.conf` or `ssd.conf` file:**

**Note:** Steps 1-3 only apply to Solaris 11. For Solaris 10, skip to [step 4](#).

1. Run the following command to check if the `sd.conf` file exists
 

```
#ls /etc/driver/drv/sd.conf
```
2. If the file is not in the expected location, run the following command to copy it from `/kernel/drv`:
 

```
#cp /kernel/drv/sd.conf /etc/driver/drv
```
3. Run the following commands to create a backup copy:
 

```
#cp /etc/driver/drv/sd.conf /etc/driver/drv/sd.conf_back
```
4. Modify the appropriate `sd.conf` or `ssd.conf` file (according to Solaris OS version and Solaris architecture on the host), and append the appropriate entry:
  - Solaris 10:
    - For a Solaris x86 host:
 

Append the following line to the `/kernel/drv/sd.conf` file:

```
sd-config-list = "XtremIO XtremApp", "throttle-max:64,
delay-busy:30000000000, retries-busy:90,
retries-timeout:30, retries-notready:30";
```
    - For a SPARC host:
 

Append the following line to the `/kernel/drv/ssd.conf` file:

```
ssd-config-list = "XtremIO XtremApp", "throttle-max:64,
delay-busy:30000000000, retries-busy:90,
retries-timeout:30, retries-notready:30";
```
  - Solaris 11:
    - For a Solaris x86 host:
 

Append the following line to the `/etc/driver/drv/sd.conf` file:

```
sd-config-list = "XtremIO XtremApp", "throttle-max:64,
delay-busy:30000000000, retries-busy:90,
retries-timeout:30, physical-block-size:8192,
retries-notready:30";
```
    - For a SPARC host:
 

Append the following line to the `/etc/driver/drv/ssd.conf` file:

```
ssd-config-list = "XtremIO XtremApp", "throttle-max:64,
delay-busy:30000000000, retries-busy:90,
retries-timeout:30, physical-block-size:8192,
retries-notready:30";
```

**Note:** Following these modifications, the Solaris host requires a reboot.



# Multipathing Software Configuration

This section describes the necessary steps for configuring multipathing on a Solaris host. Implementing these steps enables the host to continuously run I/Os even if a SCSI target goes off line due to NDU or HA.

## Configuring STMS to Manage MultiPath under Solaris

STMS (formerly known as MPxIO) is the built-in Path Management software, used by Solaris and supported by Oracle. This section describes the required steps for configuring STMS on Solaris 11 and Solaris 10 hosts.

---

**Note:** Use the STMS devices for all application configurations.

---

### Configuring MPxIO for XtremIO Devices When Another Storage Array is Configured for MPxIO on the Solaris Host

This section provides directions on how to configure MPxIO for XtremIO devices, when another storage array is already configured for MPxIO in the `scsi_vhci.conf` file of the Solaris host.

Oracle always adds support for new storage products to the MPxIO support list. Therefore, before adding storage manually in the `scsi_vhci.conf` file, it is important to check if the particular storage type already appears on the supported list.

**To check if a particular storage type appears on the Solaris supported list:**

- ◆ run the following command:

```
# mpathadm show mpath-support libmpscsi_vhci.so
```

Example:

```
# mpathadm show mpath-support libmpscsi_vhci.so
(...)
Vendor:    DGC
Product:   RAID
Revision:
Supported Load Balance Types:
           round-robin
(...)
```

In the example, support for EMC VNX storage is already added to the installed version of MPxIO.

If the storage to be added is not on the support list, you need to manually define a multi-pathing policy for the storage array in the `scsi_vhci.conf` file. The used syntax is slightly different for Solaris 10 and 11.

---

**Note:** The following examples assume that EMC VMAX storage is already used on the Solaris server and an XtremIO storage needs to be added to the MPxIO configuration.

---

**Note:** The following applies to both Solaris 10 and 11:

- ◆ The first value in the `scsi_vhci.conf` file entries, VID (Vendor ID), must occupy the first eight characters. If it is shorter than eight characters, the missing characters should be filled with spaces. For example, there should be five spaces between EMC and SYMMETRIX, and a single space between XtremIO and XtremApp.
- ◆ The second value in the file entries, PID (Product ID), can have up to sixteen characters, reported as the response to SCSI INQ command.
- ◆ Using an '=' sign in entries can be done with or without spaces around it.
- ◆ The `scsi_vhci.conf` file must not contain tabs or other white spaces other than the 'space' used to separate Vendor ID from Product ID.
- ◆ **For Solaris 10:**

The existing EMC VMAX entry in `/kernel/drv/scsi_vhci.conf`:

```
device-type-scsi-options-list =
"EMC      SYMMETRIX", "symmetric-option";
symmetric-option = 0x1000000;
```

The entry in `/kernel/drv/scsi_vhci.conf` after adding information for the XtremIO storage:

```
device-type-scsi-options-list =
"EMC      SYMMETRIX", "symmetric-option",
"XtremIO XtremApp", "symmetric-option";
symmetric-option = 0x1000000;
```

- ◆ **For Solaris 11:**

The existing EMC VMAX entry in `/etc/driver/drv/scsi_vhci.conf`:

```
scsi-vhci-failover-override =
"EMC      SYMMETRIX", "f_sym_emc";
```

The entry in `/etc/driver/drv/scsi_vhci.conf` after adding information for the XtremIO storage:

```
scsi-vhci-failover-override =
"EMC      SYMMETRIX", "f_sym_emc",
"XtremIO XtremApp", "f_sym_emc";
```

---

**Note:** There should be only one variable, 'scsi-vhci-failover-override', defined in `/etc/driver/drv/scsi_vhci.conf` file. If there are multiple variables, only the first will be used.

---

## Configuring PowerPath Multipathing

**Note:** For the most updated information on PowerPath support with XtremIO storage, refer to the *XtremIO Simple Support Matrix*.

XtremIO supports multipathing using EMC PowerPath on Solaris. PowerPath versions 6.0 and above provide Native Loadable Array Module (LAM) for XtremIO Array devices. With this support, XtremIO devices running versions 2.4 and above are managed under the XtremIO class.

PowerPath provides enhanced path management capabilities for up to 32 paths per logical device, as well as intelligent dynamic I/O load-balancing functionalities. Having multiple paths enables the host to access a storage device even if a specific path is unavailable. Multiple paths share the I/O traffic to a storage device, using intelligent load-balancing policies which enhance I/O performance and increase application availability. EMC PowerPath is the recommended multipathing choice.

PowerPath features include:

- ◆ Multiple paths - enable higher availability and I/O performance.
  - Includes control domain in Solaris LDOM.
- ◆ Path management insight capabilities - PowerPath characterizes I/O patterns and aides in diagnosing I/O problems due to flaky paths or unexpected latency values. Metrics are provided on:
  - Read and write - in MB/seconds per LUN
  - Latency distribution - the high and low watermarks per path
  - Retries - the number of failed I/Os on a specific path
- ◆ Autostandby feature - automatically detects intermittent I/O failures and places paths into autostandby (also known as flaky paths).
- ◆ PowerPath Migration Enabler - is a host-based migration tool that allows migrating data between storage systems. PowerPath Migration Enabler works in conjunction with the host operating system (also called Host Copy) and other underlying technologies, such as Open Replicator (OR).
- ◆ Remote monitoring:
  - PowerPath Virtual Appliance 2.0 (vApp 2.0)
  - SNMP management daemon

Further PowerPath related information:

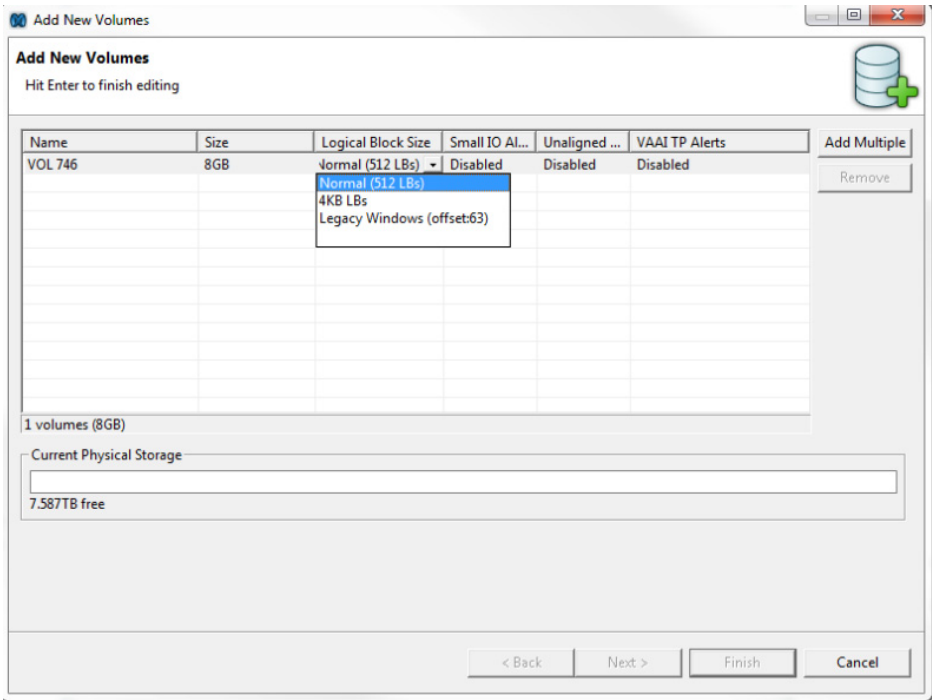
- ◆ For details on the PowerPath releases supported for your Solaris host, refer to the *XtremIO Simple Support Matrix*.
- ◆ For details on native class support with XtremIO for your host, refer to the *EMC PowerPath release notes* document for the PowerPath version you are installing.
- ◆ For details on installing and configuring PowerPath with XtremIO native class support on your host, refer to the *EMC PowerPath on Solaris Installation and Administration Guide* for the PowerPath version you are installing. This guide provides the required information for placing XtremIO volumes under PowerPath control.

# Post-Configuration Steps - Using the XtremIO Cluster

After the host configuration is completed, you can use the XtremIO storage from the host. For details on creating, presenting and managing volumes accessed from the host via either the GUI or the CLI, refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

## Disk Formatting

When creating volumes in XtremIO for a Solaris host, the logical block (LB) size of a new XtremIO volume is 512B. This parameter can be adjusted to 4KB.



**Note:** In XtremIO version 4.0.0 (and above), the Legacy Windows option is not supported.

A 512B logical block size should be used for a new XtremIO volume.

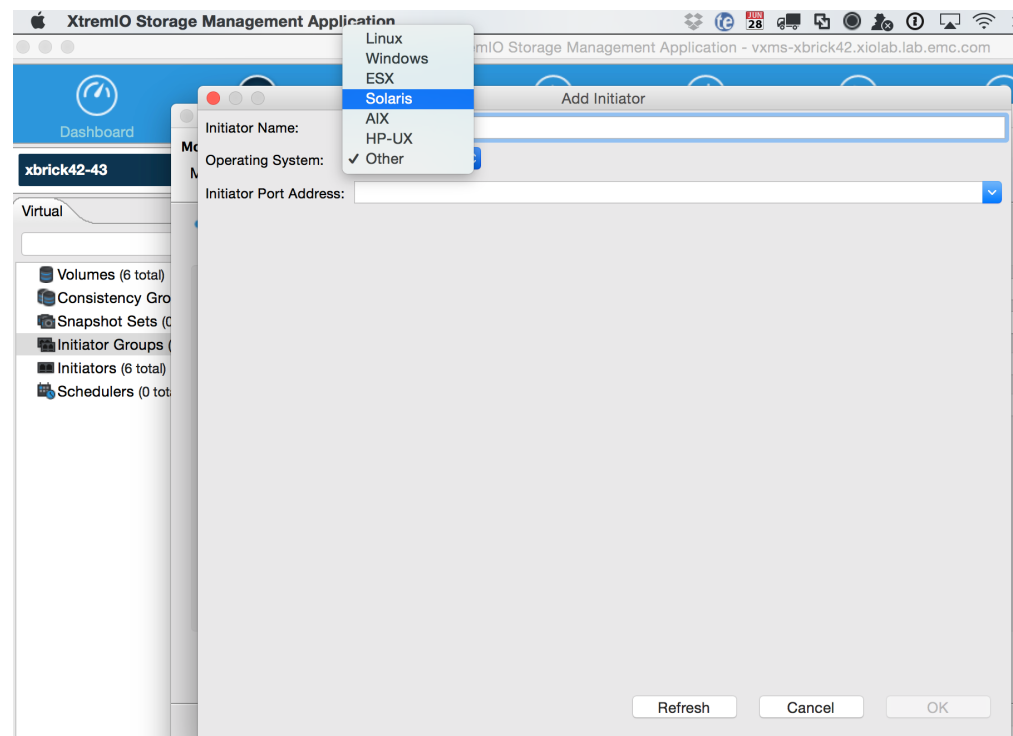
**Note:** The usage of 4KB logical block size is not recommended for new XtremIO volumes with Solaris. UFS file system does not support a 4KB block size.

## Presenting XtremIO Volumes to the Host

**Note:** It is recommended to avoid using host-level software for deduplication, compression and encryption of data stored in XtremIO. These operations are performed natively on the XtremIO cluster, thus releasing host computing resources. Furthermore, using these operations may severely impact data efficiency on the XtremIO cluster, as they can lead to a higher physical or logical capacity utilization.

**Note:** The following information in this section applies only to XtremIO version 4.0 and above.

When adding Initiator Groups and Initiators to allow Solaris hosts to access XtremIO volumes, specify **Solaris** as the operating system for the newly-created Initiators, as described in the figure below.



**Note:** Setting the Initiator's Operating System is required for optimal interoperability and stability of the host with XtremIO storage. The setting enhances SCSI target responses to the host for Solaris multipathing. You can adjust the setting while the host is online and connected to the XtremIO cluster with no I/O impact.

**Note:** Refer to the *XtremIO User Guide* for details.

When adding an initiator via CLI, specify the `solaris` as the `operating-system` parameter.

Specifying the operating system ensures that the host can maintain access to the XtremIO volumes during various cluster high-availability events.

When upgrading a cluster from XtremIO version 3.0.x to version 4.0 (or higher), make sure to modify the operating system for each initiator connected to a Solaris host.

## Overriding Disk Block Size Using `ashift` and Solaris 11.2

---

**Note:** This section applies only to Solaris 11.2 and ZFS file system. For other Solaris versions or other file systems, skip this section.

---

To align the OS LUN physical sector size with ZFS logical structure, the `ashift` value should be configured in a Solaris host that is connected to an XtremIO cluster and is running version 3.0.0-44 (or later). The default configuration of the physical block size (512B) results in high latency and very low dedup ratios.

The following procedures override the disk block size for physical (non LDOM) and virtual (LDOM), respectively.

### To reconfigure the `ashift` value on physical (non LDOM) Solaris 11.2:

1. Run one of the following commands to identify the `ashift` value on `ssd/sd` driver level:

- For Solaris x86 11.2:

```
echo ::sd_state | mdb -k | egrep ' (^un|_blocksize) '
```

- For Solaris SPARC 11.2:

```
echo ::ssd_state | mdb -k | egrep ' (^un|_blocksize) '
```

The command result is one of the following:

- `un_phy_blocksize = 0x2000` [for `physical-block-size:8192`]
- `un_phy_blocksize = 0x1000` [for `physical-block-size:4096`]
- `un_phy_blocksize = 0x200` [for default of `physical-block-size:512`]

2. Run the following command to identify the `ashift` value on ZPOOL level:

```
zdb | egrep 'ashift| name'
```

The command result is one of the following:

- `ashift: 13` [for `physical-block-size:8192`]
- `ashift: 12` [for `physical-block-size:4096`]
- `ashift: 9` [for default of `physical-block-size:512`]

3. Edit the Solaris sd driver configuration file, `/etc/driver/drv/sd.conf`, to define the physical block size to 8KB.

---

**Note:** For Sparc hosts (ssd driver), edit the `/etc/driver/drv/ssd.conf` ssd driver configuration file.

---

In this file, add the following parameter:

- For Solaris x86 11.2:

```
sd-config-list = "XtremIO XtremApp", "throttle-max:64,
delay-busy:30000000000, retries-busy:90,
retries-timeout:30, physical-block-size:8192,
retries-notready:30";
```

- For Solaris SPARC 11.2:

```
ssd-config-list = "XtremIO XtremApp", "throttle-max:64,
delay-busy:30000000000, retries-busy:90,
retries-timeout:30, physical-block-size:8192,
retries-notready:30";
```

4. Reconfigure (reload) the sd driver on device level as follows:

- Run the following command to identify the XtremIO devices:

```
cfgadm -al
```

- Run the following command to re-configure the ssd/sd driver on device level for each of the XtremIO devices:

```
for i in `cfgadm -al | grep c*::210 | awk '{print $1}'`; do
cfgadm -c unconfigure $i; cfgadm -c configure $i; done
```

---

**Note:** In the above command, `c*::210` is the prefix of XtremIO device paths. To run this command only on XtremIO volumes, replace the ``c*::210`` part of the command according to the output of the ``cfgadm -al`` command.

---

If needed, re-run this command for all XtremIO volumes device paths.

5. Reconfigure (reload) the ssd/sd driver on the entire system level:

- For Solaris x86 11.2:

```
update_drv -vf sd
```

- For Solaris SPARC 11.2:

```
update_drv -vf ssd
```

6. Run the following command to verify that the new settings are applied:

```
echo ::sd_state | mdb -k | egrep ' (^un|_blocksize) '
```

---

**Note:** For Sparc hosts (ssd driver), replace `sd_state` with `ssd_state`.

---

7. If the `ashift` value reported did not change, reboot the host with a non-fast reboot as follows:

- For Solaris x86 11.2:

```
reboot -p
```

- For Solaris SPARC 11.2:

```
touch /reconfigure; reboot -- -r
```

8. Verify that OS and ZFS are now using the new disk block size values, as follows:

1. Run the following command to verify the `un_phy_blocksize` parameter value:

- For Solaris x86 11.2:

```
echo ::sd_state | mdb -k | egrep ' (^un|_blocksize) '
```

- For Solaris SPARC 11.2

```
echo ::ssd_state | mdb -k | egrep ' (^un|_blocksize) '
```

The reported `un_phy_blocksize` value should be 2000 for physical-block-size:8192.

2. Run the following command, to verify the `ashift` parameter value:

```
zdb | egrep 'ashift| name'
```

The reported `ashift` value should be 13 for physical-block-size:8192.

#### To reconfigure the `ashift` value on virtual (LDM) Solaris 11.2:

1. Create a vdisk, based on a whole XtremIO volume, and add it to the guest LDM.
2. Configure the physical block size of the vdisk in `vdc.conf`. Add the following entry in `/platform/sun4v/kernel/drv/vdc.conf`, to set the block size to 8K for vdc instance 1:

```
block-size-list="1: 8192";
```

3. Create the zpool in the guest ldom.
4. Verify the `ashift` value as follows:

```
zdb | egrep 'ashift| name'
```

The reported `un_phy_blocksize` should be 2000 for block-size-list=8192.



## Logical Volume Manager (LVM)

---

**Note:** Logical Volume Manager (LVM) configuration and management are out of the scope of this document.

---

When configuring Solaris Volume Manager with a Solaris host connected to XtremIO storage, the following is recommended:

- ◆ LVM stripe size - The LVM stripe size should be set to a value greater than that of the application block size. XtremIO storage supports physical extent size of 8KB and above. the recommended physical extent size is 1MB or above.

## Creating a File System

---

**Note:** File system configuration and management are out of the scope of this document.

---

It is recommended to create the file system using its default block size (using a non-default block size may lead to unexpected behavior). Refer to your operating system and file system documentation.

---

**Note:** Some file systems (such as ZFS in Solaris) may require you to properly align the file system on the XtremIO volume. It is recommended to use specified tools (such as the ZFS configuration) to optimally match your host and application requirements.

---

## Configuring Boot from SAN with an XtremIO Volume

For guidelines on configuring the host to boot from an XtremIO volume, refer to [“Configuring Boot from SAN with XtremIO” on page 155](#).



# CHAPTER 6

## Host Configuration for AIX

This chapter includes the following topics:

- ◆ [Fibre Channel HBA Configuration.....](#) 124
- ◆ [EMC AIX ODM Installation Requirements and Procedure .....](#) 128
- ◆ [Multipathing Software Configuration .....](#) 131
- ◆ [Post-Configuration Steps - Using the XtremIO Cluster.....](#) 139

---

**Note:** AIX is supported with XtremIO version 2.4 (or above).

---

---

**Note:** In hosts running VIOS with NPIV, it is important to verify that the settings of HBA tunable values, `num_cmd_elems=2048`, `fc_err_recov=fast_fail` and `dyntrk=yes` are the same on both the VIOS and VIOC.

---

## Fibre Channel HBA Configuration

When using Fibre Channel with XtremIO, the FC Host Bus Adapters (HBA) issues described in this section should be addressed for optimal performance.

### Queue Depth

---

**Note:** Changing queue depth settings is designed for advanced users. Increasing the queue depth may cause the host to over-stress other arrays connected to the AIX host, resulting in performance degradation while communicating with them. Therefore, especially in mixed environments with multiple array types connected to the AIX host, compare the XtremIO recommendations for queue depth with those of other platforms before applying them.

---

---

**Note:** For further details on setting queue depth in AIX, refer to the *AIX/VIOS Disk and Adapter IO Queue Tuning* document (<http://www-03.ibm.com/support/techdocs/atsmastr.nsf/WebIndex/TD105745>).

---

Queue depth is the amount of SCSI commands (including I/O requests) that can be handled by a storage device at a given time. Queue depth can be set on either of the following:

- ◆ Initiator level - HBA queue depth
- ◆ LUN level - LUN queue depth

The LUN queue depth setting controls the amount of outstanding I/O requests per a single path.

The HBA queue depth (also referred to as execution throttle) setting controls the amount of outstanding I/O requests per HBA port.

For optimal operation with XtremIO and AIX, it is recommended to adjust the HBA queue depth setting of the FC HBA.

The LUN queue depth settings are controlled by the driver module for the card at the OS level. The LUN queue depth default value (256) should be set to a lower value only if I/O throttling is required.

## Setting the Queue Depth

### To set the HBA queue depth:

1. Run the following `chdev` command for each HBA in the AIX host to set the HBA firmware level queue depth:

```
chdev -l fcs# -a num_cmd_elems=2048 -P
```

---

**Note:** Replace `fcs#` with the HBA number (e.g. `fcs0`, `fcs1`, etc.).

---

**Note:** As of AIX versions 6.1 TL 09, 7.1 TL 04, and 7.2, the `num_cmd_elems` attribute value of VFC client adapter is increased to 2048. For details, refer to the following IBM KB articles:

<http://www-01.ibm.com/support/docview.wss?uid=isg1IV76258>

<http://www-01.ibm.com/support/docview.wss?uid=isg1IV76270>

---

2. Reboot the AIX host to apply the HBA queue depth settings.

### To set the LUN queue depth:

- ◆ Run the `chdev` command for each XtremIO volume presented by PowerPath to the AIX host (called `hdiskpower`):

```
chdev -l hdiskpower# -a queue_depth=256
```

---

**Note:** Replace `hdiskpower#` with the volume number (e.g. `hdiskpower0`, `hdiskpower1`, etc.) or, when using MPIO, replace it with the `hdisk` number (e.g. `hdisk10`).

---

## Fast I/O Failure for Fibre Channel Devices

AIX supports Fast I/O Failure for Fibre Channel devices after link events in a switched environment.

When the FC adapter driver detects a link event, such as a lost link between a storage device and a switch, it waits for the fabric to stabilize (approximately 15 seconds). If the device is not on the fabric, the FC adapter driver fails all new I/Os or future retries of the failed I/Os until the device rejoins the fabric.

Fast I/O Failure is controlled by the `fscsi` device attribute, `fc_err_recov` (default value is `delayed_fail`). To enable Fast I/O Failure for all `fscsi` devices, set this attribute to `fast_fail`, as shown in the following example:

---

**Note:** In the example, the `fscsi` device instance is `fscsi0`.

---

```
chdev -l fscsi0 -a fc_err_recov=fast_fail -P
```

To verify that the setting was enabled in the ODM, run the following command:

```
lasttr -El fscsi0
```

---

**Note:** The “-P” flag only modifies the setting in the ODM and requires a system reboot for the changes to apply.

---

Fast fail logic is applied when the switch sends a Registered State Change Notification (RSCN) to the adapter driver, indicating a link event with a remote storage device port.

Fast I/O Failure is useful when multipathing software is used. Setting the `fc_err_recov` attribute to `fast_fail` can decrease I/O failure due to link loss between the storage device and switch by supporting faster failover to alternate paths.

Earlier AIX versions required the user to de-configure the FC storage device and adapter device instances before modifying the system area network (SAN) settings that may result in an N\_Port ID (SCSI ID) change of any remote storage ports.

If dynamic tracking of FC devices is enabled, the FC adapter driver detects when the Fibre Channel N\_Port ID of a device changes. The FC adapter driver then reroutes the traffic that is destined for that device to the new address, while the devices are still online.

Events that can cause an N\_Port ID to change include the following:

- ◆ Moving a cable that connects a switch to a storage device from one switch port to another
- ◆ Connecting two separate switches using an inter-switch link (ISL)
- ◆ Rebooting a switch

Dynamic tracking of FC devices is controlled by a `fscsi` device attribute, `dyntrk` (default value is `no` for non-NPIV configurations). To enable dynamic tracking of FC devices, change all `fscsi` device attributes to `dyntrk=yes`, as shown in the example:

---

**Note:** In this example, the `fscsi` device instance is `fscsi0`.

---

```
chdev -l fscsi0 -a dyntrk=yes -P
```

Run the following command to verify that the setting was enabled in the ODM:

```
lasttr -El fscsi0
```

---

**Note:** The “-P” flag only modifies the setting in the ODM and requires a system reboot for the changes to apply.

---

Dynamic tracking logic is applied when the adapter driver receives an indication from the switch that a link event with a remote storage device port has occurred.

## Fibre Channel Adapter Device Driver Maximum I/O Size

The `max_xfer_size` FC HBA adapter device driver attribute for the `fscsi` device controls the maximum I/O size the adapter device driver can handle. This attribute also controls a memory area used by the adapter for data transfers.

### For optimal AIX host operation over FC with XtremIO:

1. Run the following command on all FC adapters connected to XtremIO:

```
chdev -l fcs0 -a max_xfer_size=0x200000 -P
```

2. Reboot the AIX host to apply the `max_xfer_size` setting adjustments.

---

**Note:** For virtualized AIX hosts, make sure to apply the `max_xfer_size` setting adjustments on all LPARs of the host connected to XtremIO storage.

---

For further details on this setting with AIX, refer to the AIX/VIOS Disk and Adapter IO Queue Tuning paper from IBM at:

[https://www-03.ibm.com/support/techdocs/atsmastr.nsf/5cb5ed706d254a8186256c71006d2e0a/d1f54f4cd1431d5a8625785000529663/\\$FILE/AIX-VIOS\\_DiskAndAdapterQueueTuningV1.2.pdf](https://www-03.ibm.com/support/techdocs/atsmastr.nsf/5cb5ed706d254a8186256c71006d2e0a/d1f54f4cd1431d5a8625785000529663/$FILE/AIX-VIOS_DiskAndAdapterQueueTuningV1.2.pdf)

## EMC AIX ODM Installation Requirements and Procedure

The Object Data Manager (ODM) is a database of system and device configuration information, integrated into the AIX operating system. Information is stored and maintained as objects with associated characteristics. The EMC ODM support package contains a series of installable filesets that are used to update the AIX ODM with customized EMC storage device configuration attributes.

To meet the EMC storage array requirements, the correct EMC ODM filesets need to be installed so as to support Fibre Channel attachment to the XtremIO array. This section outlines these requirements.

The minimum ODM and AIX operating system versions that are supported with PowerPath are:

- ◆ EMC.AIX.6.0.0.4.tar.Z => For AIX 7.2 and 7.1 operating system
- ◆ EMC.AIX.5.3.0.9.tar.Z => For AIX 6.1 operating system

The minimum ODM and AIX operating system versions that are supported with AIX native MPIO are:

- ◆ EMC.AIX.6.0.0.5.tar.Z => For AIX 7.2 operating system version 7.2 SP1 (7200-00-01)
- ◆ EMC.AIX.6.0.0.5.tar.Z => For AIX 7.1 operating system version 7.1 TL3 SP5 (7100-03-05-1524).
- ◆ EMC.AIX.5.3.1.0.tar.Z => For AIX 6.1 operating system version 6.1 TL9 SP5 (6100-09-05-1524).



**To install the EMC XtremIO fileset:**

1. Log in to the AIX system and verify the operating system version number.

```
> oslevel -s
6100-09-01-1341
>
```

2. Download the correct ODM fileset version from the EMC FTP server.

```
ftp://ftp.emc.com
Connected to ftp.emc.com

Name (ftp.emc.com:root):anonymous
331 Username okay, need password.

Password:<enter your email address>
230 User logged in,proceed.

ftp> cd pub/elab/aix/ODM_DEFINITIONS
250 File action OK.

ftp> bin
200 Command OK.

ftp> get EMC.AIX.5.3.1.0.tar.Z
200 Command OK.

ftp> by
```

3. Uncompress the EMC.AIX.5.3.1.0.tar.Z file, using the following commands:

```
uncompress EMC.AIX.5.3.1.0.tar.Z
tar -xvf EMC.AIX.5.3.1.0.tar
```

4. Create a table on content file, using the following commands:

```
rm .toc
inutoc.
```

5. Install the ODM fileset EMC.XtremIO.aix.rtc to support XtremIO. Use the following command:

```
installp -ad . EMC.XtremIO.aix.rtc
```

To verify that the fileset is installed, run the following check:

```
bash-4.2# lslpp -L | grep EMC
EMC.XtremIO.aix.rtc      6.0.0.5  C  F  EMC XtremIO AIX Support
EMC.XtremIO.fcp.rtc     6.0.0.5  C  F  EMC XtremIO FCP Support
```

6. Install the ODM fileset EMC.XtremIO.fcp.rtc to support PowerPath multipathing software. Use the following command:

```
installp -ad . EMC.XtremIO.fcp.rtc
```

```

+-----+
Pre-installation Verification
+-----+
Verifying selections...done
Verifying requisites...done
Results...

SUCSESSES
-----
Filesets listed in this section passed pre-installation verification and will be installed.

Selected Filesets
-----
EMC.XtremIO.fcp.rte 5.3.1.0          # EMC XtremIO FCP Support Soft...

<< End of Success Section >>
+-----+
BUILDDATE Verification ...
+-----+
Verifying build dates...done
FILESET STATISTICS
-----
1  Selected to be installed, of which:
1  Passed pre-installation verification
----
1  Total to be installed

0503-409 installp:  bosboot verification starting...
installp:  bosboot verification completed.
+-----+
Installing Software...
+-----+

installp:  APPLYING software for:
EMC.XtremIO.fcp.rte 5.3.1.0

. . . . << Copyright notice for EMC >> . . . . .
Copyright (c) EMC Corporation 2013
All Rights Reserved.
EMC.XtremIO.fcp.rte
Version 5309 Build 13.0

. . . . << End of copyright notice for EMC >>. . . .

Finished processing all filesets.  (Total time:  13 secs).

0503-409 installp:  bosboot verification starting...
installp:  bosboot verification completed.
0503-408 installp:  bosboot process starting...

bosboot: Boot image is 55324 512 byte blocks.
0503-292 This update will not fully take effect until after a system reboot.

* * *  A T T E N T I O N  * * *
System boot image has been updated. You should reboot the system as soon as possible to properly integrate the
changes and to avoid disruption of current functionality.

installp:  bosboot process completed.
+-----+
Summaries:
+-----+
Installation Summary
-----
Name                               Level          Part          Event          Result
-----
EMC.XtremIO.fcp.rte                5.3.1.0        USR           APPLY          SUCCESS

```

7. Following the completion of the XtremIO ODM fileset installation, reboot the host.

# Multipathing Software Configuration

## Configuring AIX Native Multipathing

## Installing EMC AIX ODM

## To install EMC AIX ODM:

1. Run the following command to verify that at least AIX levels 6100-09-05-1524 or 7100-03-05-1524 are supported:

```
oslevel -s
```

```
> oslevel -s
6100-09-05-1524
```

2. Run the following commands to uncompress the EMC ODM AIX 5.3.1.0 package, EMC.AIX.5.3.1.0.tar.Z:

```
uncompress EMC.AIX.5.3.1.0.tar.Z
```

```
tar -xvf EMC.AIX.5.3.1.0.tar
```

3. Run the following commands to create a table of contents file:

```
rm .toc
```

```
inutoc .
```

4. Install the "EMC.XtremIO.fcp.MPIO.rte" ODM fileset, using the following commands, to support AIX native MPIO multipathing software:

```
installp -ad . EMC.XtremIO.fcp.MPIO.rte
```

```
>-installp -ad . EMC.XtremIO.fcp.MPIO.rte
+-----+
+           Pre-installation Verification...
+-----+
Verifying selections...done
Verifying requisites...done
Results...

SUCCESES
-----

Filesets listed in this section passed pre-installation
verification and will be installed.

Selected Filesets
-----

EMC.XtremIO.fcp.MPIO.rte.5.3.1.0      # EMC XtremIO fcp MPIO
support...

<<End of Success Section >>
```

```

+-----+
          BUILDDATE Verification...
+-----+
Verifying build dates...done
FILESET STATISTICS
-----
      1  Selected to be installed, of which:
        1  Passed pre-installation verification
      ----
      1-Total to be installed

0503-409 installp:-bosboot verification starting...
installp:-bosboot verification completed.
+-----+
          Installing Software...
+-----+

installp: APPLYING software for:
          EMC.XtremIO.fcp.MPIO.rte 5.3.1.0

. . . . << Copyright notice for EMC >> . . . . .
          Copyright (c) EMC Corporation 2014
          All Rights Reserved
          EMC.XtremIO.fcp.MPIO.rte

          Version 5310 Build 10

. . . . << End of copyright notice for EMC >> . . . . .

Finished processing all filesets (Total time: 10 secs).

0503-409 installp: bosboot verification starting...
installp: bosboot verification completed.
0503-40 installp: bosboot process starting...

bosboot: Boot image is 51228 512 byte blocks.
0503-292 This update will not fully take effect until after a
system reboot.

* * * A T T E N T I O N * * *
      System boot image has been updated. You should reboot the
system as soon as possible to properly integrate the changes and to
avoid disruption of current functionality.

installp: bosboot process completed.
+-----+
          Summaries:
+-----+

Installation Summary
-----
Name                               Level      Part      Event      Result
-----
EMC.XtremIO.fcp.MPIO.rte          5.3.1.0    USR        APPLY      SUCCESS

```

## Configuring AIX Native MPIO

**Note:** For the most updated information on MPIO support with XtremIO Storage, refer to the *XtremIO Simple Support Matrix*.

XtremIO supports AIX MPIO multipathing, using EMC ODM version 6.0.0.5 with AIX version 7100-03-05-1524 or above, and EMC ODM version 5.3.1.0 with AIX version 6100-09-05-1524 or above.

This section describes the procedure for configuring AIX MPIO multipathing for XtremIO volumes.

Configuring an MPIO-capable device uses the same commands as a non-MPIO device. At AIX, the `cfgmgr`, `chdev`, `rmdev` and `lsdev` commands support managing MPIO device instances and display their attributes. An MPIO device instance also has path instances associated with the device instance. The `mkpath`, `chpath`, `rmpath` and `lspath` commands manage path instances and display their attributes. A path instance can be added or removed from an MPIO device without unconfiguring the device.

### To configure AIX native MPIO:

1. Run the following command to configure the XtremIO devices to the AIX system:

```
cfgmgr
```

2. Run the following command to verify that the devices are correctly configured to AIX:

```
lsdev -Cc disk
```

Output example:

```
hdisk3 Available 44-T1-01 EMC XtremIO FCP MPIO Disk
hdisk4 Available 44-T1-01 EMC XtremIO FCP MPIO Disk
hdisk5 Available 44-T1-01 EMC XtremIO FCP MPIO Disk
hdisk6 Available 44-T1-01 EMC XtremIO FCP MPIO Disk
hdisk7 Available 44-T1-01 EMC XtremIO FCP MPIO Disk
hdisk8 Available 44-T1-01 EMC XtremIO FCP MPIO Disk
hdisk9 Available 44-T1-01 EMC XtremIO FCP MPIO Disk
hdisk10 Available 44-T1-01 EMC XtremIO FCP MPIO Disk
```

## MPIO Device Attributes

This section describes the attributes that are available only by multipath devices. The attributes can be displayed or changed using the commands (in particular, the `lsattr` and `chdev` commands).

The required device attribute that all MPIO devices must support is `reserve_policy`. The default `reserve_policy` value is `no_reserve` and the default `algorithm` value is `round_robin`.

Other device-specific attributes are as follows:

- ◆ **reserve\_policy** - Defines whether a reservation methodology is employed when the device is opened. Possible values are:
  - **no\_reserve** - Does not apply a reservation methodology for the device. The device may be accessed by other initiators which may be on other host systems.

- **single\_path** - Applies a SCSI2 reserve methodology for the device, i.e. the device can be accessed only by the initiator that issued the reserve. This policy uses the SCSI2 reserve to lock the device to a single initiator (path), and commands that are routed through other paths result in a reservation conflict.
- **PR\_exclusive** - Applies a SCSI3 persistent-reserve, exclusive-host methodology when the device is opened. The `PR_key` attribute is used to prevent initiators from other host systems from accessing the device.
- **PR\_shared** - Applies a SCSI3 persistent-reserve, shared-host methodology when the device is opened. The `PR_key` value must have a unique value for each system. Initiators from other host systems are required to register before they can access the device.
- **PR\_key** - Required only if the device supports any of the persistent reserve policies (`PR_exclusive` or `PR_shared`).
- ◆ **Algorithm** - Determines the methodology by which the I/O is distributed across the paths for a device. possible values are:
  - **failover** - Sends all I/O operation to a single path. If the path is marked as failed or disabled, the next available path is selected for sending all I/O operations. This algorithm maintains all enabled paths in an ordered list based on the ascending values of the `path_priority` attribute. The valid path that has the lowest path priority value is selected for each I/O operation.
  - **round\_robin** - Distributes the I/O operations across multiple enabled paths. For devices that have active and passive paths, or preferred and non-preferred paths, only a subset of the paths are used for I/O operations. If a path is marked as failed or disabled, it is no longer used for sending I/O operations. The I/O operation is distributed based on the `path_priority` attribute. By default, paths that have a higher path priority value receive a greater share of the I/O operations.
  - **shortest\_queue** - Distributes the I/O operations across multiple enabled paths. For devices that have active and passive paths, or preferred and non-preferred paths, only a subset of the paths are used for I/O operations. This algorithm is similar to the `round_robin` algorithm. However, it distributes I/O operations based on the number of pending I/O operations on each path. The path that currently has the fewest pending I/O operations is selected for the next operations. When the algorithm is set to `shortest_queue`, the `path_priority` attribute is ignored.

---

**Note:** When SCSI2 reserves or vSCSI disks are used, `failover` must be used. For other situations, it is recommended to use `shortest_queue` (if available) or `round_robin`, as they enable a maximal use of the SAN resources.

---

## Configuring PowerPath Multipathing

---

**Note:** For the most updated information on PowerPath support with XtremIO storage, refer to the *XtremIO Simple Support Matrix*.

---



---

**Note:** When switching from native multipathing to PowerPath multipathing on AIX with XtremIO, run the following command to remove the installed fileset before configuration:

```
installp -u EMC.XtremIO.fcp.MPIO.rte
```

---

XtremIO supports multipathing using EMC PowerPath on AIX. PowerPath versions 5.7 SP1 and above provide Native Loadable Array Module (LAM) for XtremIO Array devices. With this support, XtremIO devices running versions 2.4 and above are managed under the XtremIO class.

PowerPath provides enhanced path management capabilities for up to 32 paths per logical device as well as intelligent dynamic I/O load-balancing functionalities. Having multiple paths enables the host to access a storage device even if a specific path is unavailable. Multiple paths share the I/O traffic to a storage device, using intelligent load-balancing policies which enhance I/O performance and increase application availability. EMC PowerPath is the recommended multipathing choice.

PowerPath features include:

- ◆ Multiple paths - provides higher availability and I/O performance.
  - Includes LPAR, Virtual I/O Server (includes dual VIOS) and PowerHA/HACMP.
- ◆ Path management insight capabilities - PowerPath characterizes I/O patterns and aides in diagnosing I/O problems due to flaky paths or unexpected latency values. Metrics are provided on:
  - Read and write - in MB/seconds per LUN
  - Latency distribution - the high and low watermarks per path
  - Retries - the number of failed I/Os on a specific path
- ◆ Autostandby - automatically detects intermittent I/O failures and places paths into autostandby (also known as flaky paths).
- ◆ PowerPath Migration Enabler - is a host-based migration tool that allows migrating data between storage systems. PowerPath Migration Enabler works in conjunction with the host operating system (also called Host Copy) and other underlying technologies, such as Open Replicator (OR).
- ◆ Remote monitoring:
  - PowerPath Virtual Appliance 2.0 (vApp 2.0)
  - SNMP management daemon

This section describes the procedure required for configuring PowerPath multi-pathing for XtremIO volumes.

## Upgrading from PowerPath 5.7 or 5.7 P01 to 5.7 SP1 with XtremIO Devices

When upgrading from PowerPath 5.7 or 5.7 P01 to 5.7 SP1 in an environment that includes XtremIO devices, run the following command to manage the XtremIO devices after the upgrade:

```
powermt manage class=xtremio
```

Neglecting to run the `manage` command prevents the devices from being displayed in the `powermt display` and causes them to appear as unmanaged.

In fresh installations of PowerPath 5.7 SP1 the XtremIO devices are automatically managed by PowerPath.

### To configure PowerPath multi-pathing for XtremIO volumes:

1. Uncompress, untar, and install the PowerPath fileset by running the following commands:

- a. 

```
gunzip EMCPower.AIX.5.7.SP1.b013.tar.gz
```

```
tar -xvf EMCPower.AIX.5.7.SP1.b013.tar
```

```
x EMCpower_install, 36915200 bytes, 72100 tape blocks
```
- b. 

```
rm .toc
```
- c. 

```
inutoc .
```
- d. 

```
installp -ad . EMCpower
```



```
+-----+
|                                     | Pre-installation Verification      |
+-----+-----+
Verifying selections...done
Verifying requisites...done
Results...

SUCSESSES
-----
Filesets listed in this section passed pre-installation verification and will be installed.

Selected Filesets
-----
EMCpower.MgmtComponent 5.7.1.0          # PowerPath Management Compone...
EMCpower.base 5.7.1.0                   # PowerPath Base Driver and Ut...
EMCpower.migration_enabler 5.7.1.0     # PowerPath Migration Enabler ...
EMCpower.mpx 5.7.1.0                   # PowerPath Multi_Pathing Exte...
<< End of Success Section >>

+-----+-----+
|                                     | BUILDDATE Verification ...        |
+-----+-----+
Verifying build dates...done
FILESET STATISTICS
-----
4 Selected to be installed, of which:
4 Passed pre-installation verification
----
4 Total to be installed

+-----+-----+
|                                     | Installing Software...            |
+-----+-----+
installp: APPLYING software for:
    EMCpower.base 5.7.1.0
    EMCpower.mpx 5.7.1.0
    EMCpower.migration_enabler 5.7.1.0
    EMCpower.MgmtComponent 5.7.1.0

. . . . . << Copyright notice for EMCpower >> . . . . .
Copyright (c) 2014 EMC Corporation. All right reserved.
All trademarks used herein are the property of their respective owner.

License registration is NOT required to manage the CLARiiON AX series array.
Build 013
. . . . . << End of copyright notice for EMCpower >>. . . .

0518-307 odmdelete: 0 object deleted.
Finished processing all filesets. (Total time: 3 secs).

+-----+-----+
|                                     | Summaries:                        |
+-----+-----+
Installation Summary
-----
Name                               Level           Part            Event           Result
EMCpower.base                      5.7.1.0         USR              APPLY           SUCCESS
EMCpower.mpx                       5.7.1.0         USR              APPLY           SUCCESS
EMCpower.migration_enabler         5.7.1.0         USR              APPLY           SUCCESS
EMCpower.MgmtComponent             5.7.1.0         USR              APPLY           SUCCESS
```

2. Verify the installed PowerPath version.

```
> powermt version
EMC powermt for PowerPath (c) Version 5.7 SP 1 (build 13)
```

3. Register PowerPath on the host.

```
> emcpreg -install
===== EMC PowerPath Registration =====
Do you have a new registration key or keys to enter? [n]
```

4. Type **y** and press **Enter**.

```
Enter the registration key(s) for your product(s), one per
line, pressing Enter after each key.
After typing all keys, press Enter again.
Key (Enter if done):
```

5. Type the PowerPath registration key and press **Enter**. You can also add keys for other applications, such as PowerPath, Migration Enabler or PowerPath Encryption with RSA. If you typed a valid registration key, the following output appears.

```
1 key(s) successfully added.
Key successfully installed.
Key (Enter if done):
```

---

**Note:** If you type an invalid registration key, an error message appears and you are prompted to type a valid key.

---



---

**Note:** The *EMC PowerPath Family CLI and System Messages Reference Guide* provides a list of error messages returned by the **emcpreg** license registration utility.

---

6. Press **Enter**.

```
1 key(s) successfully registered.
```

## Post-Configuration Steps - Using the XtremIO Cluster

After the host configuration is completed, you can access the XtremIO cluster from the host. Refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster, for further information on creating, presenting, and managing volumes accessed from the host via either the GUI or the CLI.

### Disk Formatting

When you create volumes in XtremIO for an AIX host, the logical block (LB) size of a new XtremIO volume is 512B. 4KB LB volumes are not supported for use with AIX SAN attached volumes since the system does not support the 4096 sector size. When you create the XtremIO volume, verify that it is created using the default 512 LB size.

Name	Size	Logical Block Size	Small IO AL...	Unaligned ...	VAAI TP Alerts
VOL 746	8GB	Normal (512 LBs)	Disabled	Disabled	Disabled

1 volumes (8GB)

Current Physical Storage

7.587TB free

< Back   Next >   Finish   Cancel

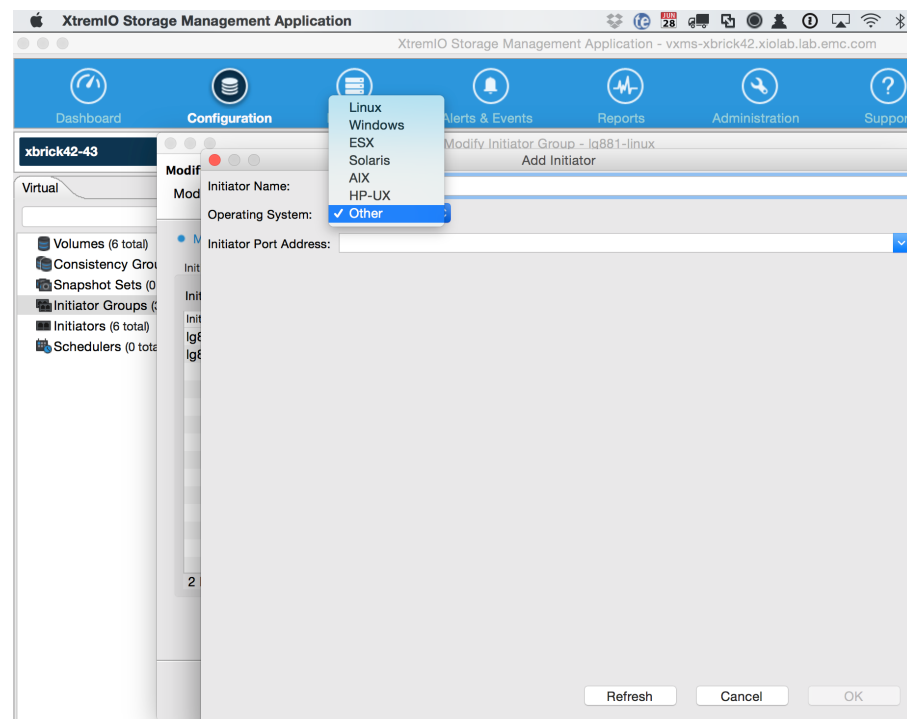
**Note:** In XtremIO version 4.0.0 (and above), the Legacy Windows option is not supported.

## Presenting XtremIO Volumes to the AIX Host

**Note:** It is recommended to avoid using host-level software for deduplication, compression and encryption of data stored in XtremIO. These operations are performed natively on the XtremIO cluster, thus releasing host computing resources. Furthermore, using these operations may severely impact data efficiency on the XtremIO cluster, as they can lead to a higher physical or logical capacity utilization.

**Note:** The following information in this section applies only to XtremIO version 4.0 and above.

When adding Initiator Groups and Initiators to allow AIX hosts to access XtremIO volumes, specify **AIX** as the operating system for the newly-created Initiators, as described in the following figure.



**Note:** Setting the Initiator's Operating System is required for optimal interoperability and stability of the host with XtremIO storage. You can adjust the setting while the host is online and connected to the XtremIO cluster with no I/O impact.

**Note:** Refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

Following a cluster upgrade from XtremIO version 3.0.x to version 4.0 (or above), make sure to modify the operating system for each initiator connected to an AIX host.

## Logical Volume Manager (LVM)

---

**Note:** Logical Volume Manager (LVM) configuration and management are out of the scope of this document.

---

When configuring LVM with an AIX host connected to XtremIO storage, follow these recommendations:

- ◆ LVM stripe size - The LVM stripe size should be set to a value greater than that of the application block size. XtremIO storage supports physical extent size of 8KB and above. the recommended physical extent size is 1MB or above.

## Creating a File System

---

**Note:** File system configuration and management are out of the scope of this document.

---

It is recommended to create the file system using its default block size (using a non-default block size may lead to unexpected behavior). Refer to your operating system and file system documentation.

---

**Note:** Some file systems may require you to properly align the file system on the XtremIO volume. It is recommended to use specified tools to optimally match your host with application requirements.

---

## Volume Configuration

After provisioning volumes on the XtremIO storage, configure them to the AIX system.

### To configure volumes to the AIX system:

1. Configure the volumes in the system by running the following command:

```
cfgmgr
```

2. Configure the volumes to PowerPath by running the following command:

```
powermt config
```

3. Display the configured volumes by running the following command:

```
lsdev -Cc disk
```

```
> lsdev -Cc disk
hdisk0 Available 44-T1-01 EMC XtremIO FCP Disk
hdisk1 Available 44-T1-01 EMC XtremIO FCP Disk
hdisk2 Available 45-T1-01 EMC XtremIO FCP Disk
hdisk3 Available 45-T1-01 EMC XtremIO FCP Disk
hdisk4 Available 46-T1-01 EMC XtremIO FCP Disk
hdisk5 Available 46-T1-01 EMC XtremIO FCP Disk
hdiskpower0 Available 44-T1-01 PowerPath Device
hdiskpower1 Available 45-T1-01 PowerPath Device
hdiskpower2 Available 46-T1-01 PowerPath Device
```

4. Display the configured PowerPath volumes by running the following command:

```
powermt display
```

```
> powermt display
XtremIO logical device count=16
=====
Host Bus Adapters  I/O Paths  Stats    HW Path  Summary Total Dead IO/Sec Q-IOs Errors
=====
0fscsi1            optimal    17       0        -        0       0
1fscsi2            optimal    17       0        -        0       0
2fscsi6            optimal    17       0        -        0       0
3fscsi7            optimal    17       0        -        0       0
```

No path should appear degraded or missing. Such a state can be due to SCD device configuration. SCD devices are configured when there are no XtremIO volumes assigned to an initiator and target port pair, or when LUN0 is not configured and appears as “EMC XtremIO FCP SCD Disk” in the command output. When this is the case, delete the SCD device by running the `rmdev -dl hdiskX` command, assign the correct volumes, including LUN0 and run `cfgmgr` to configure them to AIX.

## Using LUN 0 with XtremIO Storage

This section details considerations and steps that must be performed for LUN 0, that is to be used with AIX.

The following notes on the use of LUN numbering should be considered:

- ◆ In XtremIO version 4.0.0 (or above), volumes are numbered by default, starting from LUN id 1 (and not 0 as was the case in previous XtremIO versions).
- ◆ When a cluster is updated from XtremIO version 3.0.x to 4.0.x, an XtremIO volume with a LUN id 0 remains accessible following the upgrade.

By default, not assigning any XtremIO volumes or not assigning a volume with a LUN id 0 automatically creates an SACD LUN 0, that is inaccessible to the AIX host.

### To access a volume with a LUN 0 on an AIX host:

1. If XtremIO volumes are not yet mapped, map them to the AIX host, including LUN 0.
2. If the LUNs are not yet configured in the AIX host, configure them using the following command:

```
cfgmgr
```

3. Run the following command to display all volumes seen by the AIX host:

```
lsdev -Cc disk
```

---

**Note:** The command displays all mapped XtremIO volumes, except for the LUN 0 volume.

---

Example: Five mapped volumes were presented to the AIX host (including LUN0). However, only four volumes are seen (with `hdiskpower` in their names), indicating that LUN0 is not accessible to the host.

```
-bash -4.2# lsdev -Cc disk
hdisk0 Available Virtual SCSI Disk Drive
hdisk1 Available C5-T1-01 EMC XtremIO FCP SACD Disk
hdisk2 Available C5-T1-01 EMC XtremIO FCP SACD Disk
hdisk3 Available C5-T1-01 EMC XtremIO FCP SACD Disk
hdisk4 Available C5-T1-01 EMC XtremIO FCP SACD Disk
hdisk5 Available C6-T1-01 EMC XtremIO FCP SACD Disk
hdisk6 Available C6-T1-01 EMC XtremIO FCP SACD Disk
hdisk7 Available C6-T1-01 EMC XtremIO FCP SACD Disk
hdisk8 Available C6-T1-01 EMC XtremIO FCP SACD Disk
...
hdiskpower0 Available C6-T1-01 PowerPath Device
hdiskpower1 Available C6-T1-01 PowerPath Device
hdiskpower2 Available C6-T1-01 PowerPath Device
hdiskpower3 Available C6-T1-01 PowerPath Device
```

4. Run the following command to delete all SCAD devices displayed by `lsdev`:

```
rmdev -Rdl <device name>
```

5. Run the following command again to render the LUN0 volume accessible to the AIX host:

```
cfdmgrp
```

6. Run the following command (as done in [step 3](#)) to confirm that the LUN0 volume is seen by the AIX host:

```
lsdev -Cc disk
```

Example: At this stage, all five volumes are seen by the AIX host, including LUN 0.

```
-bash -4.2# lsdev -Cc disk
hdisk0 Available Virtual SCSI Disk Drive
hdisk2 Available C5-T1-01 EMC XtremIO FCP Disk
hdisk3 Available C5-T1-01 EMC XtremIO FCP Disk
hdisk4 Available C5-T1-01 EMC XtremIO FCP Disk
hdisk5 Available C5-T1-01 EMC XtremIO FCP Disk
hdisk6 Available C6-T1-01 EMC XtremIO FCP Disk
hdisk7 Available C6-T1-01 EMC XtremIO FCP Disk
hdisk8 Available C6-T1-01 EMC XtremIO FCP Disk
hdisk9 Available C6-T1-01 EMC XtremIO FCP Disk
...
hdiskpower0 Available C6-T1-01 PowerPath Device
hdiskpower1 Available C6-T1-01 PowerPath Device
hdiskpower2 Available C6-T1-01 PowerPath Device
hdiskpower3 Available C6-T1-01 PowerPath Device
hdiskpower4 Available C6-T1-01 PowerPath Device
```

## Configuring Boot from SAN with an XtremIO Volume

For guidelines on configuring the host to boot from an XtremIO volume, refer to [“Configuring Boot from SAN with XtremIO” on page 155](#).



# CHAPTER 7

## Host Configuration for HP-UX

This chapter includes the following topics:

◆ Fibre Channel HBA Configuration.....	146
◆ Multipathing Software Configuration .....	147
◆ Post-Configuration Steps - Using the XtremIO Cluster .....	150

---

**Note:** HP-UX 11v3 is supported with XtremIO version 2.2.3 (or above). Refer to the *XtremIO Simple Support Matrix* for the most updated information.

---

## Fibre Channel HBA Configuration

When using Fibre Channel with XtremIO, the FC Host Bus Adapters (HBA) issues described in this section should be addressed for optimal performance.

### Pre-Requisites

Follow the HP recommendations for installation and setup of the appropriate host bus adapter (HBA) for your system. It is recommended to install the latest driver version (patch), as described in the HP support site for each specific FC HBA.

Refer to [E-Lab Interoperability Navigator](#) for supported HBAs in the HP-UX environment.

### Queue Depth

---

**Note:** Changing the queue depth is designed for advanced users. Increasing the queue depth may cause the host to over-stress other arrays connected to the HP-UX host, resulting in performance degradation while communicating with them. Therefore, especially in mixed environments with multiple array types connected to the HP-UX host, compare the XtremIO recommendations for queue depth with those of other platforms before applying them.

---

Queue depth is the amount of SCSI commands (including I/O requests) that can be handled by a storage device at a given time. A queue depth can be set on either of the following:

- ◆ Initiator level - HBA queue depth
- ◆ LUN level - LUN queue depth

The LUN queue depth setting controls the amount of outstanding I/O requests per a single path.

The HBA queue depth (also referred to as execution throttle) setting controls the amount of outstanding I/O requests per HBA port.

With XtremIO and HP-UX, HBA and LUN queue depth settings should retain their default values.

If required, LUN queue depth in HP-UX can be adjusted by setting the `max_q_depth` parameter.

## Setting the LUN Queue Depth

### To set the LUN queue depth for a specific device:

- ◆ Issue the following command:

```
#scsimgr save_attr -D <special device file name> -a
max_q_depth=,value>
```

### To set the LUN queue depth for all esdisk devices presented to the HP-UX host:

- ◆ Issue the following command:

```
#scsimgr save_attr -N /escsi/esdisk -a max_q_depth=<value>
```

---

**Note:** `esdisk` is the default driver for direct access devices starting at HP-UX 11i version 3. Refer to the man page `scsi_disk(7)` for details.

---



---

**Note:** For details on the `scsimgr` command, refer to *SCSI Management and Diagnostics utility on HP-UX 11i v3* at:

[http://h20565.www2.hp.com/hpsc/doc/public/display?docId=emr\\_na-c01915618](http://h20565.www2.hp.com/hpsc/doc/public/display?docId=emr_na-c01915618)

---

## Multipathing Software Configuration

### Configuring HP-UX Native Multipathing

XtremIO supports the HP-UX Native Multipathing technology. This section describes the procedure required for configuring native multipathing for XtremIO volumes.

For optimal operation, it is recommended to keep the following HP-UX native multipathing parameters at their default values:

- ◆ `load_bal_policy` - I/O load balancing policy:  
This parameter should have the default value of `round_robin`.
- ◆ `path_fail_secs` - Timeout in seconds before declaring a LUN path offline:  
This parameter should have the default value of 120 seconds.

---

**Note:** The man page of `scsimgr_esdisk(7)` includes the list of parameters related to HP-UX native multipathing.

---

These settings ensure optimal distribution and availability of load between I/O paths from the HP-UX host to the XtremIO storage.

## Configuring PowerPath Multipathing

---

**Note:** For updated information on PowerPath support with XtremIO storage, refer to the *XtremIO Simple Support Matrix*.

---

XtremIO supports multipathing using EMC PowerPath on HP-UX. PowerPath treats XtremIO devices as generic devices. The generic Loadable Array Module (LAM) support needs to be enabled so that PowerPath can recognize and manage XtremIO devices (see [“Configuring PowerPath multipathing with XtremIO” on page 149](#)).

- ◆ For details on the PowerPath releases supported for your HP-UX host, refer to the *XtremIO Simple Support Matrix*.
- ◆ For details on generic class support for your host, refer to the *EMC PowerPath Release Notes* document for the PowerPath version you are installing.
- ◆ For details on installing and configuring PowerPath generic class support on your host, refer to the *EMC PowerPath on HP-UX Installation and Administration Guide* for the PowerPath version you are installing.

The following section provides details on configuring EMC PowerPath generic LAM with XtremIO.

## Configuring PowerPath multipathing with XtremIO

### To configure PowerPath generic LAM with XtremIO:

1. Edit the `/etc/emc/mpaa.lams` file by adding the following line:

```
managed:generic:XtremIO:XtremApp
```

2. Run PowerPath initialization, using the following command:

```
# powermt init
```

3. Run the following command, to confirm that all XtremIO volumes presented to the host were detected for PowerPath management:

```
# powermt display dev=all
```

Example:

```
# powermt display dev=all
Pseudo name=disk257
Generic ID=
Logical device ID=514F0C5902200006
state=alive; policy=ADaptive; queued-I/Os=0
=====
----- Host ----- Stor ----- I/O Path----- Stats ---
### HW Path          I/O Paths          Interf.          Mode State Q-I/Os Errors
 2 0/0/0/4/0/0/2.0x514f0c5041c74c00.0x1000000000000 c71t0d1 active alive 0 0
 0 0/0/0/3/0/0/2.0x514f0c5041c74c04.0x1000000000000 c69t0d1 active alive 0 0
 2 0/0/0/4/0/0/2.0x514f0c5041c74c04.0x1000000000000 c73t0d1 active alive 0 0
 0 0/0/0/3/0/0/2.0x514f0c5041c74c00.0x1000000000000 c67t0d1 active alive 0 0
 3 0/0/0/4/0/0/3.0x514f0c5041c74c05.0x1000000000000 c77t0d1 active alive 0 0
 1 0/0/0/3/0/0/3.0x514f0c5041c74c01.0x1000000000000 c79t0d1 active alive 0 0
 1 0/0/0/3/0/0/3.0x514f0c5041c74c05.0x1000000000000 c81t0d1 active alive 0 0
 3 0/0/0/4/0/0/3.0x514f0c5041c74c01.0x1000000000000 c75t0d1 active alive 0 0
```

---

**Note:** If not all XtremIO volumes are detected, there is a problem with the PowerPath configuration on the host. For further details on troubleshooting this issue, refer to the *EMC PowerPath on HP-UX Installation and Administration Guide* for the PowerPath version you are installing.

---

## Post-Configuration Steps - Using the XtremIO Cluster

After the host configuration is completed, you can access the XtremIO cluster from the host. For details on creating, presenting and managing volumes accessed from the host via either the GUI or the CLI, refer to *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

### Device Addressing

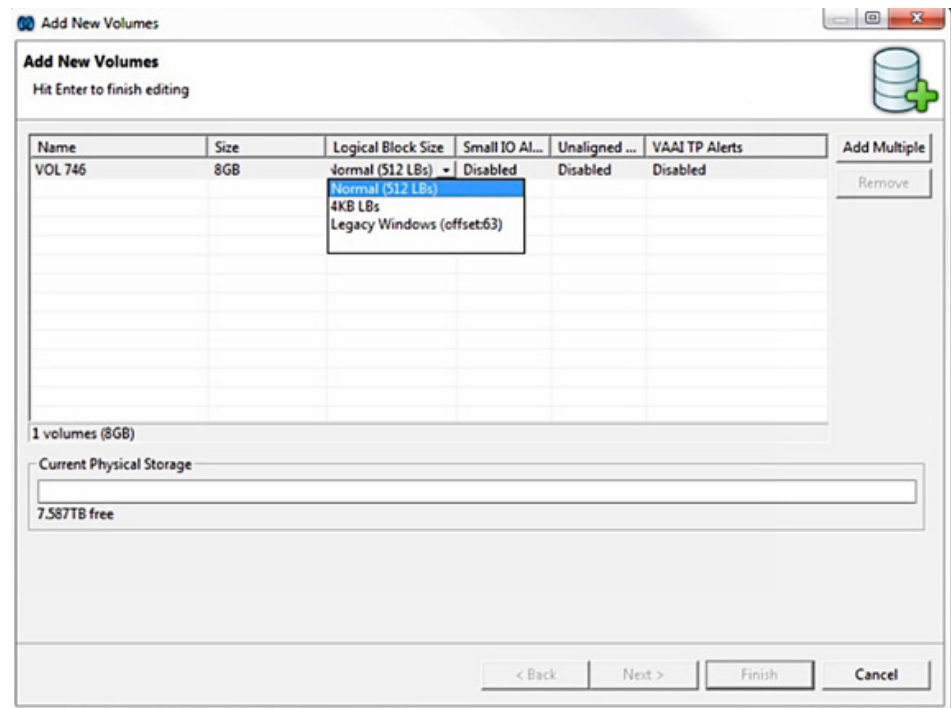
XtremIO supports only peripheral device addressing, and the only HP-UX version that can support peripheral device addressing with XtremIO is HP-UX 11iv3”.

**Note:** For further details, refer to The *Next Generation Mass Storage Stack HP-UX 11i v3* at: [http://h20565.www2.hp.com/hpsc/doc/public/display?docId=emr\\_na-c01915618](http://h20565.www2.hp.com/hpsc/doc/public/display?docId=emr_na-c01915618)

### Disk Formatting

When creating volumes in XtremIO for an HP-UX host, the following consideration should be made:

- ◆ Disk logical block size - The only logical block (LB) size, supported by HP-UX for presenting XtremIO volumes, is 512 bytes.



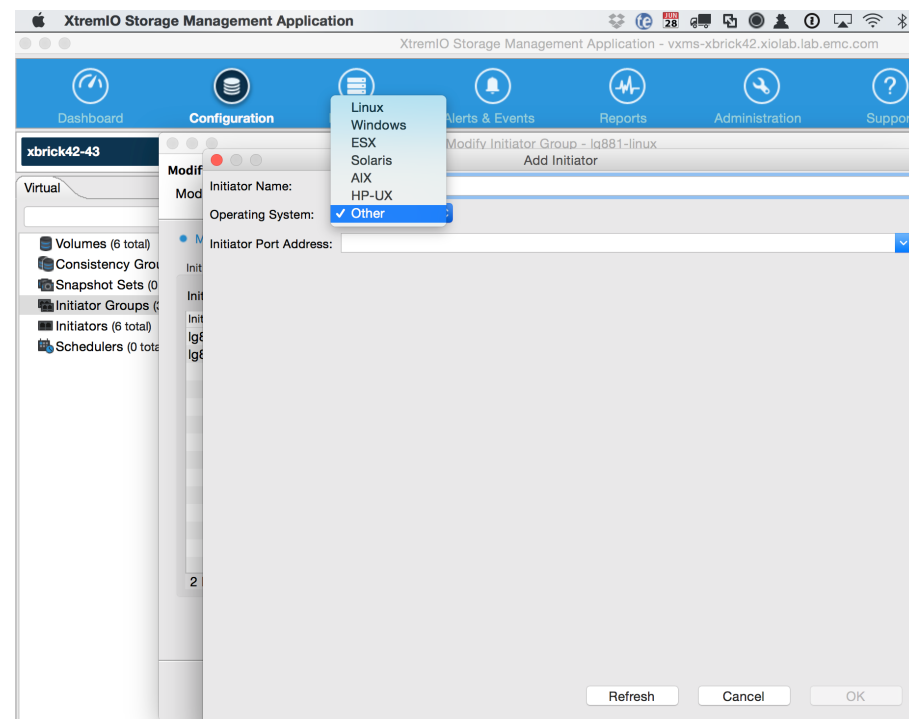
**Note:** In XtremIO version 4.0.0 (and above), the Legacy Windows option is not supported.

## Presenting XtremIO Volumes to the HP-UX Host

**Note:** It is recommended to avoid using host-level software for deduplication, compression and encryption of data stored in XtremIO. These operations are performed natively on the XtremIO cluster, thus releasing host computing resources. Furthermore, using these operations may severely impact data efficiency on the XtremIO cluster, as they can lead to a higher physical or logical capacity utilization.

**Note:** The following information in this section applies only to XtremIO version 4.0 and above.

When adding Initiator Groups and Initiators to allow HP-UX hosts to access XtremIO volumes, specify **HP-UX** as the operating system for the newly-created Initiators, as shown in the following figure.



**Note:** Setting the Initiator's Operating System is required for optimal interoperability and stability of the host with XtremIO storage. You can adjust the setting while the host is online and connected to the XtremIO cluster with no I/O impact.

**Note:** Refer to the *XtremIO Storage Array User Guide* that matches the version running on your XtremIO cluster.

Following a cluster upgrade from XtremIO version 3.0.x to version 4.0 (or above), make sure to modify the operating system for each initiator connected to an HP-UX host.

## Logical Volume Manager (LVM)

---

**Note:** Logical Volume Manager (LVM) configuration and management are out of the scope of this document.

---

When configuring LVM with an HP-UX host connected to XtremIO storage, follow these recommendations:

- ◆ LVM stripe size - The LVM stripe size should be set to a value greater than that of the application block size. XtremIO storage supports physical extent size of 8KB and above. the recommended physical extent size is 1MB or above.

## Creating a File System

---

**Note:** File system configuration and management are out of the scope of this document.

---

With XtremIO, it is recommended to create the file system using an 8KB block size. Refer to your HP-UX and file system documentation. For example, for Veritas file system, use the following command:

```
# nefws -F vxfs -o largefiles -b 8192 /dev/vgxiol/rlvxiol
```

Verify the file system block size, using the following command:

```
# fstyp -v /dev/vgxiol/rlvxiol
```

```
vxfs
```

```
version: 6
```

```
f_bsize: 8192
```

```
f_fsize: 8192
```

## Space Reclamation

Data space reclamation helps to achieve optimal XtremIO capacity utilization. This section provides steps for achieving optimal capacity utilization on the XtremIO array when connected to an HP-UX host.

With HP-UX, space reclamation is a user-initiated operation. A suggested script for initiating space reclamation is provided.



## Space Reclamation script

The following example describes an HP-UX bash space reclaim script.

```
#!/bin/bash
for i in $1
do
size=$(df -g $i |grep "total blocks" |awk '{print $4}')
name=$(df -g /mnt/space_reclamation |head -n 1 |awk -F\ ( '{print $2}'
|awk -F\ ) '{print $1}')
block_size=$(df -g /mnt/space_reclamation |grep "file system block
size" |awk '{print $1}')
reclaim=$(echo $size |awk '{printf "%.f\n", $1 * 95 / 100}')
echo $i $name $size $reclaim Test Plan
dd count=$reclaim bs=$block_size if=/dev/zero of=$i/zf
sleep 15
/bin/sync
rm -rf $i/zf
Done
```

---

**Note:** This script should also be used for Veritas File System on HP-UX, as XtremIO thin provisioning capabilities are not supported with Veritas File System.

---

## Configuring Boot from SAN with an XtremIO Volume

For guidelines on configuring the host to boot from an XtremIO volume, refer to [“Configuring Boot from SAN with XtremIO” on page 155](#).



# APPENDIX A

## Configuring Boot from SAN with XtremIO

This section provides instructions for configuring boot from SAN with XtremIO.

This section includes the following topics:

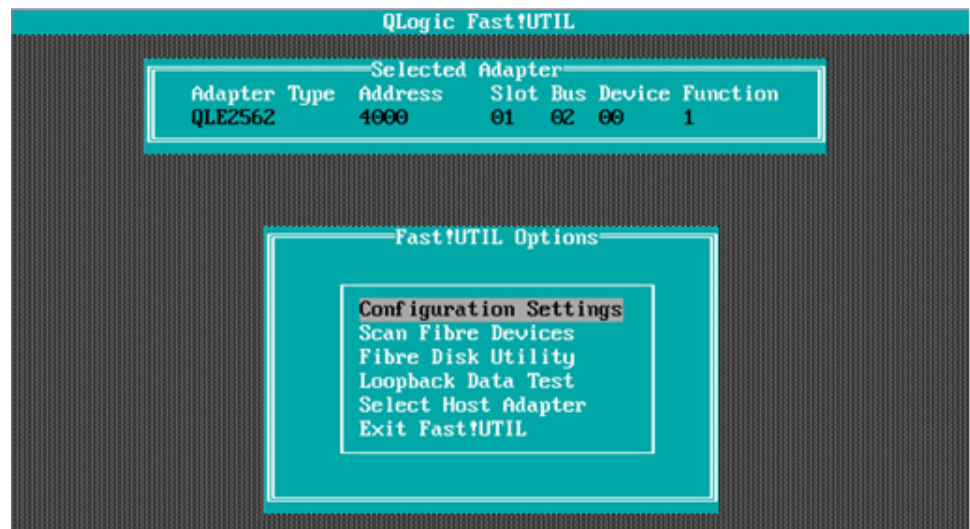
- ◆ [Configuring the Qlogic HBA BIOS for SAN Boot ..... 156](#)
- ◆ [Configuring the Emulex HBA BIOS for SAN Boot ..... 159](#)

## Configuring the Qlogic HBA BIOS for SAN Boot

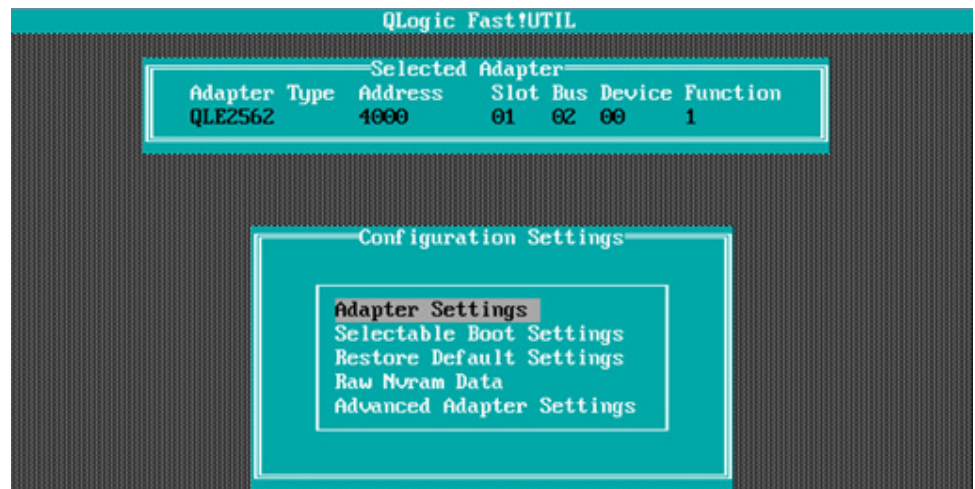
When the BIOS is installed and enabled, it needs to be configured to enable booting from the SAN.

### To configure the BIOS:

1. Connect the XtremIO Storage Controller port to the adapter in the lowest-numbered PCI slot in the server (e.g. if there are three adapters in the system, located in slots 2, 4 and 5, connect the cable to the adapter in slot 2). Do not connect cables to the other adapters.
2. Boot the server and press **Ctrl-Q** when the Qlogic banner is displayed. The banner display corresponds to the BIOS revision pertinent to the adapters that are currently installed. For specific adapters and BIOS revisions, refer to the *EMC Support Matrix*.
3. When Fast!UTIL loads, a list of addresses occupied by the adapters appears. Select the adapter from the list and press **Enter**. The Fast!UTIL Options menu appears.
4. From the Fast!UTIL Options menu, select **Configuration Settings** and press **Enter**.



5. From the Configuration Settings menu, select **Adapter Settings** and press **Enter**.

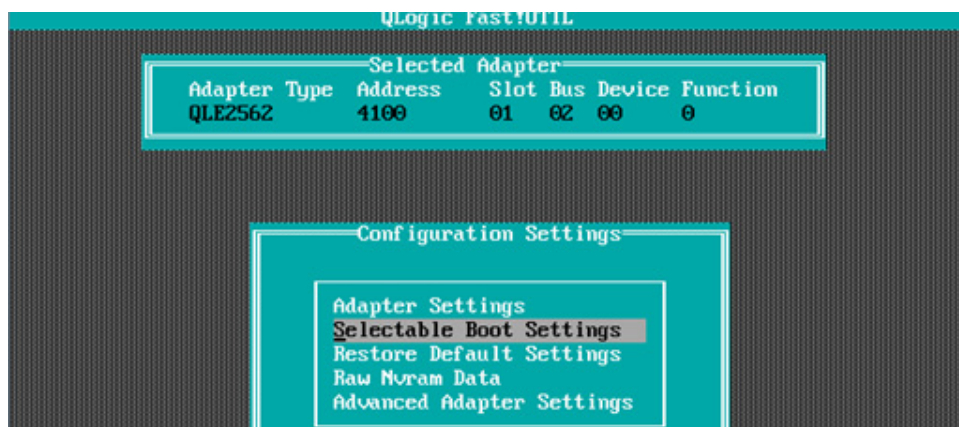


6. From the Host Adapter menu, select **Host Adapter BIOS** and press **Enter** to enable it (if it is not already enabled).

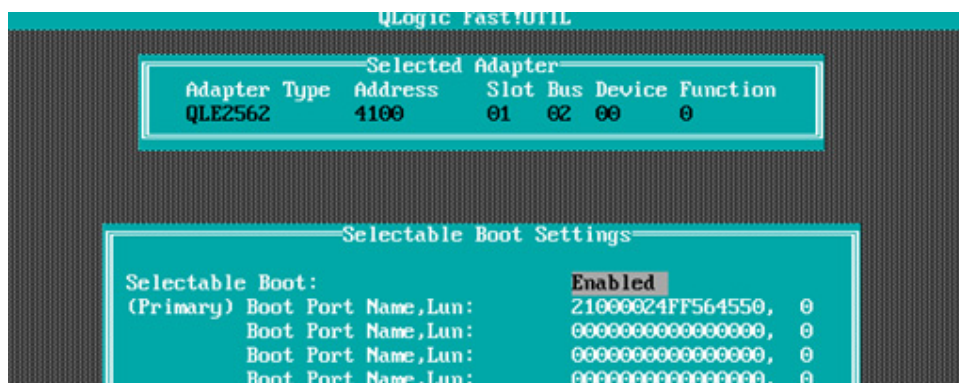


7. Press **ESC** to exit the Adapter Settings menu.

8. From the Configuration Setting menu, select **Selectable Boot Settings** and press **Enter**.



9. From the Selectable Boot Settings menu, select **Selectable Boot Device** and press **Enter** to enable this option (if it is not already enabled).



10. Select Current Boot Node Name and press Enter. The adapter scans for attached storage devices which are then displayed on the screen.
11. Select the XtremIO Storage Array port from which the system will boot. The port's entry is similar to the following example:

Select Fibre Channel Device					
ID	Vendor	Product	Rev	Port Name	Port ID
0	XtremIO	XtremApp	3020	21000024FF564550	6E0B00
1	XtremIO	XtremApp	3020	21000024FF567B9C	6E3700

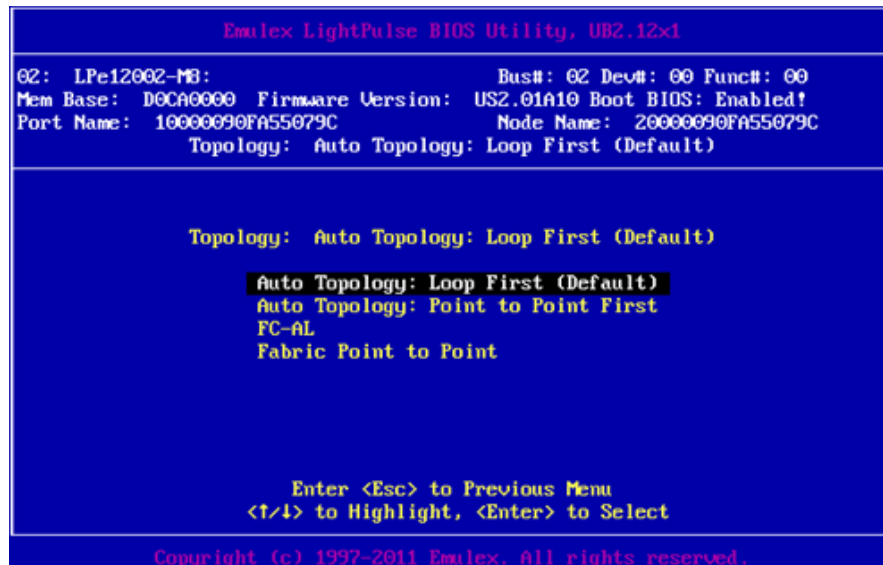
12. From the displayed list of LUNs, select the LUN to be used as the boot device and press **Enter**.
13. In the Selectable Boot Settings menu, press **ESC** to return to the Configuration Settings menu.
14. In the Configuration Settings menu, press **ESC** to return to the Options menu.
15. Select **Save Changes** and press **Enter**.
16. Press **ESC** to exit the Fast!UTIL menu.
17. Reboot the host.

## Configuring the Emulex HBA BIOS for SAN Boot

When the boot BIOS is installed and enabled, it needs to be configured to enable booting from the SAN.

### To configure the boot BIOS:

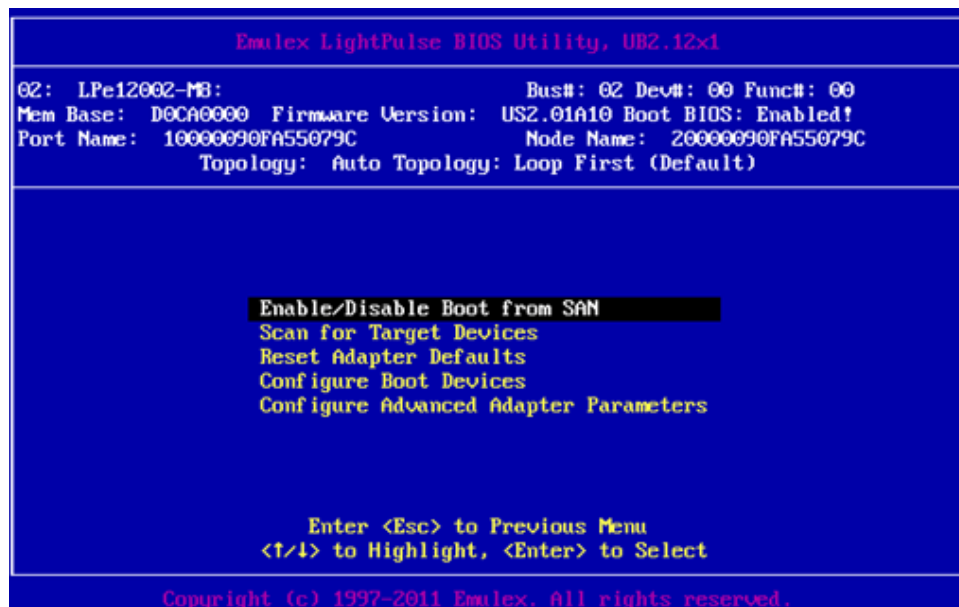
1. Connect the XtremIO Storage Controller port to the adapter in the lowest-numbered PCI slot in the server (e.g. if there are three adapters in the system, located in slots 2, 4 and 5, connect the cable to the adapter in slot 2). Do not connect cable to the other adapters.
2. Boot the server and press **ALT-E** when the Emulex banner appears, to display the BIOS setup utility. The banner display corresponds to the BIOS revision pertinent to the adapters currently installed. For specific adapters and BIOS revisions, refer to the *EMC Support Matrix*.
3. From the displayed list of adapters with the boot BIOS installed, select one HBA port and press **Enter**.
4. Verify that the displayed topology is **Auto Topology: Loop First (Default)**. If a different topology is displayed, perform the following steps:
  - a. Select **Configure Advanced Adapter Parameters**.
  - b. Select **Topology Selection**.
  - c. From the Topology Selection menu, select **Auto Topology: Loop First (Default)** and press **Enter**.



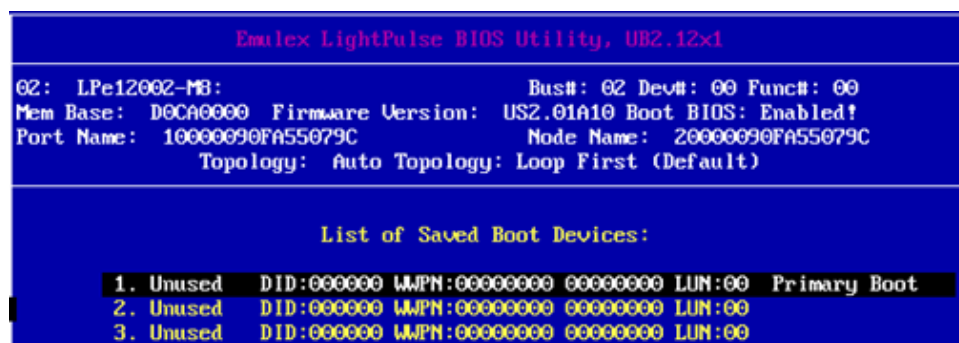
- d. Press **ESC** and then press **ESC** again to return to the main screen.



5. On the main page, select **Enable/Disable Boot from SAN** and press **Enter**.



6. If the BIOS is listed as **Disabled**, select **1** to enable it and press **Enter**.
7. Press **ESC** to return to the main screen.
8. From the main screen, select **Configure Boot Devices** and press **Enter**; the screen that appears displays a list of devices available for use as a boot LUN. Initially, the values are set to zero. However, once the information regarding the boot LUN is obtained, it is logged in the field.
9. At the prompt, select the boot entry **1** (this entry is used to fill in the information regarding the target boot device).





10. In the next screen, select **00** to clear the selected priority slot, or select the device from which the host will reboot.

In the example below, entries 01 and 02 refer to devices from XtremIO version 4.0 code, while entries 03 and 04 refer to devices from XtremIO 2.4 code.

00. Clear selected boot entry!!						
01.	DID:6E5900	WWPN:514F0C50	0217C800	LUN:01	XtremIO XtremApp	4000
02.	DID:6E5A00	WWPN:514F0C50	0217C804	LUN:01	XtremIO XtremApp	4000
03.	DID:8C0300	WWPN:21000024	FF423FE4	LUN:00	XtremIO XtremApp	2411
04.	DID:8C1900	WWPN:21000024	FF4240EA	LUN:00	XtremIO XtremApp	2411

11. Select the boot device (e.g. 01); the displayed list also provides the LUN number of the first LUN visible on the device.
12. Select the LUN that you want to boot from and press **Enter**.

```

DID:6E0B00 WWPN:21000024 FF564550
Enter two digits of starting LUN (Hex): 00

<ESC> to Previous Menu

```

13. When the next screen appears, press **Enter**.
14. When prompted to select whether to boot via WWPN or DID, select **Boot this device via WWPN**.

```

DID:6E0B00 WWPN:21000024 FF564550 LUN:00

Boot this device via WWPN
Boot this device via DID

<ESC> to Previous Menu
<↑/↓> to Highlight, <Enter> to Select

```

15. Press Enter to return to the List of Saved Boot Devices screen; the device with the corresponding target WWN is displayed.

Emulex LightPulse BIOS Utility, UB2.12x1						
02: LPe12002-MB: Bus#: 02 Dev#: 00 Func#: 00						
Mem Base: D0CA0000 Firmware Version: US2.01A10 Boot BIOS: Enabled!						
Port Name: 10000090FA55079C Node Name: 20000090FA55079C						
Topology: Auto Topology: Loop First (Default)						
List of Saved Boot Devices:						
1. Used	DID:000000	WWPN:514F0C50	0217C800	LUN:01	Primary Boot	
2. Unused	DID:000000	WWPN:00000000	00000000	LUN:00		
3. Unused	DID:000000	WWPN:00000000	00000000	LUN:00		
4. Unused	DID:000000	WWPN:00000000	00000000	LUN:00		

16. Press **ESC** to exit from the menu.
  17. Save the configuration and reboot the system.
  18. Reboot the host.
- The HBA is configured to boot from the desired LUN on the X-Brick