User's Guide

Intelligent Ethernet Adapter

QL45212

Third party information brought to you courtesy of Dell EMC.

BC0154503-00 E

Document Revision History	
Revision A, January 11, 2015	
Revision B, April 19, 2016	
Revision C, January 27, 2017	
Revision D, August 24, 2017	
Revision E, January 31, 2018	
Changes	**Sections Affected**
Updated EMI/EMC requirements: ■ CE Mark EMC Directive Compliance ■ Replaced EN55022 with EN55032. ■ VCCI ■ AS/NZS ■ Added CNS 13438:2006 Class A	"EMI and EMC Requirements" on page xvi
Added VCCI statement	"VCCI: Class A" on page xviii
Updated product safety compliance: ■ Updated UL and UL CSA compliance for 2014 ■ Deleted the 2006/95/EC low voltage directive ■ Updated TUV EN60950-1 and TUV IEC 60950-1	"Product Safety Compliance" on page xviii
Updated Table 3-5: ■ Added a column showing ESXi driver versions. ■ Added a footnote explaining that the NIC and RoCE drivers are packaged together for ESXi 6.5.	"VMware Drivers and Driver Packages" on page 28
Added a note to Step 4 stating that NPAR is not available on ports with a maximum speed of 1G.	"Getting Started" on page 41
Added a note explaining the difference between the source MAC address used by LLDP packets and the factory-assigned adapter Ethernet MAC address.	"Configuring Data Center Bridging" on page 49
Updated Table 6-1 to add a column for OFED 4.8-1 GA. Remove RHEL 7.2, SLES 11 SP4, and SLES 12 SP2. Add RHEL 7.4, SLES 12 SP3, CentOS 7.3, and CentOS 7.4	"Supported Operating Systems and OFED" on page 56
Updated Step 2 to add the **Dynamic** option.	"Preparing the Adapter" on page 58

Removed RoCE Applications and Usage subsections and promoted "Configuring a Paravirtual RDMA Device (PVRDMA)" on page 76	"Configuring RoCE on the Adapter for ESX" on page 73
Updated Step 13 to obtain the most recent adapter drivers for your OS, and not to use inbox drivers.	"Configuring SR-IOV on Windows" on page 86
Updated Step 16 to obtain the most recent adapter drivers for your OS, not to use inbox drivers, and to match driver versions on the host and VM.	"Configuring SR-IOV on Linux" on page 93
Updated Step 17 to obtain the most recent adapter drivers for your OS, not to use inbox drivers, and to match driver versions on the host and VM.	"Configuring SR-IOV on VMware" on page 98
Added Link Layer Discovery Protocol (LLDP) to the glossary	"Glossary" on page 155

Table of Contents

Preface

Supported Products xii
Intended Audience xii
What Is in This Guide xii
Documentation Conventions xiii
License Agreements xv
Legal Notices xv
Warranty xv
Laser Safety—FDA Notice xvi
Agency Certification xvi
EMI and EMC Requirements xvi
KCC: Class A xvii
VCCI: Class A xviii
Product Safety Compliance xviii

1 Product Overview

Functional Description 1
Features 1
Dell QL45212-specific Features 3
Adaptive Interrupt Frequency 3
ASIC with Embedded RISC Processor 3
Adapter Specifications 3
Physical Characteristics 3
Standards Specifications 4

2 Hardware Installation

System Requirements 6
Safety Precautions 7
Preinstallation Checklist 7
Installing the Adapter 8

3 Driver Installation

Installing Linux Driver Software 9

Installing the Linux Drivers Without RDMA 11
Removing the Linux Drivers 11
Installing Linux Drivers Using the src RPM Package 13
Installing Linux Drivers Using the kmp/kmod RPM Package . . . 14
Installing Linux Drivers Using the TAR File. 15
Installing the Linux Drivers with RDMA 15
Linux Driver Optional Parameters 16
Linux Driver Operation Defaults 17
Linux Driver Messages 17
Statistics 17
Installing Windows Driver Software 17
Installing the Windows Drivers 18
Running the DUP in the GUI. 18
DUP Installation Options. 24
DUP Installation Examples 25
Removing the Windows Drivers 25
Managing Adapter Properties 25
Setting Power Management Options. 27
Installing VMware Driver Software 27
VMware Drivers and Driver Packages. 28
Installing VMware Drivers 29
VMware Driver Optional Parameters. 30
VMware Driver Parameter Defaults. 32
Removing the VMware Driver 33

4 Upgrading the Firmware

Running the DUP by Double-Clicking 34
Running the DUP from a Command Line. 37
Running the DUP Using the .bin File 38

5 Adapter Preboot Configuration

Getting Started 41
Displaying Firmware Image Properties 44
Configuring Device-level Parameters. 45
Configuring NIC Parameters 46
Configuring Data Center Bridging 49
Configuring Partitions. 51

6 RoCE Configuration

Supported Operating Systems and OFED 56
Planning for RoCE 57

Preparing the Adapter 58
Preparing the Ethernet Switch 58
Configuring the Cisco Nexus 6000 Ethernet Switch 58
Configuring the Dell Z9100 Ethernet Switch 60
Configuring RoCE on the Adapter for Windows Server 60
Configuring RoCE on the Adapter for Linux 62
RoCE Configuration for RHEL 63
RoCE Configuration for SLES 63
Verifying the RoCE Configuration on Linux 64
VLAN Interfaces and GID Index Values 66
RoCE v2 Configuration for Linux 67
Identifying RoCE v2 GID Index or Address 67
Verifying RoCE v1 or RoCE v2 GID Index and Address from sys and class Parameters 68
Verifying RoCE v1 or RoCE v2 Function Through perftest Applications 69
Configuring RoCE on the Adapter for ESX 73
Configuring RDMA Interfaces 73
Configuring MTU 74
RoCE Mode and Statistics 75
Configuring a Paravirtual RDMA Device (PVRDMA) 76

7 iSER Configuration

Before You Begin 79
Configuring iSER for RHEL 79
Configuring iSER for SLES 12 82
Optimizing Linux Performance 83
Configuring CPUs to Maximum Performance Mode 84
Configuring Kernel sysctl Settings 84
Configuring IRQ Affinity Settings 84
Configuring Block Device Staging 85

8 SR-IOV Configuration

Configuring SR-IOV on Windows 86
Configuring SR-IOV on Linux 93
Configuring SR-IOV on VMware 98

9 Windows Server 2016

Configuring RoCE Interfaces with Hyper-V 103
Creating a Hyper-V Virtual Switch with an RDMA Virtual NIC 104
Adding a VLAN ID to Host Virtual NIC 105

Verifying If RoCE is Enabled . . . 106
Adding Host Virtual NICs (Virtual Ports) . . . 106
Mapping the SMB Drive and Running RoCE Traffic . . . 107
RoCE over Switch Embedded Teaming . . . 108
Creating a Hyper-V Virtual Switch with SET and RDMA Virtual NICs . . . 109
Enabling RDMA on SET . . . 109
Assigning a VLAN ID on SET . . . 109
Running RDMA Traffic on SET . . . 110
Configuring QoS for RoCE . . . 110
Configuring QoS by Disabling DCBX on the Adapter . . . 110
Configuring QoS by Enabling DCBX on the Adapter . . . 115
Configuring VMMQ . . . 118
Enabling VMMQ on the Adapter . . . 119
Setting the VMMQ Max QPs Default and Non-Default VPort . . . 119
Creating a Virtual Machine Switch with or Without SR-IOV . . . 120
Enabling VMMQ on the Virtual Machine Switch . . . 121
Getting the Virtual Machine Switch Capability . . . 121
Creating a VM and Enabling VMMQ on VMNetworkadapters in the VM . . . 122
Default and Maximum VMMQ Virtual NIC . . . 123
Enabling and Disabling VMMQ on a Management NIC . . . 123
Monitoring Traffic Statistics . . . 124
Configuring VXLAN . . . 124
Enabling VXLAN Offload on the Adapter . . . 124
Deploying a Software Defined Network . . . 125
Configuring Storage Spaces Direct . . . 125
Configuring the Hardware . . . 126
Deploying a Hyper-Converged System . . . 126
Deploying the Operating System . . . 127
Configuring the Network . . . 127
Configuring Storage Spaces Direct . . . 129
Deploying and Managing a Nano Server . . . 132
Roles and Features . . . 132
Deploying a Nano Server on a Physical Server . . . 134
Deploying a Nano Server in a Virtual Machine . . . 136
Managing a Nano Server Remotely . . . 138
Managing a Nano Server with Windows PowerShell Remoting . . . 138
Adding the Nano Server to a List of Trusted Hosts . . . 138
Starting the Remote Windows PowerShell Session . . . 139

Managing QLogic Adapters on a Windows Nano Server 139
RoCE Configuration. 139

10 Troubleshooting

Troubleshooting Checklist 143
Verifying that Current Drivers Are Loaded 144
Verifying Drivers in Windows 144
Verifying Drivers in Linux 144
Verifying Drivers in VMware 145
Testing Network Connectivity 145
Testing Network Connectivity for Windows 145
Testing Network Connectivity for Linux 146
Microsoft Virtualization with Hyper-V 146
Linux-specific Issues 146
Miscellaneous Issues 146
Collecting Debug Data 147

A Adapter LEDS

B Cables and Optical Modules

Supported Specifications 149
Tested Cables and Optical Modules 150
Tested Switches 152

C Dell Z9100 Switch Configuration

Glossary

List of Figures

Figure		Page
3-1	Dell Update Package Window	18
3-2	QLogic InstallShield Wizard: Welcome Window	19
3-3	QLogic InstallShield Wizard: License Agreement Window	20
3-4	InstallShield Wizard: Setup Type Window	21
3-5	InstallShield Wizard: Custom Setup Window	22
3-6	InstallShield Wizard: Ready to Install the Program Window	22
3-7	InstallShield Wizard: Completed Window	23
3-8	Dell Update Package Window	24
3-9	Setting Advanced Adapter Properties	26
3-10	Power Management Options	27
4-1	Dell Update Package: Splash Screen	34
4-2	Dell Update Package: Continue Update	35
4-3	Dell Update Package: Loading New Firmware	36
4-4	Dell Update Package: Installation Results	36
4-5	Dell Update Package: Finish Installation	37
4-6	DUP Command Line Options	38
5-1	System Setup	41
5-2	System Setup: Device Settings	41
5-3	Main Configuration Page	42
5-4	Main Configuration Page, Setting Partitioning Mode to NPAR	43
5-5	Main Configuration Page, Setting NPAR Partitioning Mode	43
5-6	Firmware Image Properties	45
5-7	System Setup: Device Level Configuration Page	45
5-8	NIC Configuration	47
5-9	System Setup: Data Center Bridging (DCB) Settings	49
5-10	NIC Partitioning Configuration, Global Bandwidth Allocation	51
5-11	Global Bandwidth Allocation, Partition 1 Configuration	52
5-12	Global Bandwidth Allocation Page: NPAReP Mode Enabled	52
5-13	Partition 1 Configuration	54
5-14	Partition 2 Configuration: FCoE Offload	54
6-1	Configuring RoCE Properties	61
6-2	Switch Settings, Server	71
6-3	Switch Settings, Client	71
6-4	Configuring RDMA_CM Applications: Server	72
6-5	Configuring RDMA_CM Applications: Client	72
6-6	Configuring a New Distributed Switch	76
6-7	Assigning a vmknic for PVRDMA	77
6-8	Setting the Firewall Rule	78
7-1	RDMA Ping Successful	80
7-2	iSER Portal Instances	81
7-3	Iface Transport Confirmed	82
7-4	Checking for New iSCSI Device	82
8-1	System Setup for SR-IOV: Integrated Devices	87

8-2 System Setup for SR-IOV: Device Level Configuration 87
8-3 Adapter Properties, Advanced: Enabling SR-IOV 88
8-4 Virtual Switch Manager: Enabling SR-IOV 89
8-5 Settings for VM: Enabling SR-IOV 91
8-6 Device Manager: VM with QLogic Adapter 92
8-7 Windows PowerShell Command: Get-NetadapterSriovVf 92
8-8 System Setup: Processor Settings for SR-IOV 93
8-9 System Setup for SR-IOV: Integrated Devices 94
8-10 Editing the grub.conf File for SR-IOV 95
8-11 Command Output for sriov_numvfs 96
8-12 Command Output for ip link show Command 96
8-13 RHEL68 Virtual Machine 97
8-14 Add New Virtual Hardware 98
8-15 VMware Host Edit Settings 101
9-1 Enabling RDMA in Host Virtual NIC 104
9-2 Hyper-V Virtual Ethernet Adapter Properties 105
9-3 Windows PowerShell Command: Get-VMNetworkAdapter 105
9-4 Windows PowerShell Command: Get-NetAdapterRdma 106
9-5 Add Counters Dialog Box 107
9-6 Performance Monitor Shows RoCE Traffic 108
9-7 Windows PowerShell Command: New-VMSwitch 109
9-8 Windows PowerShell Command: Get-NetAdapter 109
9-9 Advanced Properties: Enable QoS 111
9-10 Advanced Properties: Setting VLAN ID 112
9-11 Advanced Properties: Enabling QoS 116
9-12 Advanced Properties: Setting VLAN ID 117
9-13 Advanced Properties: Enabling Virtual Switch RSS 119
9-14 Virtual Switch Manager 120
9-15 Windows PowerShell Command: Get-VMSwitch 121
9-16 Advanced Properties: Enabling VXLAN 124
9-17 Example Hardware Configuration 126
9-18 Windows PowerShell Command: Get-NetAdapter 140
9-19 Windows PowerShell Command: Get-NetAdapterRdma 140
9-20 Windows PowerShell Command: New-Item 141
9-21 Windows PowerShell Command: New-SMBShare 141
9-22 Windows PowerShell Command: Get-NetAdapterStatistics 142

List of Tables

Table		Page
2-1	Host Hardware Requirements	6
2-2	Minimum Host Operating System Requirements	6
3-1	QLogic QL45212 Adapters Linux Drivers	9
3-2	qede Driver Optional Parameters	16
3-3	Linux Driver Operation Defaults	17
3-4	VMware Drivers	28
3-5	ESXi Driver Packages by Release	28
3-6	VMware Driver Optional Parameters	30
3-7	VMware Driver Parameter Defaults	32
5-1	Adapter Properties	44
6-1	OS Support for RoCE v1, RoCE v2, iWARP, and OFED	56
6-2	Advanced Properties for RoCE	60
9-1	Roles and Features of Nano Server	132
10-1	Collecting Debug Data Commands	147
A-1	Adapter Port Link and Activity LEDs	148
B-1	Tested Cables and Optical Modules	150
B-2	Switches Tested for Interoperability	152

Preface

This preface lists the supported products, specifies the intended audience, explains the typographic conventions used in this guide, and describes legal notices.

Supported Products

This user's guide describes installation, configuration, and management of the QL45212 25Gb Intelligent Ethernet Adapter.

Intended Audience

This guide is intended for system administrators and other technical staff members responsible for configuring and managing adapters installed on Dell® PowerEdge® servers in Windows®, Linux®, or VMware® environments.

What Is in This Guide

Following this preface, the remainder of this guide is organized into the following chapters and appendices:

- Chapter 1 Product Overview provides a product functional description, a list of features, and the adapter specifications.
- Chapter 2 Hardware Installation describes how to install the adapter, including the list of system requirements and a preinstallation checklist.
- Chapter 3 Driver Installation describes the installation of the adapter drivers on Windows, Linux, and VMware.
- Chapter 4 Upgrading the Firmware describes how to use the Dell Update Package (DUP) to upgrade the adapter firmware,
- Chapter 5 Adapter Preboot Configuration describes the preboot adapter configuration tasks using the Human Infrastructure Interface (HII) application.
- Chapter 6 RoCE Configuration describes how to configure the adapter, the Ethernet switch, and the host to use RDMA over converged Ethernet (RoCE).

- Chapter 7 iSER Configuration describes how to configure iSCSI Extensions for RDMA (iSER) for Linux RHEL and SLES.
- Chapter 8 SR-IOV Configuration provides procedures for configuring single root input/output virtualization (SR-IOV) on Windows, Linux, and VMware systems.
- Chapter 9 Windows Server 2016 describes the Windows Server 2016 features.
- Chapter 10 Troubleshooting describes a variety of troubleshooting methods and resources.
- Appendix A Adapter LEDS lists the adapter LEDs and their significance.
- Appendix B Cables and Optical Modules lists the cables and optical modules that the QL45212 Adapters support.
- Appendix C Dell Z9100 Switch Configuration describes how to configure the Dell Z9100 switch port for 25Gbps.

At the end of this guide is a glossary of terms.

Documentation Conventions

This guide uses the following documentation conventions:

- **NOTE** provides additional information.
- **CAUTION** without an alert symbol indicates the presence of a hazard that could cause damage to equipment or loss of data.
- **⚠ CAUTION** with an alert symbol indicates the presence of a hazard that could cause minor or moderate injury.
- **⚠ WARNING** indicates the presence of a hazard that could cause serious injury or death.
- Text in blue font indicates a hyperlink (jump) to a figure, table, or section in this guide, and links to Web sites are shown in underlined blue. For example:
  - Table 9-2 lists problems related to the user interface and remote agent.
  - See “Installation Checklist” on page 6.
  - For more information, visit www.cavium.com.

- Text in **bold** font indicates user interface elements such as a menu items, buttons, check boxes, or column headings. For example:
  - Click the **Start** button, point to **Programs**, point to **Accessories**, and then click **Command Prompt**.
  - Under **Notification Options**, select the **Warning Alarms** check box.
- Text in `Courier` font indicates a file name, directory path, or command line text. For example:
  - To return to the root directory from anywhere in the file structure: Type **`cd /root`** and press ENTER.
  - Issue the following command: `sh ./install.bin`
- Key names and key strokes are indicated with UPPERCASE:
  - Press CTRL+P.
  - Press the UP ARROW key.
- Text in *italics* indicates terms, emphasis, variables, or document titles. For example:
  - For a complete listing of license agreements, refer to the *Software End User License Agreement*.
  - What are *shortcut keys*?
  - To enter the date, type *mm/dd/yyyy* (where *mm* is the month, *dd* is the day, and *yyyy* is the year).
- Topic titles between quotation marks identify related topics either within this manual or in the online help, which is also referred to as *the help system* throughout this document.
- Command line interface (CLI) command syntax conventions include the following:
  - Plain text indicates items that you must type as shown. For example:
    - **qaucli -pr nic -ei**
  - `< >` (angle brackets) indicate a variable whose value you must specify. For example:
    - `<serial_number>`

NOTE

For CLI commands only, variable names are always indicated using angle brackets instead of *italics*.

- `[ ]` (square brackets) indicate an optional parameter. For example:
  - `[<file_name>]` means specify a file name, or omit it to select the default file name.
- `|` (vertical bar) indicates mutually exclusive options; select one option only. For example:
  - `on|off`
  - `1|2|3|4`
- `...` (ellipsis) indicates that the preceding item may be repeated. For example:
  - `x...` means *one* or more instances of `x`.
  - `[x...]` means *zero* or more instances of `x`.
- Vertical ellipses within command example output indicate where portions of repetitious output data have been intentionally omitted.
- `( )` (parentheses) and `{ }` (braces) are used to avoid logical ambiguity. For example:
  - `a|b c` is ambiguous
    `{(a|b) c}` means `a` or `b`, followed by `c`
    `{a|(b c)}` means either `a,` or `b c`

License Agreements

Refer to the *Software End User License Agreement* for a complete listing of all license agreements affecting this product.

Legal Notices

Legal notices covered in this section include warranty, laser safety (FDA notice), agency certification, and product safety compliance.

Warranty

For warranty details, please check the QLogic® Web site:

www.qlogic.com/Support/Pages/Warranty.aspx

Laser Safety—FDA Notice

This product complies with DHHS Rules 21CFR Chapter I, Subchapter J. This product has been designed and manufactured according to IEC60825-1 on the safety label of laser product.

CLASS I LASER

Class 1 Laser Product	**Caution**—Class 1 laser radiation when open Do not view directly with optical instruments
Appareil laser de classe 1	**Attention**—Radiation laser de classe 1 Ne pas regarder directement avec des instruments optiques
Produkt der Laser Klasse 1	**Vorsicht**—Laserstrahlung der Klasse 1 bei geöffneter Abdeckung Direktes Ansehen mit optischen Instrumenten vermeiden
Luokan 1 Laserlaite	**Varoitus**—Luokan 1 lasersäteilyä, kun laite on auki Älä katso suoraan laitteeseen käyttämällä optisia instrumenttej

Agency Certification

The following sections summarize the EMC and EMI test specifications performed on the QL45212 Adapters to comply with emission, immunity, and product safety standards.

EMI and EMC Requirements

FCC Part 15 compliance: Class A

FCC compliance information statement: This device complies with Part 15 of the FCC Rules. Operation is subject to the following two conditions: (1) this device may not cause harmful interference, and (2) this device must accept any interference received, including interference that may cause undesired operation.

ICES-003 Compliance: Class A

This Class A digital apparatus complies with Canadian ICES-003. Cet appareil numériqué de la classe A est conformé à la norme NMB-003 du Canada.

CE Mark 2014/30/EU, 2014/35/EU EMC Directive Compliance:

EN55032:2012/ CISPR 32:2015 Class A

EN55024:2010
EN61000-3-2 : Harmonic Current Emission
EN61000-3-3 : Voltage Fluctuation and Flicker

Immunity Standards
EN61000-4-2 : ESD
EN61000-4-3 : RF Electro Magnetic Field
EN61000-4-4 : Fast Transient/Burst
EN61000-4-5 : Fast Surge Common/ Differential
EN61000-4-6 : RF Conducted Susceptibility
EN61000-4-8 : Power Frequency Magnetic Field
EN61000-4-11 : Voltage Dips and Interrupt

VCCI: 2015-04; Class A

AS/NZS; CISPR 32: 2015 Class A

CNS 13438: 2006 Class A

KCC: Class A

Korea RRA Class A Certified

	Product Name/Model: Converged Network Adapters and Intelligent Ethernet Adapters Certification holder: QLogic Corporation Manufactured date: Refer to date code listed on product Manufacturer/Country of origin: QLogic Corporation/USA
A class equipment (Business purpose info/telecommunications equipment)	As this equipment has undergone EMC registration for business purpose, the seller and/or the buyer is asked to beware of this point and in case a wrongful sale or purchase has been made, it is asked that a change to household use be made.

Korean Language Format—Class A

A급 기기 (업무용 정보통신기기)

이 기기는 업무용으로 전자파적합등록을 한 기기이오니 판매자 또는 사용자는 이 점을 주의하시기 바라며, 만약 잘못판매 또는 구입하였을 때에는 가정용으로 교환하시기 바랍니다.

VCCI: Class A

This is a Class A product based on the standard of the Voluntary Control Council for Interference (VCCI). If this equipment is used in a domestic environment, radio interference may occur, in which case the user may be required to take corrective actions.

この装置は、クラスA情報技術装置です。この装置を家庭環境で使用すると電波妨害を引き起こすことがあります。この場合には使用者が適切な対策を講ずるよう要求されることがあります。 ＶＣＣＩ－Ａ

Product Safety Compliance

UL, cUL product safety:

UL 60950-1 (2nd Edition) A1 + A2 2014-10-14
CSA C22.2 No.60950-1-07 (2nd Edition) A1 +A2 2014-10

Use only with listed ITE or equivalent.

Complies with 21 CFR 1040.10 and 1040.11, 2014/30/EU, 2014/35/EU.

2006/95/EC low voltage directive:

TUV EN60950-1:2006+A11+A1+A12+A2 2nd Edition
TUV IEC 60950-1: 2005 2nd Edition Am1: 2009 + Am2: 2013 CB

CB Certified to IEC 60950-1 2nd Edition

1 Product Overview

This chapter provides the following information for the QL45212 Adapters:

- Functional Description
- Features
- Adapter Specifications

Functional Description

The QLogic QL45212 Adapter is a 25Gb Intelligent Ethernet Adapter that is designed to perform accelerated data networking for Dell PowerEdge systems. The QL45212 Adapter includes a 25Gb Ethernet MAC with full-duplex capability.

Features

The QL45212 Adapter provides the following features:

- NIC partitioning (NPAR)
- Single-chip solution:
  - 25Gb MAC
  - SerDes interface for direct attach copper (DAC) transceiver connection
  - PCIe® 3.0 x8
  - Zero copy capable hardware
- Performance features:
  - TCP, IP, UDP checksum offloads
  - TCP segmentation offload (TSO)
    - Large segment offload (LSO)
    - Generic segment offload (GSO)
    - Large receive offload (LRO)
    - Receive segment coalescing (RSC)

        - Microsoft® dynamic virtual machine queue (VMQ), virtual machine multiqueues (VMMQ), and Linux multiqueue
- Adaptive interrupts:
    - Transmit/receive side scaling (TSS/RSS)
    - Stateless offloads for Network Virtualization using Generic Routing Encapsulation (NVGRE) and virtual LAN (VXLAN) L2/L3 GRE tunneled traffic[1]
- Manageability:
    - System management bus (SMB) controller
    - *Advanced Configuration and Power Interface* (ACPI) 1.1a compliant (multiple power modes)
    - Network controller-sideband interface (NC-SI) support
- Advanced network features:
    - Jumbo frames (up to 9,600 bytes). The OS and the link partner must support jumbo frames.
    - Virtual LANs (VLAN)
    - Flow control (IEEE Std 802.3x)
- Logical link control (IEEE Std 802.2)
- High-speed on-chip reduced instruction set computer (RISC) processor
- Integrated 96KB frame buffer memory (not applicable to all models)
- 1,024 classification filters (not applicable to all models)
- Support for multicast addresses through 128-bit hashing hardware function
- Serial flash NVRAM memory
- *PCI Power Management Interface* (v1.1)
- 64-bit base address register (BAR) support
- EM64T processor support
- iSCSI and FCoE boot support[2]

[1] This feature requires OS or Hypervisor support to use the offloads.
[2] Hardware support limit of SR-IOV VFs varies. The limit may be lower on some OS environments; refer to the appropriate section for your OS.

Dell QL45212-specific Features

Features specific to the Dell QL45212 Intelligent Ethernet Adapter include adaptive interrupt frequency and ASIC with embedded RISC processor.

Adaptive Interrupt Frequency

The adapter driver intelligently adjusts host interrupt frequency based on traffic conditions to increase overall application throughput. When traffic is light, the adapter driver interrupts the host for each received packet to minimize latency. When traffic is heavy, the adapter issues one host interrupt for multiple, back-to-back incoming packets to preserve host CPU cycles.

ASIC with Embedded RISC Processor

The core control for the QL45212 Adapter resides in a tightly integrated, high-performance ASIC. The ASIC includes a RISC processor, which provides the flexibility to add new features to the card and adapts it to future network requirements through software downloads. The RISC processor also enables the adapter drivers to exploit the built-in host offload functions on the adapter as host operating systems are enhanced to take advantage of these functions.

Adapter Specifications

The QL45212 Adapter specifications include the adapter's physical characteristics and standards-compliance references.

Physical Characteristics

The QL45212 Adapters are standard PCIe cards and ship with either a full-height or a low-profile bracket for use in a standard PCIe slot.

Standards Specifications

Supported standards specifications include:

- *PCI Express Base Specification*, rev. 3.1
- *PCI Express Card Electromechanical Specification*, rev. 3.0
- *PCI Bus Power Management Interface Specification*, rev. 1.2
- IIEEE Specifications:
  - *802.3-2015 IEEE Standard for Ethernet* (flow control)
  - 802.1q (VLAN)
  - 802.1AX (Link Aggregation)
  - 802.1ad (QinQ)
  - 802.1p (Priority Encoding)
  - 1588-2002 PTPv1 (Precision Time Protocol)
  - 1588-2008 PTPv2
  - IEEE 802.3az Energy Efficient Ethernet (EEE)
- IPv4 (RFQ 791)
- IPv6 (RFC 2460)

2 Hardware Installation

This chapter provides the following hardware installation information:

- System Requirements
- Safety Precautions
- Preinstallation Checklist
- Installing the Adapter

System Requirements

Before you install a QLogic QL45212 Adapter, verify that your system meets the hardware and operating system requirements shown in Table 2-1 and Table 2-2. For a complete list of supported operating systems, visit the Cavium Web site:

Table 2-1. Host Hardware Requirements

Hardware	Requirement
Architecture	IA-32 or EMT64 that meets operating system requirements
Memory	8GB RAM (minimum)
Direct Attach Cables (DAC)	Amphenol® NDDCCGF-0001 Amphenol NDAQGF-0003 Amphenol NDCCGF-0005 Leoni® ParaLink® LA0SF064-SD-R

Table 2-2. Minimum Host Operating System Requirements

Operating System	Requirement
Windows Server	2012, 2012 R2, 2016 (including Nano)
Linux	RHEL® 6.8, 6.9, 7.2, 7.3, 7.4 SLES® 11 SP4, SLES 12 SP2, SLES 12 SP3
VMware	ESXi 6.0 u3 and later for 25G adapters

NOTE

Table 2-2 denotes minimum host OS requirements. For a complete list of supported operating systems, visit the Cavium Web site.

Safety Precautions

⚠ WARNING

The adapter is being installed in a system that operates with voltages that can be lethal. Before you open the case of your system, observe the following precautions to protect yourself and to prevent damage to the system components.

- Remove any metallic objects or jewelry from your hands and wrists.
- Make sure to use only insulated or nonconducting tools.
- Verify that the system is powered OFF and is unplugged before you touch internal components.
- Install or remove adapters in a static-free environment. The use of a properly grounded wrist strap or other personal antistatic devices and an antistatic mat is strongly recommended.

Preinstallation Checklist

Before installing the adapter, complete the following:

1. Verify that your system meets the hardware and software requirements listed under “System Requirements” on page 6.
2. Verify that your system is using the latest BIOS.

NOTE

If you acquired the adapter software from the Cavium Web site, verify the path to the adapter driver files.

3. If your system is active, shut it down.
4. When system shutdown is complete, turn off the power and unplug the power cord.
5. Remove the adapter from its shipping package and place it on an anti-static surface.
6. Check the adapter for visible signs of damage, particularly on the edge connector. Never attempt to install a damaged adapter.

Installing the Adapter

The following instructions apply to installing the QLogic QL45212 Adapters in most systems. For details about performing these tasks, refer to the manuals that were supplied with your system.

To install the adapter:

1. Review “Safety Precautions” on page 7 and “Preinstallation Checklist” on page 7. Before you install the adapter, ensure that the system power is OFF, the power cord is unplugged from the power outlet, and that you are following proper electrical grounding procedures.
2. Open the system case, and select the slot that matches the adapter size, which can be PCIe Gen 2 x8 or PCIe Gen 3 x8. A lesser width adapter can be seated into a greater width slot (x8 in an x16), but a greater width adapter cannot be seated into a lesser width slot (x8 in an x4). If you do not know how to identify a PCIe slot, refer to your system documentation.
3. Remove the blank cover-plate from the slot that you selected.
4. Align the adapter connector edge with the PCIe connector slot in the system.
5. Applying even pressure at both corners of the card, push the adapter card into the slot until it is firmly seated. When the adapter is properly seated, the adapter port connectors are aligned with the slot opening, and the adapter faceplate is flush against the system chassis.

CAUTION

Do not use excessive force when seating the card, as this may damage the system or the adapter. If you have difficulty seating the adapter, remove it, realign it, and try again.

6. Secure the adapter with the adapter clip or screw.
7. Close the system case and disconnect any personal anti-static devices.

3 Driver Installation

This chapter provides the following information about driver installation:

- Installing Linux Driver Software
- “Installing Windows Driver Software” on page 17
- “Installing VMware Driver Software” on page 27

Installing Linux Driver Software

This section describes how to install Linux drivers with or without remote direct memory access (RDMA). It also describes the Linux driver optional parameters, default values, messages, and statistics.

- Installing the Linux Drivers Without RDMA
- Installing the Linux Drivers with RDMA
- Linux Driver Optional Parameters
- Linux Driver Operation Defaults
- Linux Driver Messages
- Statistics

The QL45212 Adapter Linux drivers and supporting documentation are available on the Dell Support page:

dell.support.com

Table 3-1 describes the QL45212 Adapter Linux drivers.

Table 3-1. QLogic QL45212 Adapters Linux Drivers

Linux Driver	Description
qed	The qed core driver module directly controls the firmware, handles interrupts, and provides the low-level API for the protocol specific driver set. The qed interfaces with the qede, qedr, qedi, and qedf drivers. The Linux core module manages all PCI device resources (registers, host interface queues, and so on). The qed core module requires Linux kernel version 2.6.32 or later. Testing was concentrated on the x86_64 architecture.

Table 3-1. QLogic QL45212 Adapters Linux Drivers (Continued)

Linux Driver	Description
qede	Linux Ethernet driver for the QL45212 Adapter. This driver directly controls the hardware and is responsible for sending and receiving Ethernet packets on behalf of the Linux host networking stack. This driver also receives and processes device interrupts on behalf of itself (for L2 networking). The qede driver requires Linux kernel version 2.6.32 or later. Testing was concentrated on the x86_64 architecture.
qedr	Linux RDMA over converged Ethernet (RoCE) driver. This driver works in the OpenFabrics Enterprise Distribution (OFED™) environment in conjunction with the qed core module and the qede Ethernet driver. RDMA user space applications also require that the libqedr user library is installed on the server.
qedi	Linux iSCSI-Offload driver for the 45*xxx* Adapters. This driver works with the Open iSCSI library.
qedf	Linux FCoE-Offload driver for the 45*xxx* Adapters. This driver works with Open FCoE library.

The Linux drivers can be installed using a source Red Hat® Package Manager (RPM) package or a kmod RPM package. The RHEL RPM packages are as follows:

- `qlgc-fastlinq-<version>.<OS>.src.rpm`
- `qlgc-fastlinq-kmp-default-<version>.<arch>.rpm`

The SLES source and kmp RPM packages are as follows:

- `qlgc-fastlinq-<version>.<OS>.src.rpm`
- `qlgc-fastlinq-kmp-default-<version>.<OS>.<arch>.rpm`

The following kernel module (kmod) RPM installs Linux drivers on SLES hosts running the Xen Hypervisor:

- `qlgc-fastlinq-kmp-xen-<version>.<OS>.<arch>.rpm`

The following source RPM installs the RDMA library code on RHEL and SLES hosts:

- `qlgc-libqedr-<version>.<OS>.<arch>.src.rpm`

The following source code TAR BZip2 (BZ2) compressed file installs Linux drivers on RHEL and SLES hosts:

- `fastlinq-<version>.tar.bz2`

NOTE

For network installations through NFS, FTP, or HTTP (using a network boot disk), a driver disk that contains the qede driver may be needed. Linux boot drivers can be compiled by modifying the makefile and the make environment.

Installing the Linux Drivers Without RDMA

To install the Linux drivers without RDMA:

1. Download the QL45212 Adapter Linux drivers from Dell:

   dell.support.com

2. Remove the existing Linux drivers, as described in “Removing the Linux Drivers” on page 11.
3. Install the new Linux drivers using one of the following methods:
   - ❑ Installing Linux Drivers Using the src RPM Package
   - ❑ Installing Linux Drivers Using the kmp/kmod RPM Package
   - ❑ Installing Linux Drivers Using the TAR File

Removing the Linux Drivers

There are two procedures for removing Linux drivers: one for a non-RDMA environment and another for an RDMA environment. Choose the procedure that matches your environment.

To remove Linux drivers in a non-RDMA environment, unload and remove the drivers:

Follow the procedure that relates to the original installation method and the OS.

- If the Linux drivers were installed using an RPM package, issue the following commands:

```
rmmod qede
rmmod qed
depmod -a
rpm -e qlgc-fastlinq-kmp-default-<version>.<arch>
```

- If the Linux drivers were installed using a TAR file, issue the following commands:

```
rmmod qede
rmmod qed
depmod -a
```

  - For RHEL:

```
cd /lib/modules/<version>/extra/qlgc-fastlinq
rm -rf qed.ko qede.ko qedr.ko
```

  - For SLES:

```
cd /lib/modules/<version>/updates/qlgc-fastlinq
rm -rf qed.ko qede.ko qedr.ko
```

To remove Linux drivers in a non-RDMA environment:

1. To get the path to the currently installed drivers, issue the following command:

```
modinfo <driver name>
```

2. Unload and remove the Linux drivers.

   - If the Linux drivers were installed using an RPM package, issue the following commands:

```
modprobe -r qede
depmod -a
rpm -e qlgc-fastlinq-kmp-default-<version>.<arch>
```

   - If the Linux drivers were installed using a TAR file, issue the following commands:

```
modprobe -r qede
depmod -a
```

NOTE

If the qedr is present, issue the `modprobe -r qedr` command instead.

3. Delete the `qed.ko`, `qede.ko`, and `qedr.ko` files from the directory in which they reside. For example, in SLES, issue the following commands:

```
cd /lib/modules/<version>/updates/qlgc-fastlinq
rm -rf qed.ko
rm -rf qede.ko
rm -rf qedr.ko
depmod -a
```

To remove Linux drivers in an RDMA environment:

1. To get the path to the installed drivers, issue the following command:

```
modinfo <driver name>
```

2. Unload and remove the Linux drivers.

```
modprobe -r qedr
modprobe -r qede
modprobe -r qed
depmod -a
```

3. Remove the driver module files:

   - If the drivers were installed using an RPM package, issue the following command:

     ```
     rpm -e qlgc-fastlinq-kmp-default-<version>.<arch>
     ```

   - If the drivers were installed using a TAR file, issue the following commands for your operating system:

     For RHEL:

     ```
     cd /lib/modules/<version>/extra/qlgc-fastlinq
     rm -rf qed.ko qede.ko qedr.ko
     ```

     For SLES:

     ```
     cd /lib/modules/<version>/updates/qlgc-fastlinq
     rm -rf qed.ko qede.ko qedr.ko
     ```

Installing Linux Drivers Using the src RPM Package

To install Linux drivers using the src RPM package:

1. Issue the following at a command prompt:

```
rpm -ivh RPMS/<arch>/qlgc-fastlinq-<version>.src.rpm
```

2. Change the directory to the RPM path and build the binary RPM for the kernel.

   For RHEL:

   ```
   cd /root/rpmbuild
   rpmbuild -bb SPECS/fastlinq-<version>.spec
   ```

   For SLES:

   ```
   cd /usr/src/packages
   rpmbuild -bb SPECS/fastlinq-<version>.spec
   ```

3. Install the newly compiled RPM:

   ```
   rpm -ivh RPMS/<arch>/qlgc-fastlinq-<version>.<arch>.rpm
   ```

   NOTE

   The `--force` option may be needed on some Linux distributions if conflicts are reported.

   The drivers will be installed in the following paths.

   For SLES:

   ```
   /lib/modules/<version>/updates/qlgc-fastlinq
   ```

   For RHEL:

   ```
   /lib/modules/<version>/extra/qlgc-fastlinq
   ```

4. Turn on all ethX interfaces as follows:

   ```
   ifconfig <ethX> up
   ```

5. For SLES, use YaST to configure the Ethernet interfaces to automatically start at boot by setting a static IP address or enabling DHCP on the interface.

Installing Linux Drivers Using the kmp/kmod RPM Package

To install kmod RPM package:

1. Issue the following command at a command prompt:

   ```
   rpm -ivh qlgc-fastlinq-<version>.<arch>.rpm
   ```

2. Reload the driver:

   ```
   modprobe -r qede
   modprobe qede
   ```

Installing Linux Drivers Using the TAR File

To install Linux drivers using the TAR file:

1. Create a directory and extract the TAR files to the directory:

   ```
   tar xjvf fastlinq-<version>.tar.bz2
   ```

2. Change to the recently created directory, and then install the drivers:

   ```
   cd fastlinq-<version>
   make clean; make install
   ```

   The qed and qede drivers will be installed in the following paths.

   For SLES:

   ```
   /lib/modules/<version>/updates/qlgc-fastlinq
   ```

   For RHEL:

   ```
   /lib/modules/<version>/extra/qlgc-fastlinq
   ```

3. Test the drivers by loading them (unload the existing drivers first, if necessary):

   ```
   rmmod qede
   rmmod qed
   modprobe qed
   modprobe qede
   ```

Installing the Linux Drivers with RDMA

To install Linux drivers in an inbox OFED environment:

1. Download the QL45212 Adapter Linux drivers from the Dell:

   dell.support.com

2. Configure RoCE on the adapter, as described in “Configuring RoCE on the Adapter for Linux” on page 62.

3. Remove existing Linux drivers, as described in “Removing the Linux Drivers” on page 11.

4. Install the new Linux drivers using one of the following methods:

   - ❑ Installing Linux Drivers Using the kmp/kmod RPM Package
   - ❑ Installing Linux Drivers Using the TAR File

5. Install libqedr libraries to work with RDMA user space applications. The libqedr RPM is available only for inbox OFED. Issue the following command:

   `rpm –ivh qlgc-libqedr-<version>.<arch>.rpm`

6. To build and install the libqedr user space library, issue the following command:

   `'make libqedr_install'`

7. Test the drivers by loading them as follows:

   `modprobe qedr`
   `make install_libeqdr`

Linux Driver Optional Parameters

Table 3-2 describes the optional parameters for the qede driver.

Table 3-2. qede Driver Optional Parameters

Parameter	Description
`debug`	Controls driver verbosity level similar to `ethtool -s <dev> msglvl`.
`int_mode`	Controls interrupt mode other than MSI-X.
`gro_enable`	Enables or disables the hardware generic receive offload (GRO) feature. This feature is similar to the kernel's software GRO, but is only performed by the device hardware.
`err_flags_override`	A bitmap for disabling or forcing the actions taken in case of a hardware error: ■ bit #31 – An enable bit for this bitmask ■ bit #0 – Prevent hardware attentions from being reasserted ■ bit #1 – Collect debug data ■ bit #2 – Trigger a recovery process ■ bit #3 – Call WARN to get a call trace of the flow that led to the error

Linux Driver Operation Defaults

Table 3-3 lists the qed and qede Linux driver operation defaults.

Table 3-3. Linux Driver Operation Defaults

Operation	qed Driver Default	qede Driver Default
Speed	Auto-negotiation with speed advertised	Auto-negotiation with speed advertised
MSI/MSI-X	Enabled	Enabled
Flow Control	—	Auto-negotiation with RX and TX advertised
MTU	—	1500 (range is 46–9600)
Rx Ring Size	—	1000
Tx Ring Size	—	4078 (range is 128–8191)
Coalesce Rx Microseconds	—	24 (range is 0–255)
Coalesce Tx Microseconds	—	48
TSO	—	Enabled

Linux Driver Messages

To set the Linux driver message detail level, issue one of the following commands:

- **`ethtool -s <interface> msglvl <value>`**
- **`modprobe qede debug=<value>`**

  Where `<value>` represents bits 0–15, which are standard Linux networking values, and bits 16 and greater are driver-specific.

Statistics

To view detailed statistics and configuration information, use the ethtool utility. See the ethtool man page for more information.

Installing Windows Driver Software

- Installing the Windows Drivers
- Removing the Windows Drivers
- Managing Adapter Properties
- Setting Power Management Options

Installing the Windows Drivers

Install Windows driver software using the Dell Update Package (DUP):

- Running the DUP in the GUI
- DUP Installation Options
- DUP Installation Examples

Running the DUP in the GUI

To run the DUP in the GUI:

1. Double-click the icon representing the Dell Update Package file.

NOTE

The actual file name of the Dell Update Package varies.

2. In the Dell Update Package window (Figure 3-1), click **Install**.

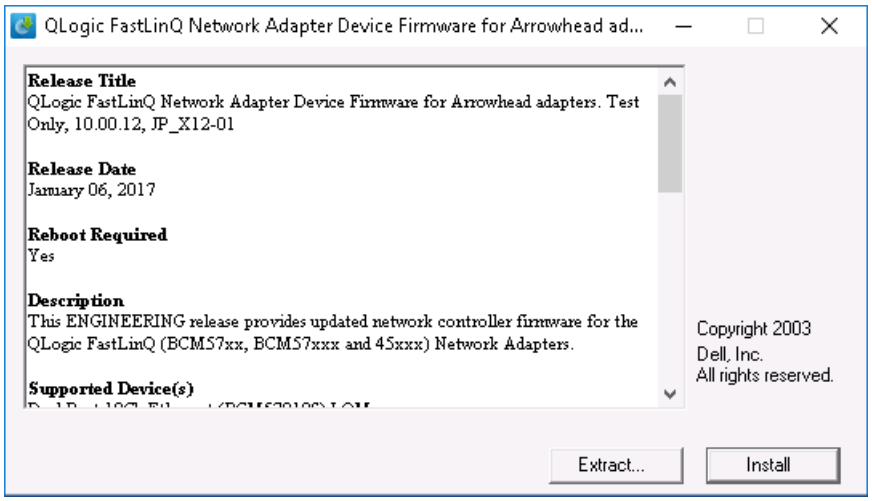


Figure 3-1. Dell Update Package Window

3. In the QLogic Super Installer—InstallShield® Wizard's Welcome window (Figure 3-2), click **Next**.



Figure 3-2. QLogic InstallShield Wizard: Welcome Window

4. Complete the following in the wizard's License Agreement window (Figure 3-3):
   a. Read the QLogic End User Software License Agreement.
   b. To continue, select **I accept the terms in the license agreement**.

c. Click **Next**.

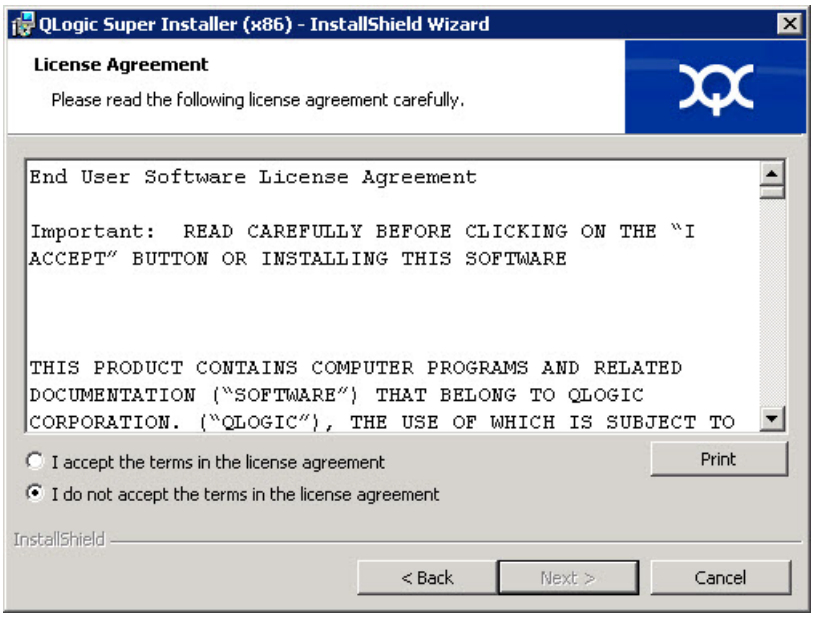


Figure 3-3. QLogic InstallShield Wizard: License Agreement Window

5. Complete the wizard's Setup Type window (Figure 3-4) as follows:
   a. Select one of the following setup types:
      - Click **Complete** to install all program features.
      - Click **Custom** to manually select the features to be installed.
   b. To continue, click **Next**.

If you clicked **Complete**, proceed directly to Step 6b.

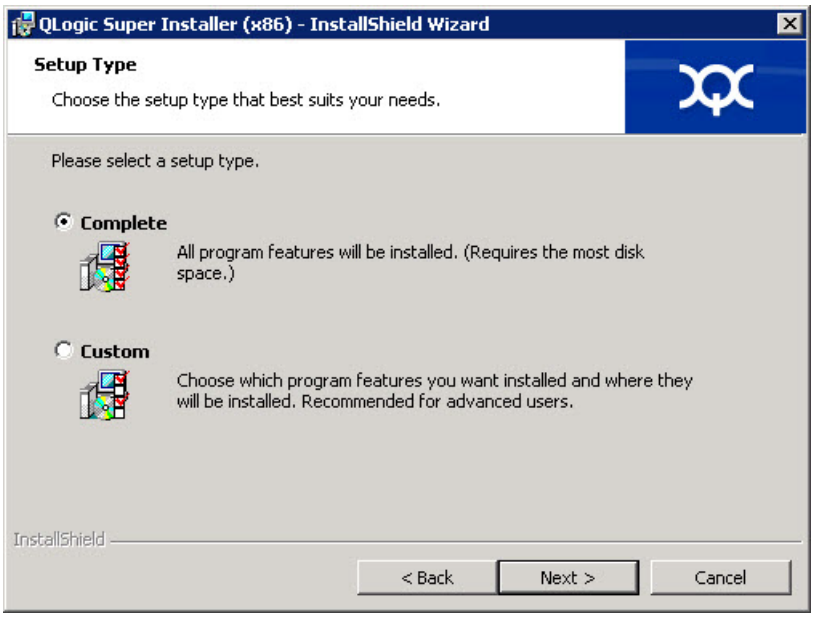


Figure 3-4. InstallShield Wizard: Setup Type Window

6. If you selected **Custom** in Step 5, complete the Custom Setup window (Figure 3-5) as follows:

   a. Select the features to install. By default, all features are selected. To change a feature's install setting, click the icon next to it, and then select one of the following options:

   - **This feature will be installed on the local hard drive**—Marks the feature for installation without affecting any of its subfeatures.
   - **This feature, and all subfeatures, will be installed on the local hard drive**—Marks the feature and all of its subfeatures for installation.
   - **This feature will not be available**—Prevents the feature from being installed.

b. Click **Next** to continue.

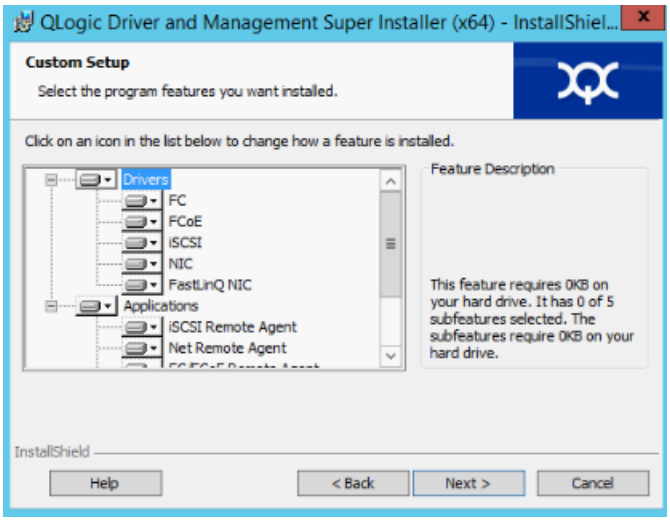


Figure 3-5. InstallShield Wizard: Custom Setup Window

7. In the InstallShield Wizard's Ready To Install window (Figure 3-6), click **Install**. The InstallShield Wizard installs the QLogic Adapter drivers and Management Software Installer.

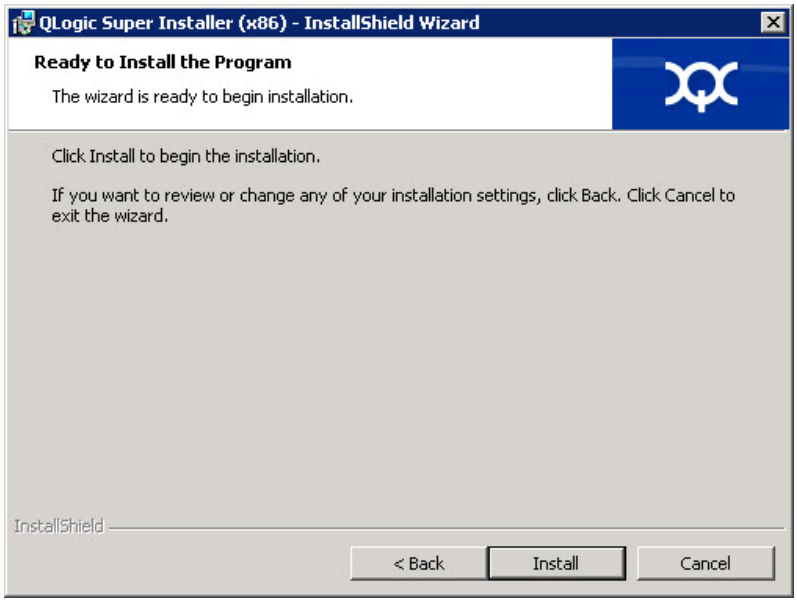


Figure 3-6. InstallShield Wizard: Ready to Install the Program Window

8. When the installation is complete, the InstallShield Wizard Completed window appears (Figure 3-7). Click **Finish** to dismiss the installer.

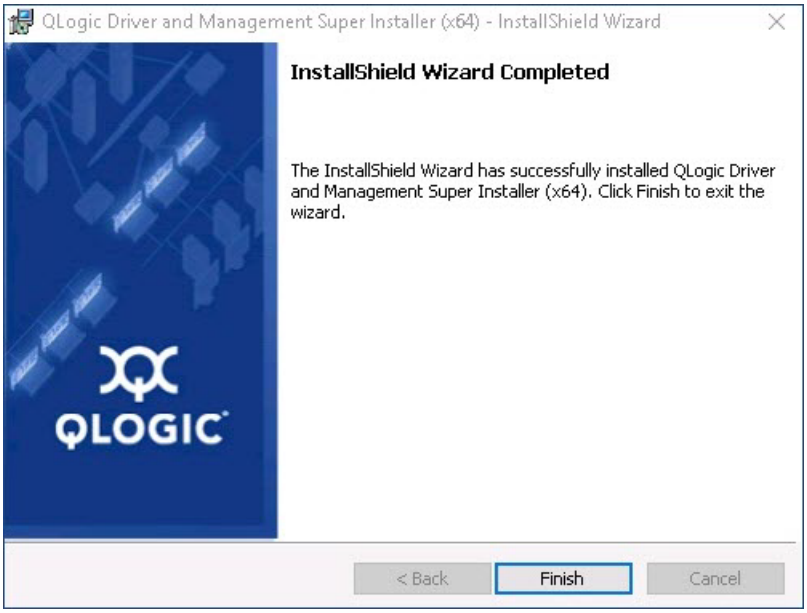


Figure 3-7. InstallShield Wizard: Completed Window

9. In the Dell Update Package window (Figure 3-8), “Update installer operation was successful” indicates completion.

   ❑ (Optional) To open the log file, click **View Installation Log**. The log file shows the progress of the DUP installation, any previous installed versions, any error messages, and other information about the installation.

❑ To close the Update Package window, click **CLOSE**.

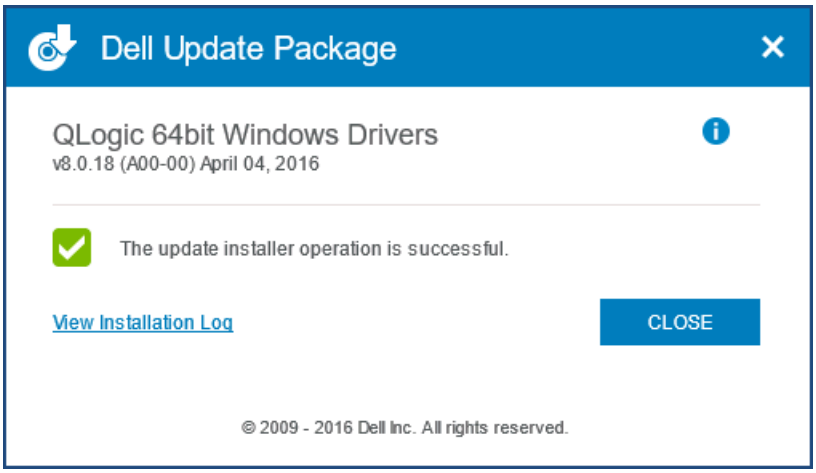


Figure 3-8. Dell Update Package Window

DUP Installation Options

To customize the DUP installation behavior, use the following command line options.

- To extract only the driver components to a directory:

  `/drivers=<path>`

  NOTE

  This command requires the **`/s`** option.

- To install or update only the driver components:

  `/driveronly`

  NOTE

  This command requires the **`/s`** option.

- (Advanced) Use the `/passthrough` option to send all text following `/passthrough` directly to the QLogic installation software of the DUP. This mode suppresses any provided GUIs, but not necessarily those of the QLogic software.

  `/passthrough`

- (Advanced) To return a coded description of this DUP's supported features:

  `/capabilities`

  NOTE

  This command requires the **`/s`** option.

DUP Installation Examples

The following examples show how to use the installation options.

To update the system silently:

`<DUP_file_name>.exe /s`

To extract the update contents to the `C:\mydir\` directory:

`<DUP_file_name>.exe /s /e=C:\mydir`

To extract the driver components to the `C:\mydir\` directory:

`<DUP_file_name>.exe /s /drivers=C:\mydir`

To install only the driver components:

`<DUP_file_name>.exe /s /driveronly`

To change from the default log location to `C:\my path with spaces\log.txt`:

`<DUP_file_name>.exe /l="C:\my path with spaces\log.txt"`

Removing the Windows Drivers

To remove the Windows drivers:

1. In the Control Panel, click **Programs**, and then click **Programs and Features**.
2. In the list of programs, select **QLogic FastLinQ Driver Installer**, and then click **Uninstall**.
3. Follow the instructions to remove the drivers.

Managing Adapter Properties

To view or change the QL45212 Adapter properties:

1. In the Control Panel, click **Device Manager**.
2. On the properties of the selected adapter, click the **Advanced** tab.

3. On the Advanced page (Figure 3-9), select an item under **Property** and then change the **Value** for that item as needed.

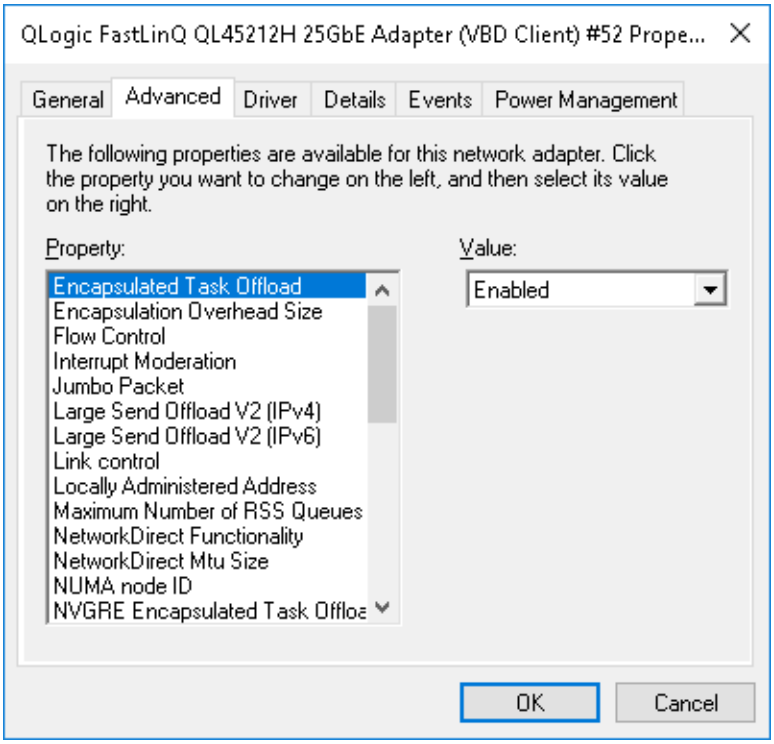


Figure 3-9. Setting Advanced Adapter Properties

Setting Power Management Options

You can set power management options to allow the operating system to turn off the controller to save power or to allow the controller to wake up the computer. If the device is busy (servicing a call, for example), the operating system will not shut down the device. The operating system attempts to shut down every possible device only when the computer attempts to go into hibernation. To have the controller remain on at all times, do not select the **Allow the computer to turn off the device to save power** check box (Figure 3-10).

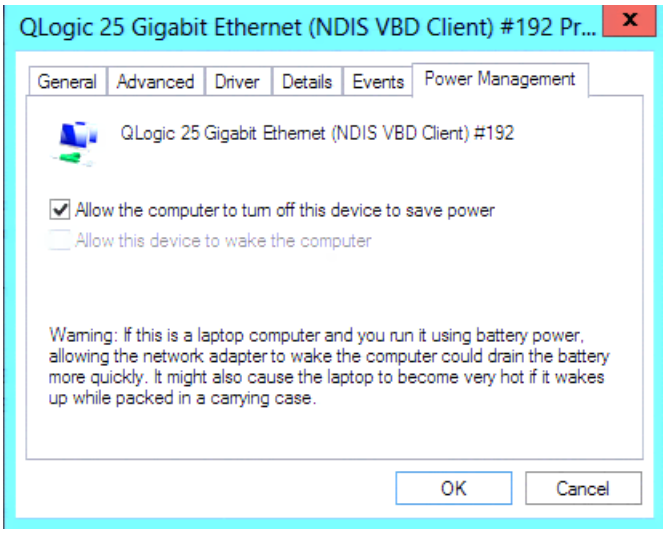


Figure 3-10. Power Management Options

NOTE

- The Power Management page is available only for servers that support power management.
- Do not select **Allow the computer to turn off the device to save power** for any adapter that is a member of a team.

Installing VMware Driver Software

This section describes the qedentv VMware ESXi driver for the QL45212 Adapters:

- VMware Drivers and Driver Packages
- Installing VMware Drivers
- VMware Driver Optional Parameters
- VMware Driver Parameter Defaults
- Removing the VMware Driver

VMware Drivers and Driver Packages

Table 3-4 lists the VMware ESXi drivers for the protocols.

Table 3-4. VMware Drivers

VMware Drivers	Description
qedentv	Native networking driver
qedrntv	Native RDMA-Offload (RoCE and RoCEv2) driver [a]
qedf	Native FCoE-Offload driver
qedil	Legacy iSCSI-Offload driver

[a] The certified RoCE driver is not included in this release. The uncertified driver may be available as an early preview.

The ESXi drivers are included as individual driver packages and are not bundled together, except as noted. Table 3-5 lists the ESXi versions and applicable driver versions.

Table 3-5. ESXi Driver Packages by Release

ESXi Release	Protocol	Driver Name	Driver Version
ESXi 6.5 [a]	NIC	qedentv	3.0.7.5
	FCoE	qedf	1.2.24.0
	iSCSI	qedil	1.0.19.0
	RoCE	qedrntv	3.0.7.5.1
ESXi 6.0u3	NIC	qedentv	2.0.7.5
	FCoE	qedf	1.2.24.0
	iSCSI	qedil	1.0.19.0

[a] For ESXi 6.5, the NIC and RoCE drivers have been packaged together and can be installed as a single offline bundle using the standard ESXi installation commands. The package name is *qedentv_3.0.7.5_qedrntv_3.0.7.5.1_signed_drivers.zip*. The recommended installation sequence is NIC and RoCE drivers, followed by FCoE and iSCSI drivers.

Install individual drivers using either:

- Standard ESXi package installation commands (see Installing VMware Drivers)
- Procedures in the individual driver Read Me files
- Procedures in the following VMware KB article:

  https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2137853

You should install the NIC driver first, followed by the storage drivers.

Installing VMware Drivers

You can use the driver ZIP file to install a new driver or update an existing driver. Be sure to install the entire driver set from the same driver ZIP file. Mixing drivers from different ZIP files will cause problems.

To install the VMware driver:

1. Download the VMware driver for the QL45212 Adapter from the VMware support page:

   www.vmware.com/support.html

2. Power up the ESX host, and then log into an account with administrator authority.
3. Unzip the driver ZIP file, and then extract the `.vib` file.
4. Use the Linux scp utility to copy a `.vib` file from a local system into the `/tmp` directory on an ESX server with IP address 10.10.10.10. For example, issue the following command:

```
#scp qedentv-1.0.3.11-1OEM.550.0.0.1331820.x86_64.vib root@10.10.10.10:/tmp
```

You can place the file anywhere that is accessible to the ESX console shell.

NOTE

If you do not have a Linux machine, you can uses the vSphere datastore file browser to upload the files to the server.

5. Place the host in maintenance mode by issuing the following command:

```
#esxcli --maintenance-mode
```

6. Select one of the following installation options:

   - **Option 1:** Install the `.vib` directly on an ESX server using either the CLI or the VMware Update Manager (VUM):

     - To install the `.vib` file using the CLI, issue the following command. Be sure to specify the full `.vib` file path.

       ```
       # esxcli software vib install -v
       /tmp/qedentv-1.0.3.11-1OEM.550.0.0.1331820.x86_64.vi
       b
       ```

     - To install the `.vib` file using the VUM, see the knowledge base article here:

       Updating an ESXi/ESX host using VMware vCenter Update Manager 4.x and 5.x (1019545)

   - **Option 2:** Install all of the individual VIBs at one time by issuing the following command:

     ```
     # esxcli software vib install –d
     /tmp/qedentv-bundle-2.0.3.zip
     ```

To upgrade an existing driver:

Follow the steps for a new installation, except replace the command in the preceding Option 1 with the following:

```
#esxcli software vib update -v
/tmp/qedentv-1.0.3.11-1OEM.550.0.0.1331820.x86_64.vib
```

VMware Driver Optional Parameters

Table 3-6 describes the optional parameters that can be supplied as command line arguments to the `esxcfg-module` command.

Table 3-6. VMware Driver Optional Parameters

Parameter	Description
`hw_vlan`	Globally enables (`1`) or disables (`0`) hardware VLAN insertion and removal. Disable this parameter when the upper layer needs to send or receive fully formed packets. `hw_vlan=1` is the default.

Table 3-6. VMware Driver Optional Parameters (Continued)

Parameter	Description
num_queues	Specifies the number of TX/RX queue pairs. num_queues can be 1–11 or one of the following: ■ –1 allows the driver to determine the optimal number of queue pairs (default). ■ 0 uses the default queue. You can specify multiple values delimited by commas for multiport or multi-function configurations.
multi_rx_filters	Specifies the number of RX filters per RX queue, excluding the default queue. multi_rx_filters can be 1–4 or one of the following values: ■ –1 uses the default number of RX filters per queue. ■ 0 disables RX filters.
disable_tpa	Enables (0) or disables (1) the TPA (LRO) feature. disable_tpa=0 is the default.
max_vfs	Specifies the number of virtual functions (VFs) per physical function (PF). max_vfs can be 0 (disabled) or 64 VFs on a single port (enabled). The 64 VF maximum support for ESXi is an OS resource allocation constraint.
RSS	Specifies the number of receive side scaling queues used by the host or virtual extensible LAN (VXLAN) tunneled traffic for a PF. RSS can be 2, 3, 4, or one of the following values: ■ –1 uses the default number of queues. ■ 0 or 1 disables RSS queues. You can specify multiple values delimited by commas for multiport or multi-function configurations.
debug	Specifies the level of data that the driver records in the vmkernel log file. debug can have the following values, shown in increasing amounts of data: ■ 0x80000000 indicates Notice level. ■ 0x40000000 indicates Information level (includes the Notice level). ■ 0x3FFFFFFF indicates Verbose level for all driver submodules (includes the Information and Notice levels).
auto_fw_reset	Enables (1) or disables (0) the driver automatic firmware recovery capability. When this parameter is enabled, the driver attempts to recover from events such as transmit timeouts, firmware asserts, and adapter parity errors. The default is auto_fw_reset=1.

Table 3-6. VMware Driver Optional Parameters (Continued)

Parameter	Description
vxlan_filter_en	Enables (1) or disables (0) the VXLAN filtering based on the outer MAC, the inner MAC, and the VXLAN network (VNI), directly matching traffic to a specific queue. The default is vxlan_filter_en=1. You can specify multiple values delimited by commas for multiport or multifunction configurations.
enable_vxlan_offld	Enables (1) or disables (0) the VXLAN tunneled traffic checksum offload and TCP segmentation offload (TSO) capability. The default is enable_vxlan_offld=1. You can specify multiple values delimited by commas for multiport or multifunction configurations.

VMware Driver Parameter Defaults

Table 3-7 lists the VMware driver parameter default values.

Table 3-7. VMware Driver Parameter Defaults

Parameter	Default
Speed	Autonegotiation with all speeds advertised. The speed parameter must be the same on all ports. If autonegotiation is enabled on the device, all of the device ports will use autonegotiation.
Flow Control	Autonegotiation with RX and TX advertised
MTU	1,500 (range 46–9,600)
Rx Ring Size	8,192 (range 128–8,192)
Tx Ring Size	8,192 (range 128–8,192)
MSI-X	Enabled
Transmit Send Offload (TSO)	Enabled
Large Receive Offload (LRO)	Enabled
RSS	Enabled (four RX queues)
HW VLAN	Enabled
Number of Queues	Enabled (eight RX/TX queue pairs)
Wake on LAN (WoL)	Disabled

Removing the VMware Driver

To remove the `.vib` file (qedentv), issue the following command:

```
# esxcli software vib remove --vibname qedentv
```

To remove the driver, issue the following command:

```
# vmkload_mod -u qedentv
```

4 Upgrading the Firmware

This chapter provides information about upgrading the firmware using the Dell Update Package (DUP).

The firmware DUP is a Flash update utility only; it is not used for adapter configuration. You can run the firmware DUP by double-clicking the executable file. Alternatively, you can run the firmware DUP from the command line with several supported command line options.

- Running the DUP by Double-Clicking
- Running the DUP from a Command Line
- Running the DUP Using the .bin File (Linux only)

Running the DUP by Double-Clicking

To run the firmware DUP by double-clicking the executable file:

1. Double-click the icon representing the firmware Dell Update Package file.

   The Dell Update Package splash screen appears, as shown in Figure 4-1. Click **Install** to continue.

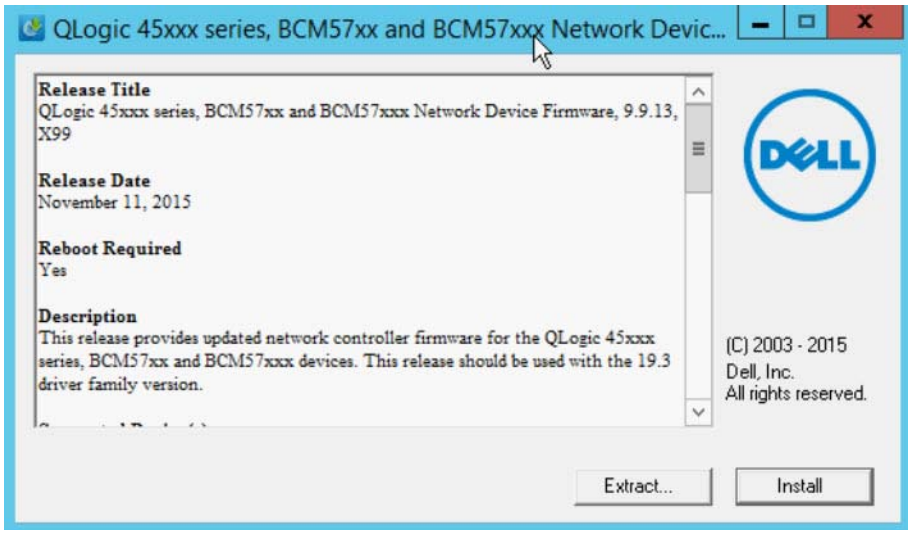


Figure 4-1. Dell Update Package: Splash Screen

2. Follow the on-screen instructions. In the Warning dialog box, click **Yes** to continue the installation, as shown in Figure 4-2.

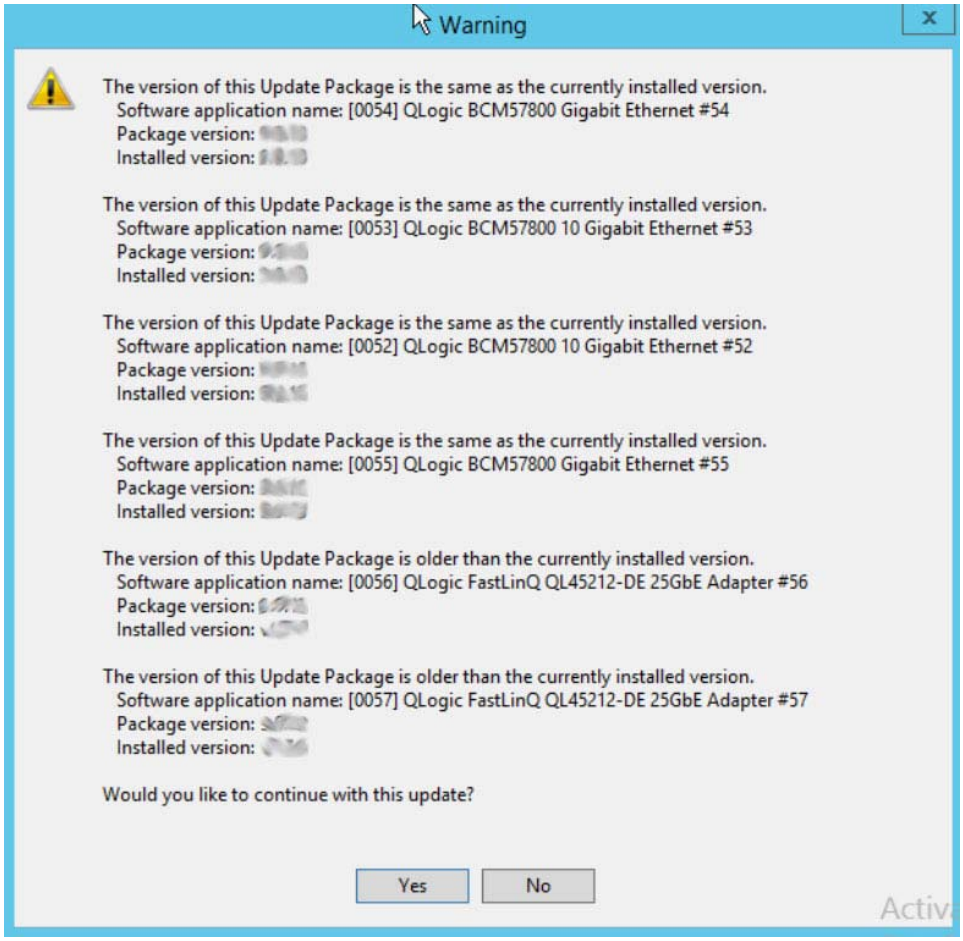


Figure 4-2. Dell Update Package: Continue Update

The installer indicates that it is loading the new firmware, as shown in Figure 4-3.

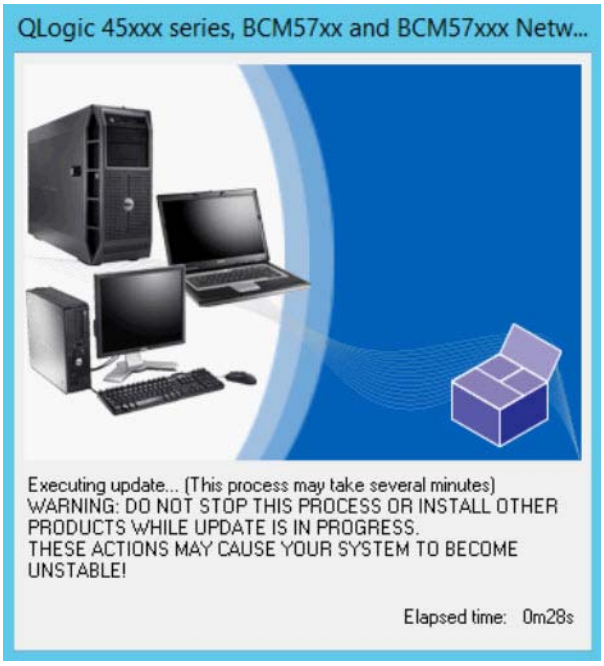


Figure 4-3. Dell Update Package: Loading New Firmware

When complete, the installer indicates the result of the installation, as shown in Figure 4-4.

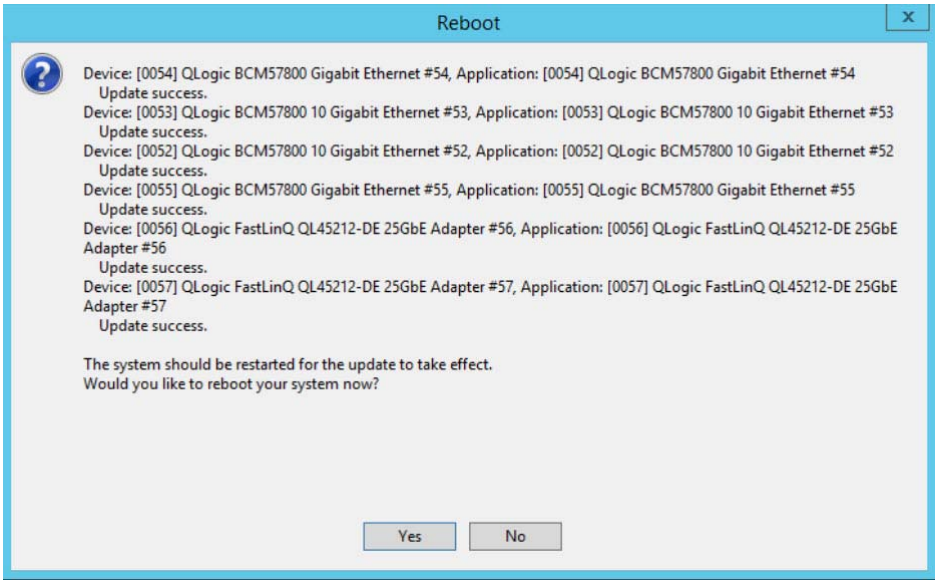


Figure 4-4. Dell Update Package: Installation Results

3. Click **Yes** to reboot the system.
4. Click **Finish** to complete the installation, as shown in Figure 4-5.

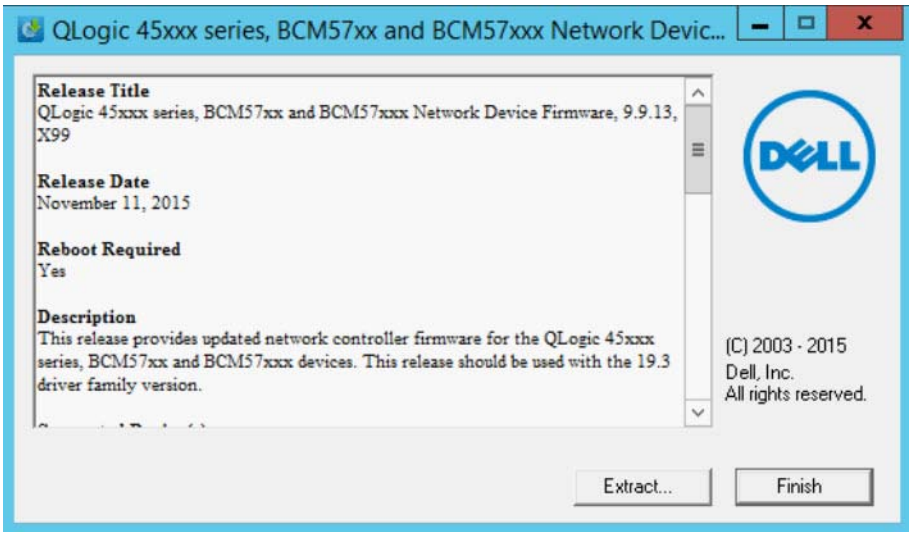


Figure 4-5. Dell Update Package: Finish Installation

Running the DUP from a Command Line

Running the firmware DUP from the command line, with no options specified, results in the same behavior as double-clicking the DUP icon. Note that the actual file name of the DUP will vary.

To run the firmware DUP from a command line:

- Issue the following command:

```
C:\> Network_Firmware_2T12N_WN32_<version>_X16.EXE
```

Figure 4-6 shows the options that you can use to customize the Dell Update Package installation.

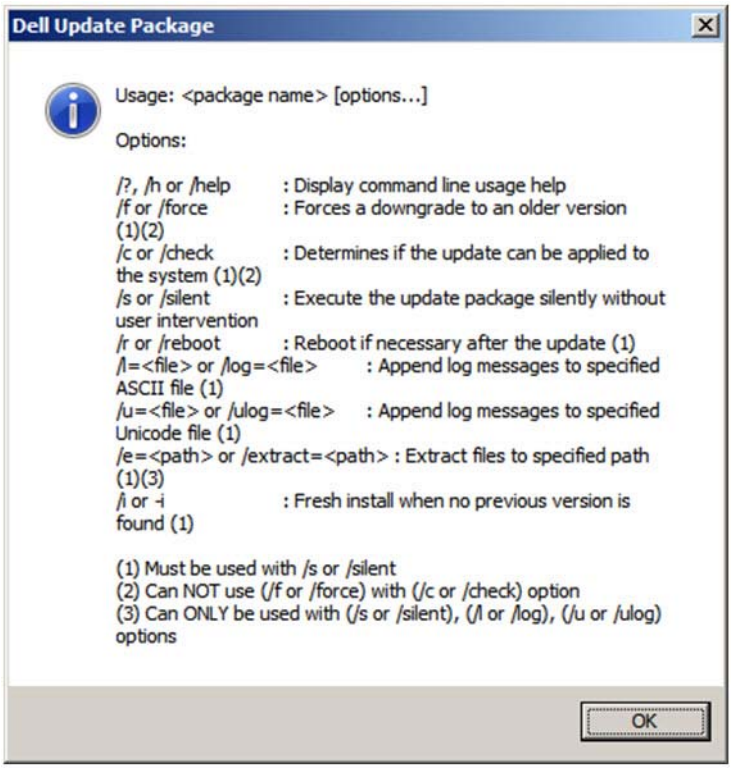


Figure 4-6. DUP Command Line Options

Running the DUP Using the .bin File

The following procedure is supported only on Linux OS.

To update the DUP using the .bin file:

1. Copy the `Network_Firmware_NJCX1_LN_X.Y.Z.BIN` file to the system under test (SUT).
2. Change the file type into an executable file as follows:

   `chmod 777 Network_Firmware_NJCX1_LN_X.Y.Z.BIN`
3. To start the update process, issue the following command:

   `./Network_Firmware_NJCX1_LN_X.Y.Z.BIN`
4. After the firmware is updated, reboot the system.

Example output from SUT during the DUP update:

```
./Network_Firmware_NJCX1_LN_08.07.26.BIN
```

```
Collecting inventory...
Running validation...
BCM57810 10 Gigabit Ethernet rev 10 (p2p1)
The version of this Update Package is the same as the currently installed
version.
Software application name: BCM57810 10 Gigabit Ethernet rev 10 (p2p1)
Package version: 08.07.26
Installed version: 08.07.26
BCM57810 10 Gigabit Ethernet rev 10 (p2p2)
The version of this Update Package is the same as the currently installed
version.
Software application name: BCM57810 10 Gigabit Ethernet rev 10 (p2p2)
Package version: 08.07.26
Installed version: 08.07.26
Continue? Y/N:Y
Y entered; update was forced by user
Executing update...
WARNING: DO NOT STOP THIS PROCESS OR INSTALL OTHER DELL PRODUCTS WHILE UPDATE
IS IN PROGRESS.
THESE ACTIONS MAY CAUSE YOUR SYSTEM TO BECOME UNSTABLE!
.........................................................................
Device: BCM57810 10 Gigabit Ethernet rev 10 (p2p1)
  Application: BCM57810 10 Gigabit Ethernet rev 10 (p2p1)
  Update success.
Device: BCM57810 10 Gigabit Ethernet rev 10 (p2p2)
  Application: BCM57810 10 Gigabit Ethernet rev 10 (p2p2)
  Update success.
Would you like to reboot your system now?
Continue? Y/N:Y
```

5 Adapter Preboot Configuration

During the host boot process, you have the opportunity to pause and perform adapter management tasks using the Human Infrastructure Interface (HII) application. These tasks include the following:

- Displaying Firmware Image Properties
- Configuring Device-level Parameters
- Configuring NIC Parameters
- Configuring Data Center Bridging
- Configuring Partitions

NOTE

The HII screen shots in this chapter are representative and may not match the screens that you see on your system.

Getting Started

To start the HII application:

1. Open the System Setup window for your platform. For information about launching the System Setup, consult the user guide for your system.
2. In the System Setup window (Figure 5-1), select **Device Settings**, and then press ENTER.



Figure 5-1. System Setup

3. In the Device Settings window (Figure 5-2), select the QL45212 Adapter port that you want to configure, and then press ENTER.

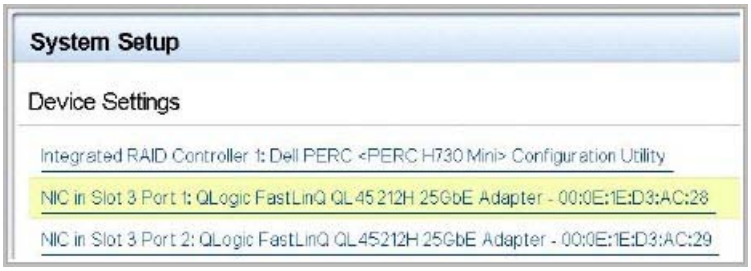


Figure 5-2. System Setup: Device Settings

4. The Main Configuration Page presents the adapter management options where you can set the partitioning mode.

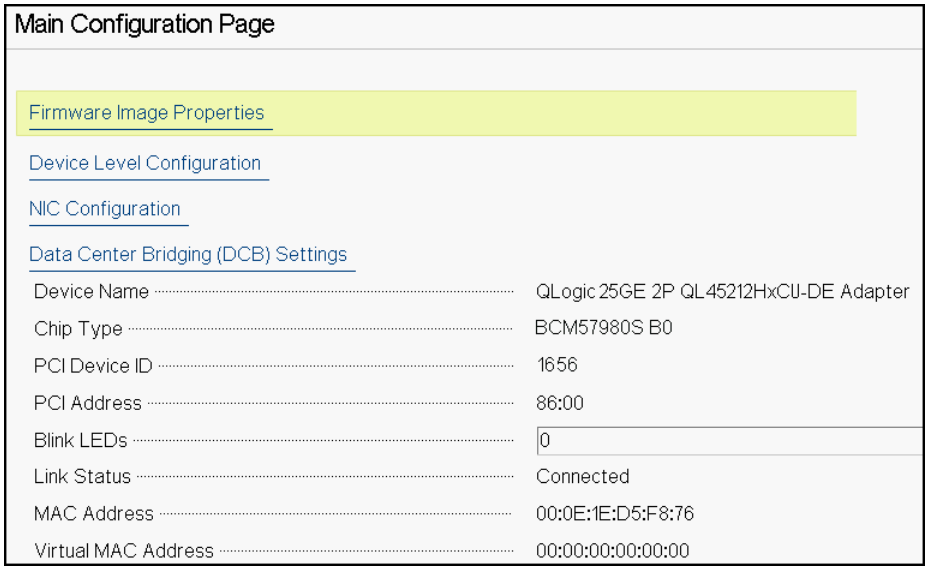


Figure 5-3. Main Configuration Page

- ❑ Under **Device Level Configuration**, set the **Partitioning Mode** to **NPAR** to add the **NIC Partitioning Configuration** option to the Main Configuration Page, as shown in Figure 5-4.

NOTE

NPAR is not available on ports with a maximum speed of 1G.

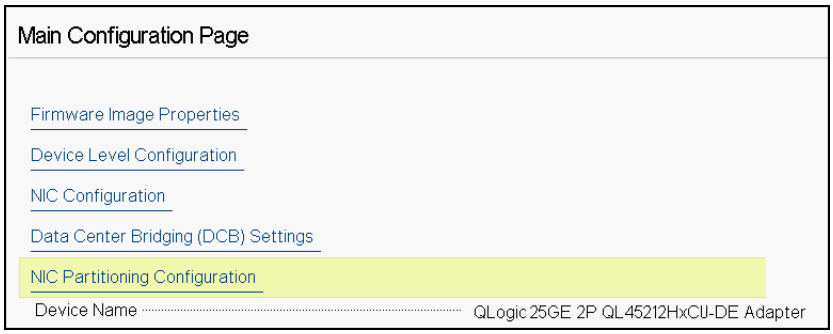


Figure 5-4. Main Configuration Page, Setting Partitioning Mode to NPAR

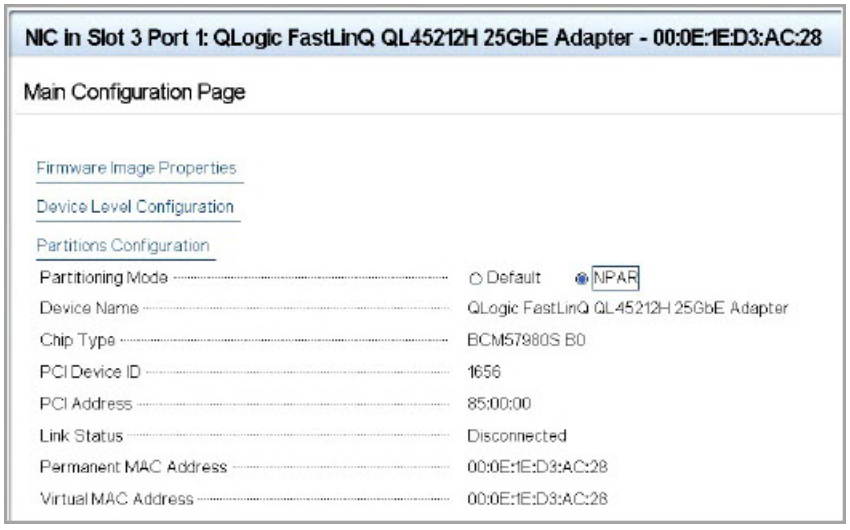


Figure 5-5. Main Configuration Page, Setting NPAR Partitioning Mode

In Figure 5-3 and Figure 5-4, the Main Configuration Page shows the following:

- **Firmware Image Properties** (see "Displaying Firmware Image Properties" on page 44)
- **Device Level Configuration** (see "Configuring Device-level Parameters" on page 45)

- **NIC Configuration** (see “Configuring NIC Parameters” on page 46)
- Data Center Bridging (DCB) Settings (see “Configuring Data Center Bridging” on page 49)
- **NIC Partitioning Configuration** (if **NPAR** is selected on the Device Level Configuration page) (see “Configuring Partitions” on page 51)

In addition, the Main Configuration Page presents the adapter properties listed in Table 5-1.

Table 5-1. Adapter Properties

Adapter Property	Description
Device Name	Factory-assigned device name
Chip Type	ASIC version
PCI Device ID	Unique vendor-specific PCI device ID
PCI Address	PCI device address in bus-device function format
Blink LEDs	User-defined blink count for the port LED
Link Status	External link status
MAC Address	Manufacturer-assigned permanent device MAC address
Virtual MAC Address	User-defined device MAC address

Displaying Firmware Image Properties

To view the properties for the firmware image, select **Firmware Image Properties** on the Main Configuration Page, and then press ENTER. The Firmware Image Properties page (Figure 5-6) specifies the following view-only data:

- **Family Firmware Version** is the multiboot image version, which comprises several firmware component images.
- **MBI Version** is the Cavium QLogic bundle image version that is active on the device.
- **Controller BIOS Version** is the management firmware version.
- **EFI Driver Version** is the extensible firmware interface (EFI) driver version.
- **L2B Firmware Version** is the NIC offload firmware version for boot.

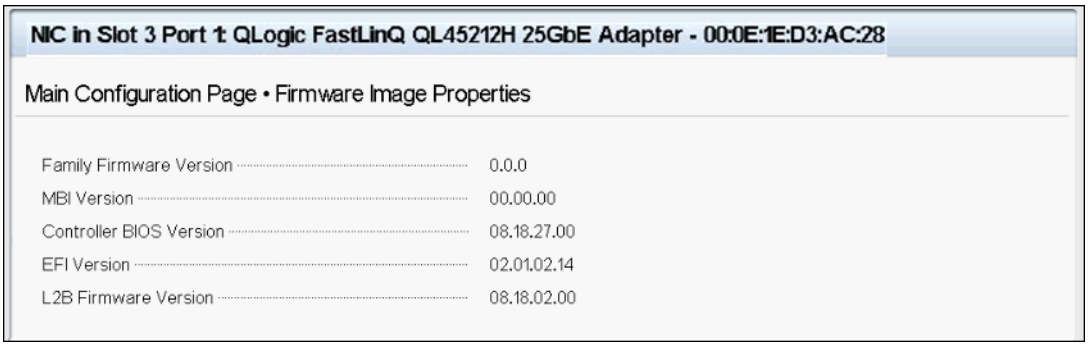


Figure 5-6. Firmware Image Properties

Configuring Device-level Parameters

Device-level configuration comprises enabling single root-I/O virtualization (SR-IOV) and NIC partitioning, and enabling or disabling NPAReP. To perform device-level configuration, select **Device Level Configuration** on the Main Configuration Page, and then click **Finish**. Figure 5-7 shows the Device Level Configuration page.

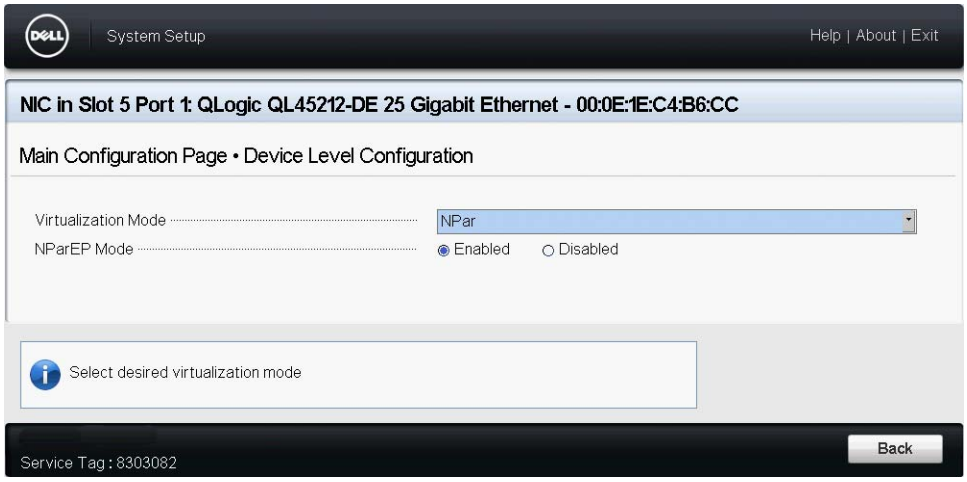


Figure 5-7. System Setup: Device Level Configuration Page

To enable NIC partitioning (NPAR), single root input/output virtualization (SR-IOV), or both:

1. On the Main Configuration Page (Figure 5-3 on page 42), select **Device Level Configuration**, and then click **Finish**.

2. On the Device Level Configuration page (Figure 5-7), click **Virtualization Mode**, and then select from the following options:
   - ❑ **NPAR** enables NPAR.
   - ❑ **SRIOV** enables SR-IOV (non-partitioned virtualization).
   - ❑ **NPAR+SRIOV** enables NPAR with SR-IOV.
   - ❑ **None** disables all virtualization.
3. If NPAR is enabled (with or without SR-IOV), and the system is capable of alternate routing ID interpretation (ARI), choose to enable or disable **NPAReP mode**:
   - ❑ Click **Disabled** to specify four partitions per port (that is, 8 total PFs per adapter), where each PF supports 16 SR-IOV VFs per PF.
   - ❑ Click **Enabled** to specify eight partitions per port (that is, 16 total PFs per adapter), where:
     - ■ 16 SR-IOV VFs are on the first 8 PFs (PFs 0– 7, which are the first 4 PFs of each port).
     - ■ 8 SR-IOV VFs are on the second 8 PFs (PFs 8–15, which are the last 4 PFs of each port).
     - ■ SR-IOV VFs are currently hard-allocated to their respective PFs.
4. Click **Back**.
5. When prompted, click **Yes** to save the changes. Changes will take effect after a system reset.

Configuring NIC Parameters

NIC configuration includes setting the following parameters:

- ■ **Link Speed**
- ■ **NIC + RDMA Mode**
- ■ **Boot Mode**
- ■ **FEC Mode**
- ■ **PXE Virtual LAN Mode**
- ■ **Virtual LAN ID**

To configure NIC parameters:

1. On the Main Configuration Page, select **NIC Configuration** (Figure 5-3 on page 42), and then click **Finish**.

   Figure 5-8 shows the NIC Configuration page.

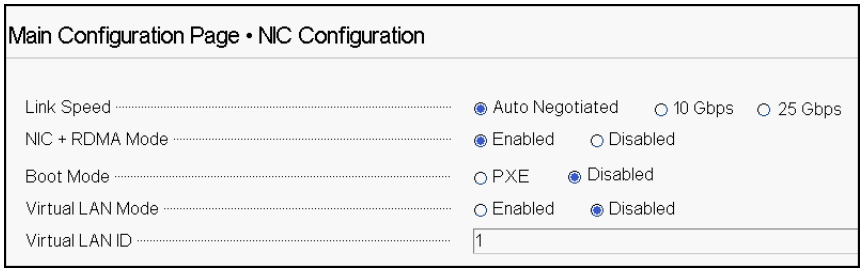


Figure 5-8. NIC Configuration

2. Select one of the following **Link Speed** options. The link speed option you choose applies to both adapter ports. The link speed and the forward error correction (FEC) must match that of the connected switch or device port.

   - **Auto Negotiated** automatically negotiates the link parameters with the connected switch or device (default). FEC is automatically enabled. Both ends must be in Auto Negotiation mode.
   - **10 Gbps** specifies the port fixed link speed at 10Gbps. FEC is not supported.
   - **25 Gbps** specifies the port fixed link speed at 25Gbps.

3. For **NIC + RDMA Mode**, select either **Enabled** or **Disabled** for RDMA on the port. This setting applies to all partitions of the port, if in NPAR mode.

4. **FEC Mode** is visible when **25 Gbps** fixed speed mode is selected as the **Link Speed** in Step 2. For **FEC Mode**, select one of the following options. Not all FEC modes are available on all adapters.

   - **Disabled** indicates that FEC is not enabled.
   - **Fire Code** indicates that the link is configured to use Fire Code FEC; the FEC sublayer operates similarly to Clause 74 FEC.

5. For **Boot Mode**, select one of the following values:

   - **PXE** enables PXE boot.
   - **Disabled** prevents this port from being used as a remote boot source.

6. The **Virtual LAN Mode** parameter applies to the entire port when in PXE remote install mode. It is not persistent after a PXE remote install finishes. Select from the following VLAN options:
   - ❑ **Enabled** enables VLAN mode on this port for PXE remote install mode.
   - ❑ **Disabled** disables VLAN mode on this port.
7. The **Virtual LAN ID** parameter specifies the VLAN tag ID to be used on this port for PXE remote install mode. This setting applies only when **Virtual LAN Mode** is enabled in the previous step.
8. Click **Back**.
9. When prompted, click **Yes** to save the changes. Changes take effect after a system reset.

To configure the port to use RDMA:

NOTE

Follow these steps to enable RDMA on all partitions of an NPAR mode port.

1. Set **NIC + RDMA Mode** to **Enabled**.
2. Click **Back**.
3. When prompted, click **Yes** to save the changes. Changes take effect after a system reset.

To configure the port's boot mode:

1. For a UEFI PXE remote installation, select **PXE** as the **Boot Mode**.
2. Click **Back**.
3. When prompted, click **Yes** to save the changes. Changes take effect after a system reset.

To configure the port's PXE remote install to use a VLAN:

NOTE

This VLAN is not persistent after the PXE remote install is finished.

1. Set the **Virtual LAN Mode** to **Enabled**.
2. In the **Virtual LAN ID** box, enter the number to be used.
3. Click **Back**.

4. When prompted, click **Yes** to save the changes. Changes take effect after a system reset.

Configuring Data Center Bridging

The data center bridging (DCB) settings comprise the DCBX protocol and the RoCE priority.

To configure the DCB settings:

1. On the Main Configuration Page (Figure 5-3 on page 42), select **Data Center Bridging (DCB) Settings**, and then click **Finish**.
2. On the Data Center Bridging (DCB) Settings page (Figure 5-9), select the appropriate **DCBX Protocol** option:
   - **Disabled** disables DCBX on this port.
   - **CEE** enables the legacy Converged Enhanced Ethernet (CEE) protocol DCBX mode on this port.
   - **IEEE** enables the IEEE DCBX protocol on this port.
   - **Dynamic** enables dynamic application of either the CEE or IEEE protocol to match the attached link partner.
3. On the Data Center Bridging (DCB) Settings page, enter the **RoCE v1 Priority** as a value from **0–7**. This setting indicates the DCB traffic class priority number used for RoCE traffic and should match the number used by the DCB-enabled switching network for RoCE traffic.
   - **0** specifies the usual priority number used by the lossy default or common traffic class.
   - **3** specifies the priority number used by lossless FCoE traffic.
   - **4** specifies the priority number used by lossless iSCSI-TLV over DCB traffic.
   - **1**, **2**, **5**, **6**, and **7** specify DCB traffic class priority numbers available for RoCE use. Follow the respective OS RoCE setup instructions for using this RoCE control.

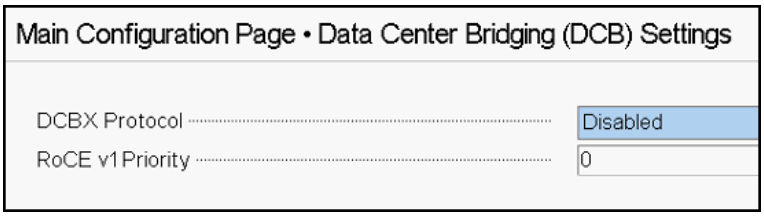


Figure 5-9. System Setup: Data Center Bridging (DCB) Settings

4. Click **Back**.
5. When prompted, click **Yes** to save the changes. Changes take effect after a system reset.

NOTE

When DCBX is enabled, the adapter periodically sends link layer discovery protocol (LLDP) packets with a dedicated unicast address that serves as the source MAC address. This LLDP MAC address is different from the factory-assigned adapter Ethernet MAC address. If you examine the MAC address table for the switch port that is connected to the adapter, you will see two MAC addresses: one for LLDP packets and one for the adapter Ethernet interface.

Configuring Partitions

You can configure bandwidth ranges for each partition on the adapter.

To configure the maximum and minimum bandwidth allocations:

1. On the Main Configuration Page, select **NIC Partitioning Configuration**, and then press ENTER.
2. On the Partitions Configuration page (Figure 5-10), select **Global Bandwidth Allocation**.

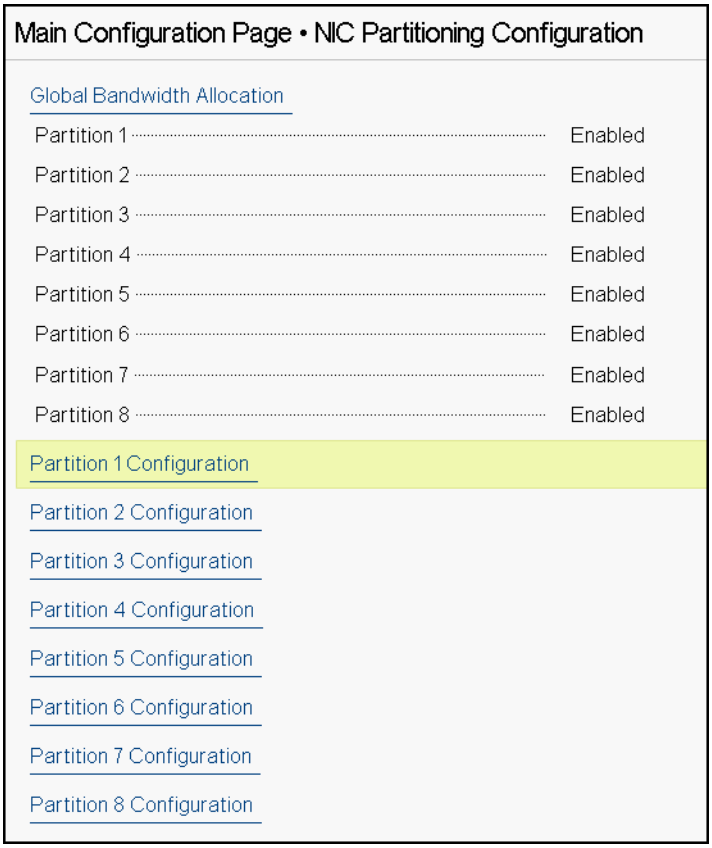


Figure 5-10. NIC Partitioning Configuration, Global Bandwidth Allocation

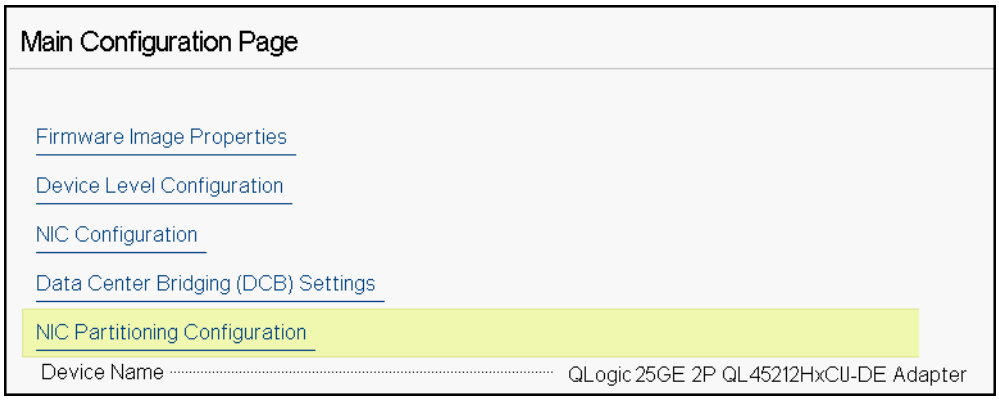


Figure 5-11. Global Bandwidth Allocation, Partition 1 Configuration

3. On the Global Bandwidth Allocation page (Figure 5-12), click each partition minimum and maximum TX bandwidth field for which you want to allocate bandwidth. There are four partitions per port if NPAReP mode is disabled and eight partitions per port if NPAReP mode is enabled. NPAReP mode requires that ARI be enabled in the BIOS for that adapter's PCI slot.

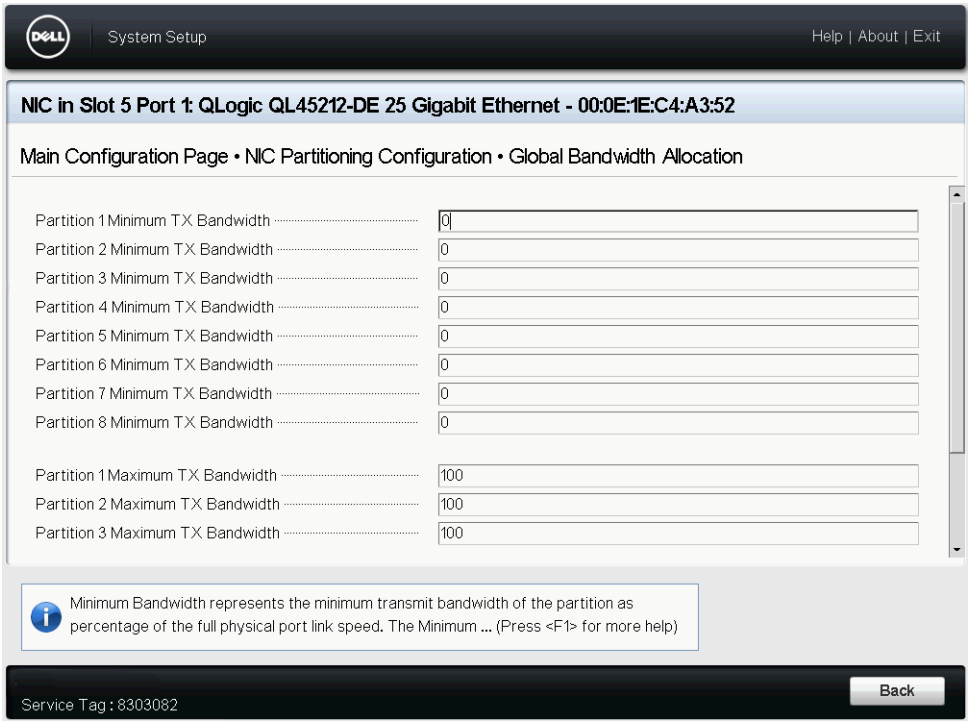


Figure 5-12. Global Bandwidth Allocation Page: NPAReP Mode Enabled

- **Partition *n* Minimum TX Bandwidth** is the minimum transmit bandwidth of the selected partition expressed as a percentage of the maximum physical port link speed. Values can be `0–100`. When DCBX ETS mode is enabled, the per-traffic class DCBX ETS minimum bandwidth value is used simultaneously with the per-partition minimum TX bandwidth value.The total of the minimum TX bandwidth values of all partitions on a single port must equal 100 or be all zeros.

  Setting the TX bandwidth to all zeros is similar to equally dividing the available bandwidth over every active partition; however, the bandwidth is dynamically allocated over all actively sending partitions. A zero value (when one or more of the other values are set to a non-zero value) allocates a minimum of one percent to that partition, when congestion (from all of the partitions) is restricting TX bandwidth.

- **Partition *n* Maximum TX Bandwidth** is the maximum transmit bandwidth of the selected partition expressed as a percentage of the maximum physical port link speed. Values can be `1–100`. The per-partition maximum TX bandwidth value applies regardless of the DCBX ETS mode setting.

Type a value in each selected field, and then click **Back**.

4. When prompted, click **Yes** to save the changes. Changes take effect after a system reset.

To configure partitions:

1. To examine a specific partition configuration, on the NIC Partitioning Configuration page (Figure 5-10 on page 51), select **Partition *n* Configuration**. If NParEP is not enabled, only four partitions exist per port.

2. To configure the first partition, select **Partition 1 Configuration** to open the Partition 1 Configuration page (Figure 5-13), which shows the following parameters:
   - **NIC Mode** (always enabled)
   - **PCI Device ID**
   - **PCI** (bus) **Address**
   - **MAC Address**
   - **Virtual MAC Address**

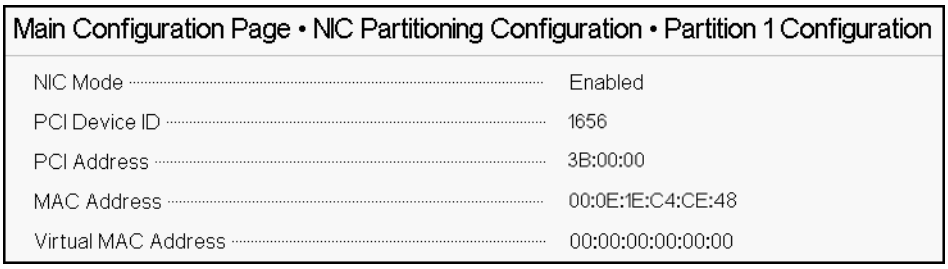


Figure 5-13. Partition 1 Configuration

3. To configure the second partition, select **Partition 2 Configuration** to open the Partition 2 Configuration page. If FCoE Offload is present, the Partition 2 Configuration (Figure 5-14) shows the following parameters:
   - **NIC Mode** enables or disables the L2 Ethernet NIC personality on Partitions 2 and greater. To disable any of the remaining partitions, set the **NIC Mode** to **Disabled**. To disable offload-capable partitions, disable both the **NIC Mode** and respective offload mode.
   - **PCI Device ID**
   - **PCI** (bus) **Address**
   - **MAC Address**
   - **Virtual MAC Address**

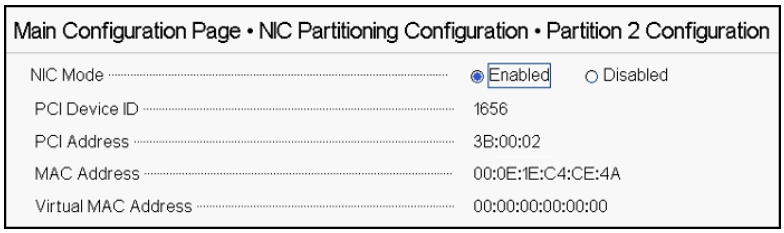


Figure 5-14. Partition 2 Configuration: FCoE Offload

6 RoCE Configuration

This chapter describes RDMA over converged Ethernet (RoCE v1 and v2) configuration on the QL45212 Adapter, the Ethernet switch, and the Windows or Linux host, including:

- Supported Operating Systems and OFED
- Planning for RoCE
- Preparing the Adapter
- Preparing the Ethernet Switch
- Configuring RoCE on the Adapter for Windows Server
- Configuring RoCE on the Adapter for Linux
- Configuring RoCE on the Adapter for ESX

NOTE

Some RoCE features may not be fully enabled in the current release.

Supported Operating Systems and OFED

Table 6-1 shows the operating system support for RoCE v1, RoCE v2, and OFED.

Table 6-1. OS Support for RoCE v1, RoCE v2, iWARP, and OFED

Operating System	Inbox	OFED 3.18-3 GA	OFED-4.8-1 GA
Windows Server 2012 R2	RoCE v1, RoCE v2	No	No
Windows Server 2016	RoCE v1, RoCE v2	No	No
RHEL 6.8	RoCE v1, iWARP	RoCE v1, iWARP	No
RHEL 6.9	RoCE v1, iWARP	No	No
RHEL 7.3	RoCE v1, RoCE v2, iWARP, iSER	No	RoCE v1, RoCE v2, iWARP
RHEL 7.4	RoCE v1, RoCE v2, iWARP, iSER	No	No
SLES 12 SP3	RoCE v1, RoCE v2, iWARP, iSER	No	No
CentOS 7.3	RoCE v1, RoCE v2, iWARP, iSER	No	RoCE v1, RoCE v2, iWARP
CentOS 7.4	RoCE v1, RoCE v2, iWARP, iSER	No	No
VMware ESXi 6.0 u3	No	N/A	
VMware ESXi 6.5, 6.5U1 [a]	RoCE v1, RoCE v2	N/A	

[a] The certified RoCE driver is not included in this release. The uncertified driver is available as an early preview.

Planning for RoCE

As you prepare to implement RoCE, consider the following limitations:

- If you are using the inbox OFED, the operating system should be the same on the server and client systems. Some of the applications may work between different operating systems, but there is no guarantee. This is an OFED limitation.
- For OFED applications (most often perftest applications), server and client applications should use the same options and values. Problems can arise if the operating system and the perftest application have different versions. To confirm the perftest version, issue the following command:

  ```
  # ib_send_bw --version
  ```

- Building libqedr in inbox OFED requires installing libibverbs-devel.
- Running user space applications in inbox OFED requires installing the InfiniBand® Support group, by yum groupinstall “InfiniBand Support” that contains libibcm, libibverbs, and more.
- OFED and RDMA applications that depend on libibverbs also require the QLogic RDMA user space library, libqedr. Install libqedr using the libqedr RPM or source packages.
- RoCE supports only little endian.
- RoCE does not work over a VF in an SR-IOV environment.

Preparing the Adapter

Follow these steps to enable DCBX and specify the RoCE priority using the HII management application. For information about the HII application, see Chapter 5 Adapter Preboot Configuration.

To prepare the adapter:

1. In the Main Configuration Page, select **Data Center Bridging (DCB) Settings**, and then click **Finish**.
2. In the Data Center Bridging (DCB) Settings window, click the **DCBX Protocol** option. The QL45212 Adapter supports both CEE and IEEE protocols. This value should match the corresponding value on the DCB switch. In this example, select **CEE** or **Dynamic**.
3. In the **RoCE Priority** box, type a priority value. This value should match the corresponding value on the DCB switch. In this example, type 5. Typically, 0 is used for the default lossy traffic class, 3 is used for the FCoE traffic class, and 4 is used for lossless iSCSI-TLV over DCB traffic class.
4. Click **Back**.
5. When prompted, click **Yes** to save the changes. Changes will not take effect until after a system reset.

   For Windows, you can configure DCBX using the HII or QoS method.The configuration shown in this section is through HII. For QoS, refer to “Configuring QoS for RoCE” on page 110“.

Preparing the Ethernet Switch

This section describes how to configure a Cisco® Nexus® 6000 Ethernet Switch and a Dell Z9100 Ethernet Switch for RoCE.

- Configuring the Cisco Nexus 6000 Ethernet Switch
- Configuring the Dell Z9100 Ethernet Switch

Configuring the Cisco Nexus 6000 Ethernet Switch

Steps for configuring the Cisco Nexus 6000 Ethernet Switch for RoCE include configuring class maps, configuring policy maps, applying the policy, and assigning a VLAN ID to the switch port.

To configure the Cisco switch:

1. Open a config terminal session as follows:

```
Switch# config terminal
switch(config)#
```

2. Configure quality of service (QoS) class map and set the RoCE priority to match the adapter (5) as follows:

```
switch(config)# class-map type qos class-roce
switch(config)# match cos 5
```

3. Configure queuing class maps as follows:

```
switch(config)# class-map type queuing class-roce
switch(config)# match qos-group 3
```

4. Configure network QoS class maps as follows:

```
switch(config)# class-map type network-qos class-roce
switch(config)# match qos-group 3
```

5. Configure QoS policy maps as follows:

```
switch(config)# policy-map type qos roce
switch(config)# class type qos class-roce
switch(config)# set qos-group 3
```

6. Configure queuing policy maps to assign network bandwidth. In this example, use a value of 50 percent:

```
switch(config)# policy-map type queuing roce
switch(config)# class type queuing class-roce
switch(config)# bandwidth percent 50
```

7. Configure network QoS policy maps to set priority flow control for no-drop traffic class as follows:

```
switch(config)# policy-map type network-qos roce
switch(config)# class type network-qos class-roce
switch(config)# pause no-drop
```

8. Apply the new policy at the system level as follows

```
switch(config)# system qos
switch(config)# service-policy type qos input roce
switch(config)# service-policy type queuing output roce
switch(config)# service-policy type queuing input roce
switch(config)# service-policy type network-qos roce
```

9. Assign a VLAN ID to the switch port to match the VLAN ID assigned to the adapter (5).

```
switch(config)# interface ethernet x/x
switch(config)# switchport mode trunk
switch(config)# switchport trunk allowed vlan 1,5
```

Configuring the Dell Z9100 Ethernet Switch

To configure the Dell Z9100 Ethernet Switch for RoCE, see the procedure in Appendix C Dell Z9100 Switch Configuration.

Configuring RoCE on the Adapter for Windows Server

Configuring RoCE on the adapter for Windows Server comprises enabling RoCE on the adapter and verifying the Network Direct MTU size.

To configure RoCE on a Windows Server host:

1. Enable RoCE on the adapter.
   a. Open the Windows Device Manager, and then open the QL45212 Adapters NDIS Miniport Properties.
   b. On the QLogic FastLinQ Adapter Properties, click the **Advanced** tab.
   c. On the Advanced page, configure the properties listed in Table 6-2 by selecting each item under **Property** and choosing an appropriate **Value** for that item. Then click **OK**.

Table 6-2. Advanced Properties for RoCE

Property	Value or Description
Network Direct Functionality	Enabled
Network Direct Mtu Size	The network direct MTU size must be less than the jumbo packet size.
RDMA Mode	**RoCE v1** or **RoCE v2**.
VLAN ID	Assign any VLAN ID to the interface. The value must be the same as is assigned on the switch.

Figure 6-1 shows an example of configuring a property value.

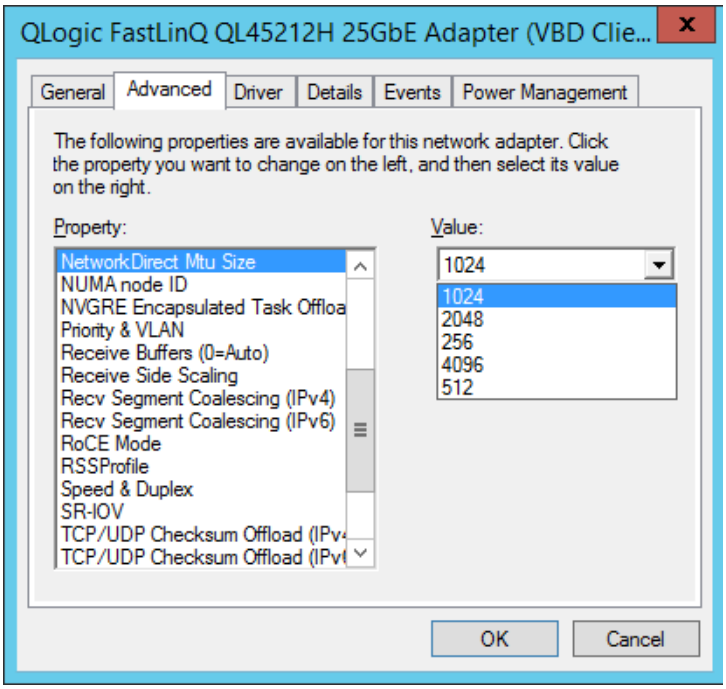


Figure 6-1. Configuring RoCE Properties

2. Using Windows PowerShell, verify that RDMA is enabled on the adapter. The `Get-NetAdapterRdma` command lists the adapters that support RDMA—both ports are enabled.

NOTE

If you are configuring RoCE over Hyper-V, do not assign a VLAN ID to the physical interface.

```
PS C:\Users\Administrator> Get-NetAdapterRdma
Name                InterfaceDescription        Enabled
-----               --------------------        -------
SLOT 4 3 Port 1     QLogic FastLinQ QL45212...  True
SLOT 4 3 Port 2     QLogic FastLinQ QL45212...  True
```

3. Using Windows PowerShell, verify that `NetworkDirect` is enabled on the host operating system. The `Get-NetOffloadGlobalSetting` command shows `NetworkDirect` is enabled.

```
PS C:\Users\Administrators> Get-NetOffloadGlobalSetting
ReceiveSideScaling            : Enabled
ReceiveSegmentCoalescing      : Enabled
Chimney                       : Disabled
TaskOffload                   : Enabled
NetworkDirect                 : Enabled
NetworkDirectAcrossIPSubnets  : Blocked
PacketCoalescingFilter        : Disabled
```

4. Connect a server message block (SMB) drive, run RoCE traffic, and verify the results.

   To set up and connect to an SMB drive, view the information available online from Microsoft:

   https://technet.microsoft.com/en-us/library/hh831795(v=ws.11).aspx

5. By default, Microsoft's SMB Direct establishes two RDMA connections per port, which provides good performance, including line rate at a higher block size (for example, 64KB). To optimize performance, you can change the quantity of RDMA connections per RDMA interface to four (or greater).

   To increase the quantity of RDMA connections to four (or more), issue the following command in Windows PowerShell:

```
PS C:\Users\Administrator> Set-ItemProperty -Path
"HKLM:\SYSTEM\CurrentControlSet\Services\LanmanWorkstation\
Parameters" ConnectionCountPerRdmaNetworkInterface -Type
DWORD -Value 4 –Force
```

Configuring RoCE on the Adapter for Linux

This section describes the RoCE configuration procedure for RHEL and SLES. It also describes how to verify the RoCE configuration and provides some guidance about using group IDs (GIDs) with VLAN interfaces.

- RoCE Configuration for RHEL
- RoCE Configuration for SLES
- Verifying the RoCE Configuration on Linux
- VLAN Interfaces and GID Index Values
- RoCE v2 Configuration for Linux

RoCE Configuration for RHEL

To configure RoCE on the adapter, the Open Fabrics Enterprise Distribution (OFED) must be installed and configured on the RHEL host.

To prepare inbox OFED for RHEL:

1. While installing or upgrading the operating system, select the InfiniBand and OFED support packages.
2. Install the following RPMs from the RHEL ISO image:

   `libibverbs-devel-x.x.x.x86_64.rpm`
   (required for libqedr library)

   `perftest-x.x.x.x86_64.rpm`
   (required for InfiniBand bandwidth and latency applications)

   or, using Yum, install the inbox OFED:

   ```
   yum groupinstall "Infiniband Support"
   yum install perftest
   yum install tcl tcl-devel tk zlib-devel libibverbs
   libibverbs-devel
   ```

   NOTE

   During installation, if you already selected the previously mentioned packages, you need not reinstall them. The inbox OFED and support packages may vary depending on the operating system version.

3. Install the new Linux drivers as described in “Installing the Linux Drivers with RDMA” on page 15.

RoCE Configuration for SLES

To configure RoCE on the adapter for an SLES host, OFED must be installed and configured on the SLES host.

To install inbox OFED for SLES Linux:

1. While installing or upgrading the operating system, select the InfiniBand support packages.
2. Install the following RPMs from the corresponding SLES SDK kit image:

   `libibverbs-devel-x.x.x.x86_64.rpm`
   (required for libqedr installation)

   `perftest-x.x.x.x86_64.rpm`
   (required for bandwidth and latency applications)

3. Install the Linux drivers, as described in “Installing the Linux Drivers with RDMA” on page 15.

Verifying the RoCE Configuration on Linux

After installing OFED, installing the Linux driver, and loading the RoCE drivers, verify that the RoCE devices were detected on all Linux operating systems.

To verify RoCE configuration on Linux:

1. Stop firewall tables using `service/systemctl` commands.
2. For RHEL only: If the RDMA service is installed (`yum install rdma`), verify that the RDMA service has started.

NOTE

For RHEL 6.x and SLES 11 SP4, you must start RDMA service after reboot. For RHEL 7.*x* and SLES 12 SPX and later, RDMA service starts itself after reboot.

On RHEL or CentOS: Use the `service rdma` status command to start service:

- ❑ If RDMA has not started, issue the following command:

```
# service rdma start
```

- ❑ If RDMA does not start, issue either of the following alternative commands:

```
# /etc/init.d/rdma start
```

or

```
# systemctl start rdma.service
```

3. Verify that the RoCE devices were detected by examining the dmesg logs:

```
# dmesg|grep qedr
  [87910.988411] qedr: discovered and registered 2 RoCE funcs
```

4. Verify that all of the modules have been loaded. For example:

```
# lsmod|grep qedr
  qedr                   89871  0
  qede                   96670  1 qedr
  qed                    2075255  2 qede,qedr
  ib_core                88311  16  qedr, rdma_cm, ib_cm,
                         ib_sa,iw_cm,xprtrdma,ib_mad,ib_srp,
```

```
                                ib_ucm,ib_iser,ib_srpt,ib_umad,
                                ib_uverbs,rdma_ucm,ib_ipoib,ib_isert
```

5. Configure the IP address and enable the port using a configuration method such as ifconfig:

   ```
   # ifconfig ethX 192.168.10.10/24 up
   ```

6. Issue the `ibv_devinfo` command. For each PCI function, you should see a separate `hca_id`, as shown in the following example:

   ```
   root@captain:~# ibv_devinfo
   hca_id: qedr0
           transport:                      InfiniBand (0)
           fw_ver:                         8.3.9.0
           node_guid:                      020e:1eff:fe50:c7c0
           sys_image_guid:                 020e:1eff:fe50:c7c0
           vendor_id:                      0x1077
           vendor_part_id:                 5684
           hw_ver:                         0x0
           phys_port_cnt:                  1
                   port:   1
                          state:                  PORT_ACTIVE (1)
                           max_mtu:                4096 (5)
                           active_mtu:             1024 (3)
                           sm_lid:                 0
                           port_lid:               0
                           port_lmc:               0x00
                           link_layer:             Ethernet
   ```

7. Verify the L2 and RoCE connectivity between all servers: one server acts as a server, another acts as a client.

   - ❑ Verify the L2 connection using a simple `ping` command.
   - ❑ Verify the RoCE connection by performing an RDMA ping on the server or client:

     On the server, issue the following command:

     ```
     ibv_rc_pingpong -d <ib-dev> -g 0
     ```

     On the client, issue the following command:

     ```
     ibv_rc_pingpong -d <ib-dev> -g 0 <server L2 IP address>
     ```

The following are examples of successful ping pong tests on the server and the client.

Server Ping:

```
root@captain:~# ibv_rc_pingpong -d qedr0 -g 0
local address:  LID 0x0000, QPN 0xff0000, PSN 0xb3e07e, GID
fe80::20e:1eff:fe50:c7c0
remote address: LID 0x0000, QPN 0xff0000, PSN 0x934d28, GID
fe80::20e:1eff:fe50:c570
8192000 bytes in 0.05 seconds = 1436.97 Mbit/sec
1000 iters in 0.05 seconds = 45.61 usec/iter
```

Client Ping:

```
root@lambodar:~# ibv_rc_pingpong -d qedr0 -g 0 192.168.10.165
local address:  LID 0x0000, QPN 0xff0000, PSN 0x934d28, GID
fe80::20e:1eff:fe50:c570
remote address: LID 0x0000, QPN 0xff0000, PSN 0xb3e07e, GID
fe80::20e:1eff:fe50:c7c0
8192000 bytes in 0.02 seconds = 4211.28 Mbit/sec
1000 iters in 0.02 seconds = 15.56 usec/iter
```

- To display RoCE statistics, issue the following commands, where ***X*** is the device number:

  ```
  > mount -t debugfs nodev /sys/kernel/debug
  > cat /sys/kernel/debug/qedr/qedrX/stats
  ```

VLAN Interfaces and GID Index Values

If you are using VLAN interfaces on both the server and the client, you must also configure the same VLAN ID on the switch. If you are running traffic through a switch, the InfiniBand applications must use the correct GID value, which is based on the VLAN ID and VLAN IP address.

Based on the following results, the GID value (-x 4 / -x 5) should be used for any perftest applications.

```
# ibv_devinfo -d qedr0 -v|grep GID
     GID[  0]:  fe80:0000:0000:0000:020e:1eff:fe50:c5b0
     GID[  1]:  0000:0000:0000:0000:0000:ffff:c0a8:0103
     GID[  2]:  2001:0db1:0000:0000:020e:1eff:fe50:c5b0
     GID[  3]:  2001:0db2:0000:0000:020e:1eff:fe50:c5b0
     GID[  4]:   0000:0000:0000:0000:0000:ffff:c0a8:0b03 IP address for VLAN interface
     GID[  5]:  fe80:0000:0000:0000:020e:1e00:0350:c5b0    VLAN ID 3
```

NOTE

The default GID value is zero (0) for back-to-back or pause settings. For server/switch configurations, you must identify the proper GID value. If you are using a switch, refer to the corresponding switch configuration documents for the proper settings.

RoCE v2 Configuration for Linux

To verify RoCE v2 functionality, you must use RoCE v2 supported kernels.

To configure RoCE v2 for Linux:

1. Ensure that you are using one of the following supported kernels:
   - ❑ SLES 12 SP2 GA
   - ❑ RHEL 7.3 GA
2. Configure RoCE v2 as follows:
   a. Identify the GID index for RoCE v2.
   b. Configure the routing address for the server and client.
   c. Enable L3 routing on the switch.

NOTE

You can configure RoCE v1 and RoCE v2 by using RoCE v2-supported kernels. These kernels allow you to run RoCE traffic over the same subnet, as well as over different subnets such as RoCE v2 and any routable environment. Only a few settings are required for RoCE v2, and all other switch and adapter settings are common for RoCE v1 and RoCE v2.

Identifying RoCE v2 GID Index or Address

To find RoCE v1- and RoCE v2-specific GIDs, use either sys or class parameters, or run RoCE scripts from the QL45212 FastLinQ source package. To check the default **RoCE GID Index** and address, issue the `ibv_devinfo` command and compare it with the sys or class parameters. For example:

```
#ibv_devinfo -d qedr0 -v|grep GID
            GID[  0]:           fe80:0000:0000:0000:020e:1eff:fec4:1b20
            GID[  1]:           fe80:0000:0000:0000:020e:1eff:fec4:1b20
            GID[  2]:           0000:0000:0000:0000:0000:ffff:1e01:010a
            GID[  3]:           0000:0000:0000:0000:0000:ffff:1e01:010a
            GID[  4]:           3ffe:ffff:0000:0f21:0000:0000:0000:0004
            GID[  5]:           3ffe:ffff:0000:0f21:0000:0000:0000:0004
```

```
GID[  6]:          0000:0000:0000:0000:0000:ffff:c0a8:6403
GID[  7]:          0000:0000:0000:0000:0000:ffff:c0a8:6403
```

Verifying RoCE v1 or RoCE v2 GID Index and Address from sys and class Parameters

Use one of the following options to verify the RoCE v1 or RoCE v2 GID Index and address from the sys and class parameters:

- **Option 1:**

```
# cat /sys/class/infiniband/qedr0/ports/1/gid_attrs/types/0
IB/RoCE v1
# cat /sys/class/infiniband/qedr0/ports/1/gid_attrs/types/1
RoCE v2

# cat /sys/class/infiniband/qedr0/ports/1/gids/0
fe80:0000:0000:0000:020e:1eff:fec4:1b20
# cat /sys/class/infiniband/qedr0/ports/1/gids/1
fe80:0000:0000:0000:020e:1eff:fec4:1b20
```

- **Option 2:**

  Use the scripts from the FastLinQ source package.

```
#/../fastlinq-8.x.x.x/add-ons/roce/show_gids.sh
DEV   PORT  INDEX  GID                                     IPv4           VER     DEV
---   ----  -----  ---                                     ------------   ---     ---
qedr0  1    0      fe80:0000:0000:0000:020e:1eff:fec4:1b20                v1      p4p1
qedr0  1    1      fe80:0000:0000:0000:020e:1eff:fec4:1b20                v2      p4p1
qedr0  1    2      0000:0000:0000:0000:0000:ffff:1e01:010a  30.1.1.10     v1      p4p1
qedr0  1    3      0000:0000:0000:0000:0000:ffff:1e01:010a  30.1.1.10     v2      p4p1
qedr0  1    4      3ffe:ffff:0000:0f21:0000:0000:0000:0004                v1      p4p1
qedr0  1    5      3ffe:ffff:0000:0f21:0000:0000:0000:0004                v2      p4p1
qedr0  1    6      0000:0000:0000:0000:0000:ffff:c0a8:6403  192.168.100.3 v1      p4p1.100
qedr0  1    7      0000:0000:0000:0000:0000:ffff:c0a8:6403  192.168.100.3 v2      p4p1.100
qedr1  1    0      fe80:0000:0000:0000:020e:1eff:fec4:1b21                v1      p4p2
qedr1  1    1      fe80:0000:0000:0000:020e:1eff:fec4:1b21                v2      p4p2
```

NOTE

You must specify the GID index values for RoCE v1- or RoCE v2-based on server or switch configuration (Pause/PFC). Use GID index for the link local IPv6 address, IPv4 address, or IPv6 address. To use VLAN tagged frames for RoCE traffic, you must specify GID index values that are derived from the VLAN IPv4 or IPv6 address.

Verifying RoCE v1 or RoCE v2 Function Through perftest Applications

This section shows how to verify RoCE v1 or RoCE v2 function through perftest applications. In this example, the following server IP and client IP are used:

- Server IP: 192.168.100.3
- Client IP: 192.168.100.4

Verifying RoCE v1

Run over the same subnet and use the RoCE v1 GID index.

```
Server# ib_send_bw -d qedr0 -F -x 0
Client# ib_send_bw -d qedr0 -F -x 0  192.168.100.3
```

Verifying RoCE v2

Run over the same subnet and use the RoCE v2 GID index.

```
Server# ib_send_bw -d qedr0 -F -x 1
Client# ib_send_bw -d qedr0 -F -x 1  192.168.100.3
```

NOTE

If you are running through a switch PFC configuration, use VLAN GIDs for RoCE v1 or v2 through the same subnet.

Verifying RoCE v2 Through Different Subnets

NOTE

You must first configure the route settings for the switch and servers. On the adapter, set the RoCE priority and DCBX mode using the HII or UEFI user interface.

To verify RoCE v2 through different subnets.

1. Set the route configuration for the server and client using the DCBX-PFC configuration.

   - **System Settings:**

     Server VLAN IP : 192.168.100.3 and **Gateway :**192.168.100.1

     Client VLAN IP : 192.168.101.3 and **Gateway :**192.168.101.1

   - **Server Configuration**:

```
#/sbin/ip link add link p4p1 name p4p1.100 type vlan id 100
#ifconfig p4p1.100 192.168.100.3/24 up
#ip route add 192.168.101.0/24 via 192.168.100.1 dev p4p1.100
```

   - **Client Configuration:**

```
#/sbin/ip link add link p4p1 name p4p1.101 type vlan id 101
#ifconfig p4p1.101 192.168.101.3/24 up
#ip route add 192.168.100.0/24 via 192.168.101.1 dev p4p1.101
```

2. Set the switch settings using the following procedure.

   - Use any flow control method (Pause, DCBX-CEE, or DCBX-IEEE), and enable IP routing for RoCE v2. See “Preparing the Ethernet Switch” on page 58 for RoCE v2 configuration, or refer to the vendor switch documents.

   - If you are using PFC configuration and L3 routing, run RoCE v2 traffic over the VLAN using a different subnet, and use the RoCE v2 VLAN GID index.

```
Server# ib_send_bw -d qedr0 -F -x 5
Client# ib_send_bw -d qedr0 -F -x 5 192.168.100.3
```

Server Switch Settings:

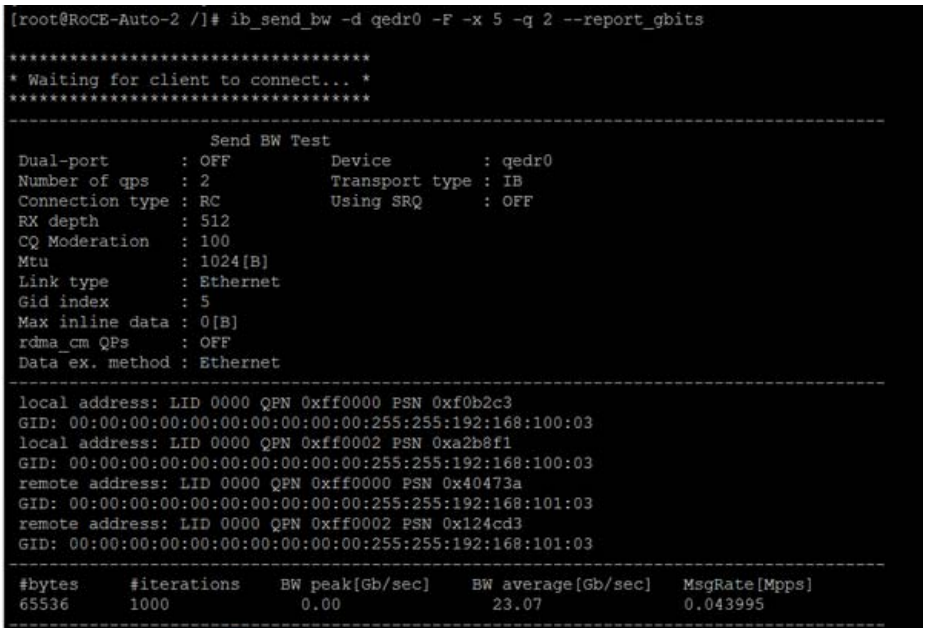


Figure 6-2. Switch Settings, Server

Client Switch Settings:

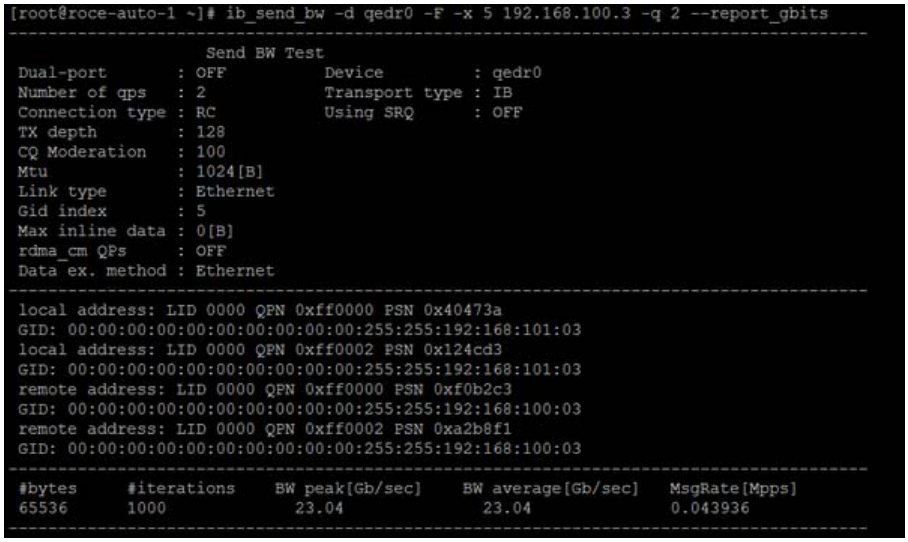


Figure 6-3. Switch Settings, Client

Configuring RoCE v1 or RoCE v2 Settings for RDMA_CM Applications

To configure RoCE, use the following scripts from the FastLinQ source package:

```
# ./show_rdma_cm_roce_ver.sh
qedr0 is configured to IB/RoCE v1
qedr1 is configured to IB/RoCE v1

# ./config_rdma_cm_roce_ver.sh v2
configured rdma_cm for qedr0 to RoCE v2
configured rdma_cm for qedr1 to RoCE v2
```

Server Settings:

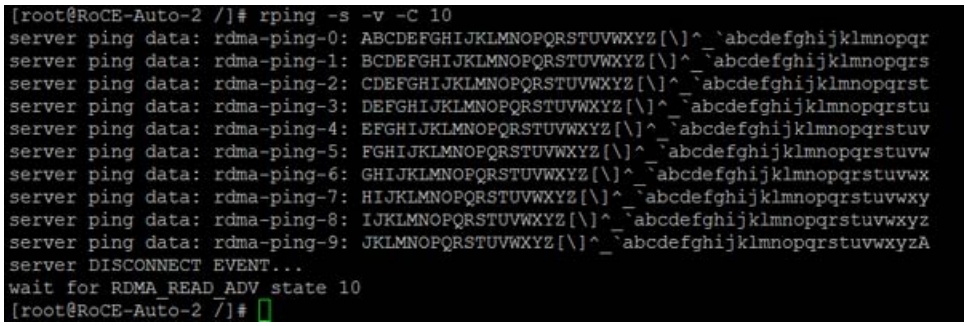


Figure 6-4. Configuring RDMA_CM Applications: Server

Client Settings:



Figure 6-5. Configuring RDMA_CM Applications: Client

Configuring RoCE on the Adapter for ESX

This section provides the following procedures and information for RoCE configuration:

- Configuring RDMA Interfaces
- Configuring MTU
- RoCE Mode and Statistics
- Configuring a Paravirtual RDMA Device (PVRDMA)

Configuring RDMA Interfaces

To configure the RDMA interfaces:

1. Install both QLogic NIC and RoCE drivers.
2. Using the module parameter, enable the RoCE function from the NIC driver by issuing the following command:

   `esxcfg-module -s 'enable_roce=1' qedentv`

   To apply the change, reload the NIC driver or reboot the system.
3. To view a list of the NIC interfaces, issue the `esxcfg-nics -l` command. For example:

```
esxcfg-nics -l
Name    PCI           Driver      Link Speed      Duplex MAC Address       MTU    Description
Vmnic0  0000:01:00.2 qedentv     Up   25000Mbps  Full   a4:5d:36:2b:6c:92 1500   QLogic Corp.
QLogic FastLinQ QL41xxx 1/10/25 GbE Ethernet Adapter
Vmnic1  0000:01:00.3 qedentv     Up   25000Mbps  Full   a4:5d:36:2b:6c:93 1500   QLogic Corp.
QLogic FastLinQ QL41xxx 1/10/25 GbE Ethernet Adapter
```

4. To view a list of the RDMA devices, issue the `esxcli rdma device list` command. For example:

```
esxcli rdma device list
Name     Driver   State    MTU  Speed    Paired Uplink  Description
-------  -------  ------  ----  -------  -------------  ------------------------------
vmrdma0  qedrntv  Active  1024  25 Gbps  vmnic0     QLogic FastLinQ QL45xxx RDMA Interface
vmrdma1  qedrntv  Active  1024  25 Gbps  vmnic1     QLogic FastLinQ QL45xxx RDMA Interface
```

5. To create a new virtual switch, issue the following command:

   `esxcli network vswitch standard add -v <new vswitch name>`

   For example:

   `# esxcli network vswitch standard add -v roce_vs`

   This creates a new virtual switch named *roce_vs*.

6. To associate the QLogic NIC port to the vSwitch, issue the following command:

   # **esxcli network vswitch standard uplink add -u <uplink device> -v <roce vswitch>**

   For example:

   # **esxcli network vswitch standard uplink add -u vmnic0 -v roce_vs**

7. To create a new port group on this vSwitch, issue the following command:

   # **esxcli network vswitch standard portgroup add -p roce_pg -v roce_vs**

   For example:

   # **esxcli network vswitch standard portgroup add -p roce_pg -v roce_vs**

8. To create a vmknic interface on this port group and configure the IP, issue the following command:

   # **esxcfg-vmknic -a -i <IP address> -n <subnet mask> <roce port group name>**

   For example:

   # **esxcfg-vmknic -a -i 192.168.10.20 -n 255.255.255.0 roce_pg**

9. To configure the VLAN ID, issue the following command:

   # **esxcfg-vswitch -v <VLAN ID> -p roce_pg**

   To run RoCE traffic with VLAN ID, configure the VLAN ID on the corresponding VMkernel port group.

Configuring MTU

To modify MTU for RoCE interface, change the MTU of the corresponding vSwitch. Set the MTU size of the RDMA interface based on MTU of the vSwitch by issuing the following command:

esxcfg-vswitch -m <new MTU> <RoCE vswitch name>

For example:

```
# esxcfg-vswitch -m 4000 roce_vs
# esxcli rdma device list
Name     Driver   State    MTU  Speed    Paired Uplink  Description
-------  -------  ------  ----  -------  -------------  ------------------------------
vmrdma0  qedrntv  Active     2048   25 Gbps  vmnic0    QLogic FastLinQ QL45xxx RDMA Interface
vmrdma1  qedrntv  Active     1024   25 Gbps  vmnic1    QLogic FastLinQ QL45xxx RDMA Interface
```

RoCE Mode and Statistics

For the RoCE mode, ESXi requires concurrent support of both RoCE v1 and v2. The decision regarding which RoCE mode to use is made during queue pair creation. The ESXi driver advertises both modes during registration and initialization. To view RoCE statistics, issue the following command:

```
# esxcli rdma device stats get -d vmrdma0
   Packets received: 0
   Packets sent: 0
   Bytes received: 0
   Bytes sent: 0
   Error packets received: 0
   Error packets sent: 0
   Error length packets received: 0
   Unicast packets received: 0
   Multicast packets received: 0
   Unicast bytes received: 0
   Multicast bytes received: 0
   Unicast packets sent: 0
   Multicast packets sent: 0
   Unicast bytes sent: 0
   Multicast bytes sent: 0
   Queue pairs allocated: 0
   Queue pairs in RESET state: 0
   Queue pairs in INIT state: 0
   Queue pairs in RTR state: 0
   Queue pairs in RTS state: 0
   Queue pairs in SQD state: 0
   Queue pairs in SQE state: 0
   Queue pairs in ERR state: 0
   Queue pair events: 0
   Completion queues allocated: 1
   Completion queue events: 0
   Shared receive queues allocated: 0
   Shared receive queue events: 0
   Protection domains allocated: 1
   Memory regions allocated: 3
   Address handles allocated: 0
   Memory windows allocated: 0
```

Configuring a Paravirtual RDMA Device (PVRDMA)

To configure PVRDMA using vCenter interface:

1. Create and configure a new distributed virtual switch as follows:
   a. In the VMware vSphere Web Client, right-click the **RoCE** node in the left pane of the Navigator window.
   b. On the Actions menu, point to **Distributed Switch**, and then click **New Distributed Switch**.
   c. Select version 6.5.0.
   d. Under **New Distributed Switch**, click **Edit settings**, and then configure the following:
      - **Number of uplinks**. Select an appropriate value.
      - **Network I/O Control**. Select **Disabled**.
      - **Default port group**. Select this check box.
      - **Port group name**. Enter a name for the port group.

      Figure 6-6 shows an example.

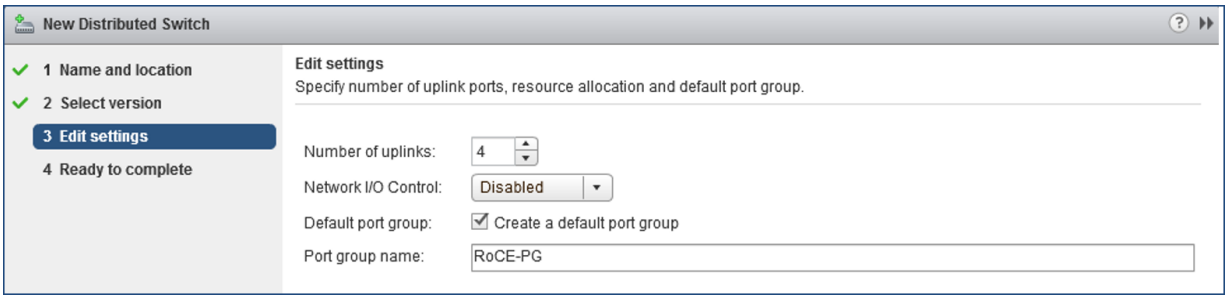


Figure 6-6. Configuring a New Distributed Switch

2. Configure a distributed virtual switch as follows:
   a. In the VMware vSphere Web Client, expand the **RoCE** node in the left pane of the Navigator window.
   b. Right-click **RoCE-VDS**, and then click **Add and Manage Hosts**.
   c. Under **Add and Manage Hosts**, configure the following:
      - **Assign uplinks**. Select from the list of available uplinks.
      - **Manage VMkernel network adapters**. Accept the default, and then click **Next**.
      - **Migrate VM networking**. Assign the port group created in Step 1.

3. Assign a vmknic for PVRDMA to use on ESX hosts:
   a. Right-click a host, and then click **Settings**.
   b. On the Settings page, expand the **System** node, and then click **Advanced System Settings**.
   c. The Advanced System Settings page shows the key-pair value and its summary. Click **Edit**.
   d. On the Edit Advanced System Settings page, filter on **PVRDMA** to narrow all the setting to just Net.PVRDMAVmknic.
   e. Set the **Net.PVRDMAVmknic** value to **vmknic**; for example, **vmk1**.

      Figure 6-7 shows an example.

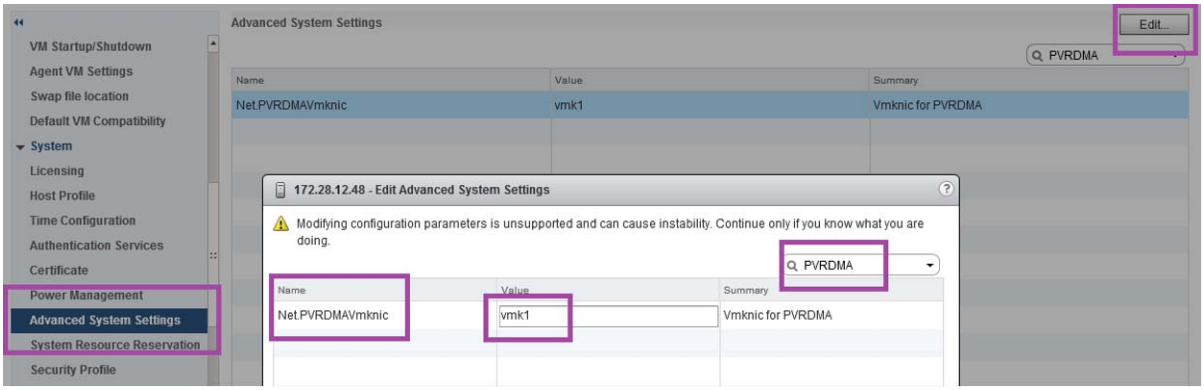


Figure 6-7. Assigning a vmknic for PVRDMA

4. Set the firewall rule for the PVRDMA:
   a. Right-click a host, and then click **Settings**.
   b. On the Settings page, expand the **System** node, and then click **Security Profile**.
   c. On the Firewall Summary page, click **Edit**.
   d. In the Edit Security Profile dialog box under **Name**, scroll down, select the **pvrdma** check box, and then select the **Set Firewall** check box.

Figure 6-8 shows an example.

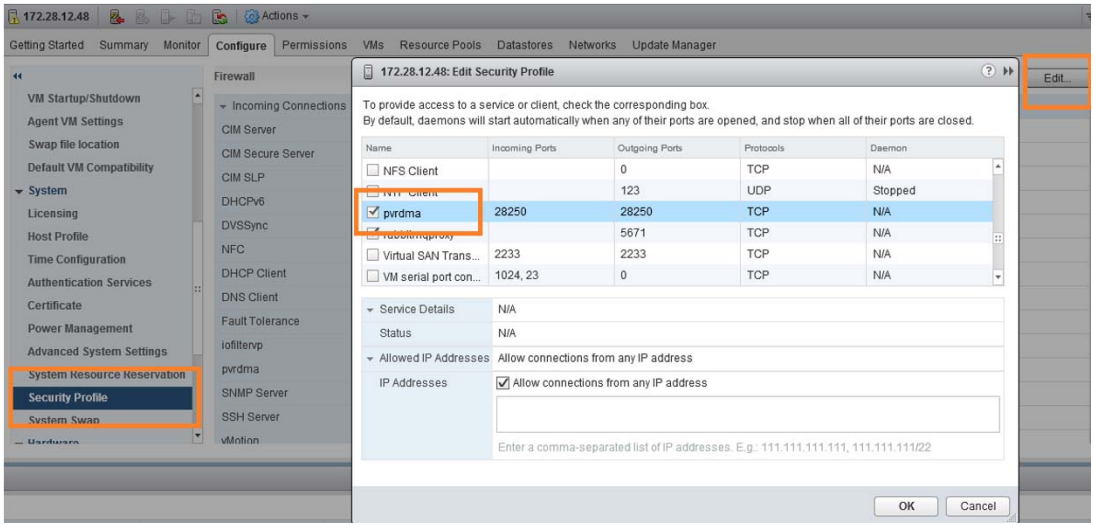


Figure 6-8. Setting the Firewall Rule

5. Set up the VM for PVRDMA as follows:
   a. Install one of the following supported guest OSs:
      - RHEL 7.2
      - Ubuntu 14.04 (kernel version 4.0)
   b. Install OFED-3.18.
   c. Compile and install the PVRDMA guest driver and library.
   d. Add a new PVRDMA network adapter to VM as follows:
      - Edit the VM settings.
      - Add a new network adapter.
      - Select the newly added DVS port group as **Network**.
      - Select **PVRDMA** as the adapter type.
   e. After the VM is booted, ensure that the PVRDMA guest driver is loaded.

7 iSER Configuration

This chapter provides procedures for configuring iSCSI Extensions for RDMA (iSER) for Linux (RHEL and SLES), including:

- Before You Begin
- Configuring iSER for RHEL
- Configuring iSER for SLES 12
- Optimizing Linux Performance

Before You Begin

As you prepare to configure iSER, consider the following:

- iSER is supported only in inbox OFED for the following operating systems:
  - RHEL 7.1 and 7.2
  - SLES 12 and 12 SP1
  - CentOS 7.2
- Out-of-box OFED does not support iSER target configuration and the ib_isert module is not available for the latest out-of-box OFED versions.
- After logging into the targets or while running I/O traffic, unloading the Linux RoCE qedr driver may crash the system.
- While running I/O, performing interface down/up tests or performing cable pull-tests can cause driver or iSER module errors that may crash the system. If this happens, reboot the system.

Configuring iSER for RHEL

To configure iSER for RHEL:

1. Install inbox OFED as described in “RoCE Configuration for RHEL” on page 63. Out-of-box OFEDs are not supported for iSER because the ib_isert module is not available in the out-of-box OFED 3.18-2 GA/3.18-3 GA versions. The inbox ib_isert module does not work with any out-of-box OFED versions.

2. Unload any existing FastLinQ drivers as described in “Removing the Linux Drivers” on page 11.
3. Install the latest FastLinQ driver and libqedr packages as described in “Installing the Linux Drivers with RDMA” on page 15.
4. Load the RDMA services.

```
systemctl start rdma
modprobe qedr
modprobe ib_iser
modprobe ib_isert
```

5. Verify that all RDMA and iSER modules loaded on the initiator and target devices by issuing the `lsmod | grep qed` and `lsmod | grep iser` commands.
6. Verify that there are separate `hca_id` instances by issuing the `ibv_devinfo` command, as shown in Step 6 on page 65.
7. Check the RDMA connection on the initiator device and the target device.
   a. On the initiator device issue the following command:

      ```
      rping -s -C 10 -v
      ```

   b. On the target device, issue the following command:

      ```
      rping -c -a 192.168.100.99 -C 10 -v
      ```

   Figure 7-1 shows an example of a successful RDMA ping.

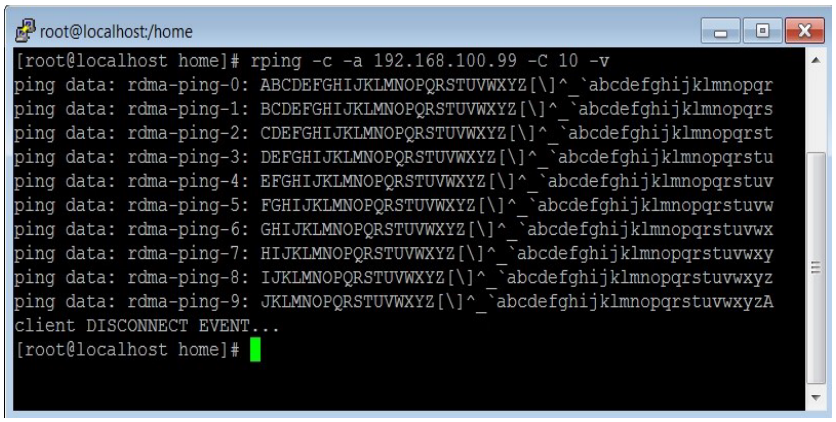


Figure 7-1. RDMA Ping Successful

8. You can use a Linux TCM-LIO target to test iSER. The setup is the same for any iSCSI target, except that you issue the command `enable_iser Boolean=true` on the applicable portals. The portal instances are identified as **`iser`** in Figure 7-2.

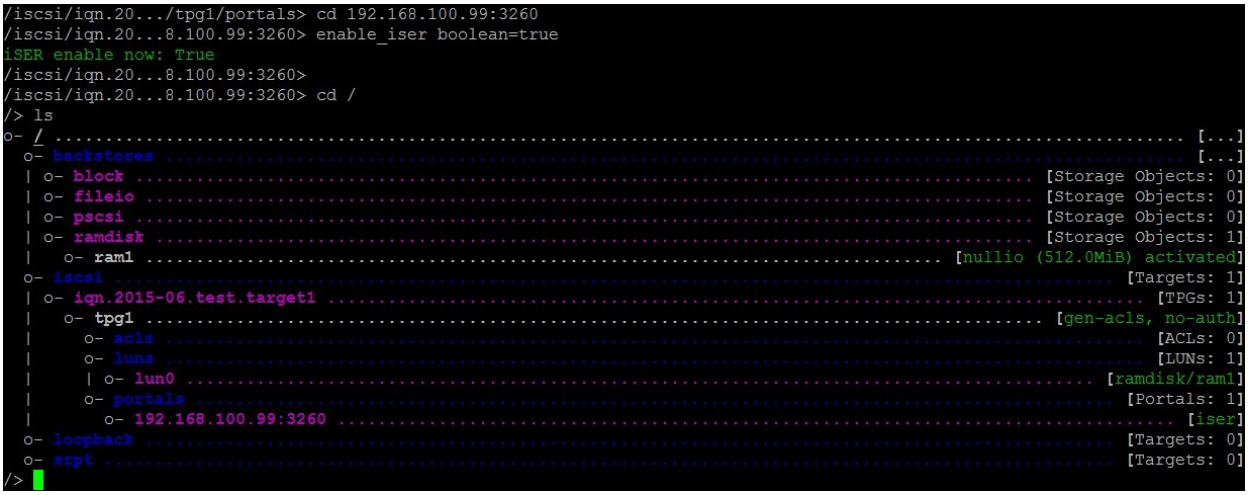


Figure 7-2. iSER Portal Instances

9. Install Linux iSCSI Initiator Utilities using the `yum install iscsi-initiator-utils` commands.

   a. To discover the iSER target, issue the `iscsiadm` command. For example:

      `iscsiadm -m discovery -t st -p 192.168.100.99:3260`

   b. To change the transport mode to iSER, issue the `iscsiadm` command. For example:

      `iscsiadm -m node -T iqn.2015-06.test.target1 -o update -n iface.transport_name -v iser`

   c. To connect to or log in to the iSER target, issue the `iscsiadm` command. For example:

      `iscsiadm -m node -l -p 192.168.100.99:3260 -T iqn.2015-06.test.target1`

   d. Confirm that the `Iface Transport` is **`iser`** in the target connection, as shown Figure 7-3. Issue the `iscsiadm` command; for example:

      `iscsiadm -m session -P2`

```
[root@localhost ~]# iscsiadm -m discovery -t st -p 192.168.100.99:3260
192.168.100.99:3260,1 iqn.2015-06.test.target1
192.168.100.99:3260,1 iqn.2015-06.test.target1
[root@localhost ~]#
[root@localhost ~]# iscsiadm -m node -T iqn.2015-06.test.target1 -o update -n iface.transport_name -v iser
[root@localhost ~]#
[root@localhost ~]#
[root@localhost ~]# iscsiadm -m node -l -p 192.168.100.99:3260 -T iqn.2015-06.test.target1
Logging in to [iface: default, target: iqn.2015-06.test.target1, portal: 192.168.100.99,3260] (multiple)
Login to [iface: default, target: iqn.2015-06.test.target1, portal: 192.168.100.99,3260] successful.
[root@localhost ~]#
[root@localhost ~]# iscsiadm -m session -P2
Target: iqn.2015-06.test.target1 (non-flash)
        Current Portal: 192.168.100.99:3260,1
        Persistent Portal: 192.168.100.99:3260,1
                **********
                Interface:
                **********
                Iface Name: default
                Iface Transport: iser
                Iface Initiatorname: iqn.1994-05.com.redhat:c672dfb8b08f
                Iface IPaddress: <empty>
                Iface HWaddress: <empty>
                Iface Netdev: <empty>
                SID: 33
                iSCSI Connection State: LOGGED IN
                iSCSI Session State: LOGGED_IN
                Internal iscsid Session State: NO CHANGE
                *********
                Timeouts:
                *********
                Recovery Timeout: 120
```

Figure 7-3. Iface Transport Confirmed

e. To check for a new iSCSI device, as shown Figure 7-4, issue the `lsscsi` command.

```
[root@localhost ~]# lsscsi
[6:0:0:0]    disk    HP       LOGICAL VOLUME   1.18  /dev/sdb
[6:0:0:1]    disk    HP       LOGICAL VOLUME   1.18  /dev/sda
[6:0:0:3]    disk    HP       LOGICAL VOLUME   1.18  /dev/sdc
[6:3:0:0]    storage HP       P440ar           1.18  -
[39:0:0:0]   disk    LIO-ORG  ram1             4.0   /dev/sdd
[root@localhost ~]#
```

Figure 7-4. Checking for New iSCSI Device

Configuring iSER for SLES 12

Because the targetcli is not inbox on SLES 12.*x*, you must complete the following procedure.

To configure iSER for SLES 12:

1. To install targetcli, copy and install the following RPMs from the ISO image (x86_64 and noarch location):

   `lio-utils-4.1-14.6.x86_64.rpm`

```
python-configobj-4.7.2-18.10.noarch.rpm
python-PrettyTable-0.7.2-8.5.noarch.rpm
python-configshell-1.5-1.44.noarch.rpm
python-pyparsing-2.0.1-4.10.noarch.rpm
python-netifaces-0.8-6.55.x86_64.rpm
python-rtslib-2.2-6.6.noarch.rpm
python-urwid-1.1.1-6.144.x86_64.rpm
targetcli-2.1-3.8.x86_64.rpm
```

2. Before starting the targetcli, load all RoCE device drivers and iSER modules as follows:

```
# modprobe qed
# modprobe qede
# modprobe qedr
# modprobe ib_iser   (Initiator)
# modprobe ib_isert  (Target)
```

3. Before configuring iSER targets, configure NIC interfaces and run L2 and RoCE traffic, as described in Step 7 on page 65.

4. Start the targetcli utility, and configure your targets on the iSER target system.

NOTE

targetcli versions are different in RHEL and SLES. Be sure to use the proper backstores to configure your targets:

- RHEL uses *ramdisk*
- SLES uses *rd_mcp*

Optimizing Linux Performance

Consider the following Linux performance configuration enhancements described in this section.

- Configuring CPUs to Maximum Performance Mode
- Configuring Kernel sysctl Settings
- Configuring IRQ Affinity Settings
- Configuring Block Device Staging

Configuring CPUs to Maximum Performance Mode

Configure the CPU scaling governor to performance by using the following script to set all CPUs to maximum performance mode:

```
for CPUFREQ in
/sys/devices/system/cpu/cpu*/cpufreq/scaling_governor; do [ -f
$CPUFREQ ] || continue; echo -n performance > $CPUFREQ; done
```

Verify that all CPU cores are set to maximum performance mode by issuing the following command:

```
cat /sys/devices/system/cpu/cpu*/cpufreq/scaling_governor
```

Configuring Kernel sysctl Settings

Set the kernel sysctl settings as follows:

```
sysctl -w net.ipv4.tcp_mem="4194304 4194304 4194304"
sysctl -w net.ipv4.tcp_wmem="4096 65536 4194304"
sysctl -w net.ipv4.tcp_rmem="4096 87380 4194304"
sysctl -w net.core.wmem_max=4194304
sysctl -w net.core.rmem_max=4194304
sysctl -w net.core.wmem_default=4194304
sysctl -w net.core.rmem_default=4194304
sysctl -w net.core.netdev_max_backlog=250000
sysctl -w net.ipv4.tcp_timestamps=0
sysctl -w net.ipv4.tcp_sack=1
sysctl -w net.ipv4.tcp_low_latency=1
sysctl -w net.ipv4.tcp_adv_win_scale=1
echo 0 > /proc/sys/vm/nr_hugepages
```

Configuring IRQ Affinity Settings

The following example sets CPU core 0, 1, 2, and 3 to IRQ XX, YY, ZZ, and XYZ respectively. Perform these steps for each IRQ assigned to a port (default is eight queues per port).

```
systemctl disable irqbalance
systemctl stop irqbalance
cat /proc/interrupts | grep qedr   Shows IRQ assigned to each port queue
echo 1 > /proc/irq/XX/smp_affinity_list
echo 2 > /proc/irq/YY/smp_affinity_list
echo 4 > /proc/irq/ZZ/smp_affinity_list
echo 8 > /proc/irq/XYZ/smp_affinity_list
```

Configuring Block Device Staging

Set the block device staging settings for each iSCSI device or target as follows:

```
echo noop > /sys/block/sdd/queue/scheduler
echo 2 > /sys/block/sdd/queue/nomerges
echo 0 > /sys/block/sdd/queue/add_random
echo 1 > /sys/block/sdd/queue/rq_affinity
```

8 SR-IOV Configuration

Single root input/output virtualization (SR-IOV) is a specification by the PCI SIG that enables a single PCI Express (PCIe) device to appear as multiple, separate physical PCIe devices. SR-IOV permits isolation of PCIe resources for performance, interoperability, and manageability.

NOTE

Some SR-IOV features may not be fully enabled in the current release.

This chapter provides instructions for:

- Configuring SR-IOV on Windows
- Configuring SR-IOV on Linux
- Configuring SR-IOV on VMware

Configuring SR-IOV on Windows

To configure SR-IOV on Windows:

1. Access the server BIOS System Setup, and then click **System BIOS Settings**.
2. On the System BIOS Settings page, click **Integrated Devices**.

3. On the Integrated Devices page (Figure 8-1):
   a. Set the **SR-IOV Global Enable** option to **Enabled**.
   b. Click **Back**.

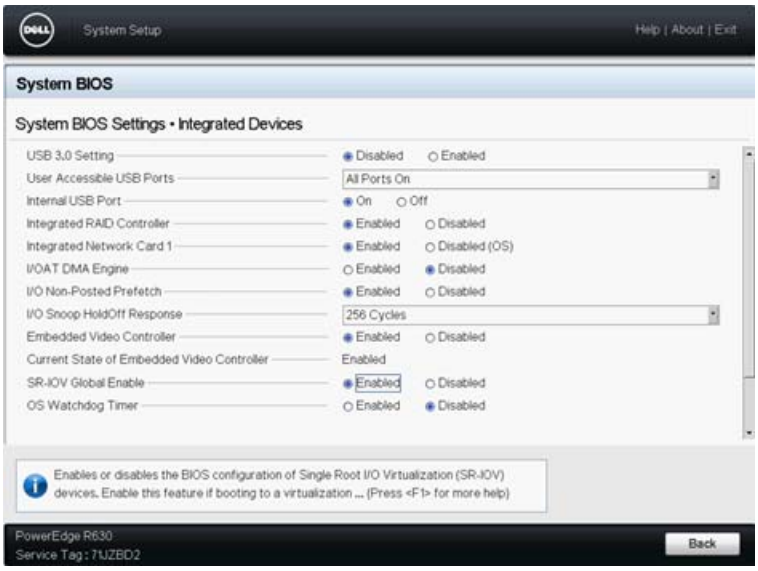


Figure 8-1. System Setup for SR-IOV: Integrated Devices

4. On the Main Configuration Page for the selected adapter, click **Device Level Configuration**.
5. On the Main Configuration Page - Device Level Configuration (Figure 8-2):
   a. Set the Virtualization Mode to **SR-IOV**.
   b. Click Back.



Figure 8-2. System Setup for SR-IOV: Device Level Configuration

6. On the Main Configuration Page, click **Finish**.
7. In the Warning - Saving Changes message box, click **Yes** to save the configuration.
8. In the Success - Saving Changes message box, click **OK**.

9. To enable SR-IOV on the miniport adapter:
   a. Access Device Manager.
   b. Open the miniport adapter properties, and then click the **Advanced** tab.
   c. On the Advanced properties page (Figure 8-3) under **Property**, select **SR-IOV**, and then set the value to **Enabled**.
   d. Click **OK**.

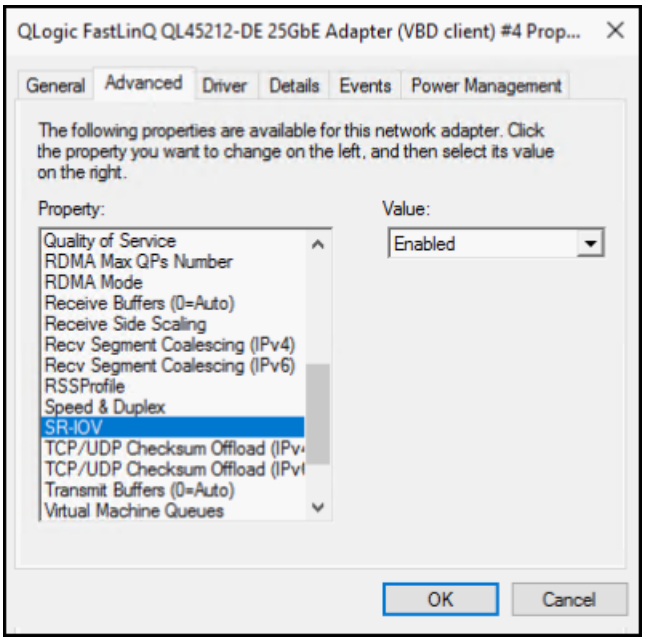


Figure 8-3. Adapter Properties, Advanced: Enabling SR-IOV

10. To create a Virtual Machine Switch with SR-IOV (Figure 8-4 on page 89):
    a. Launch the Hyper-V Manager.
    b. Select **Virtual Switch Manager**.
    c. In the **Name** box, type a name for the virtual switch.
    d. Under **Connection type**, select **External network**.

e. Select the **Enable single-root I/O virtualization (SR-IOV)** check box, and then click **Apply**.

NOTE

Be sure to enable SR-IOV when you create the vSwitch. This option is unavailable after the vSwitch is created.

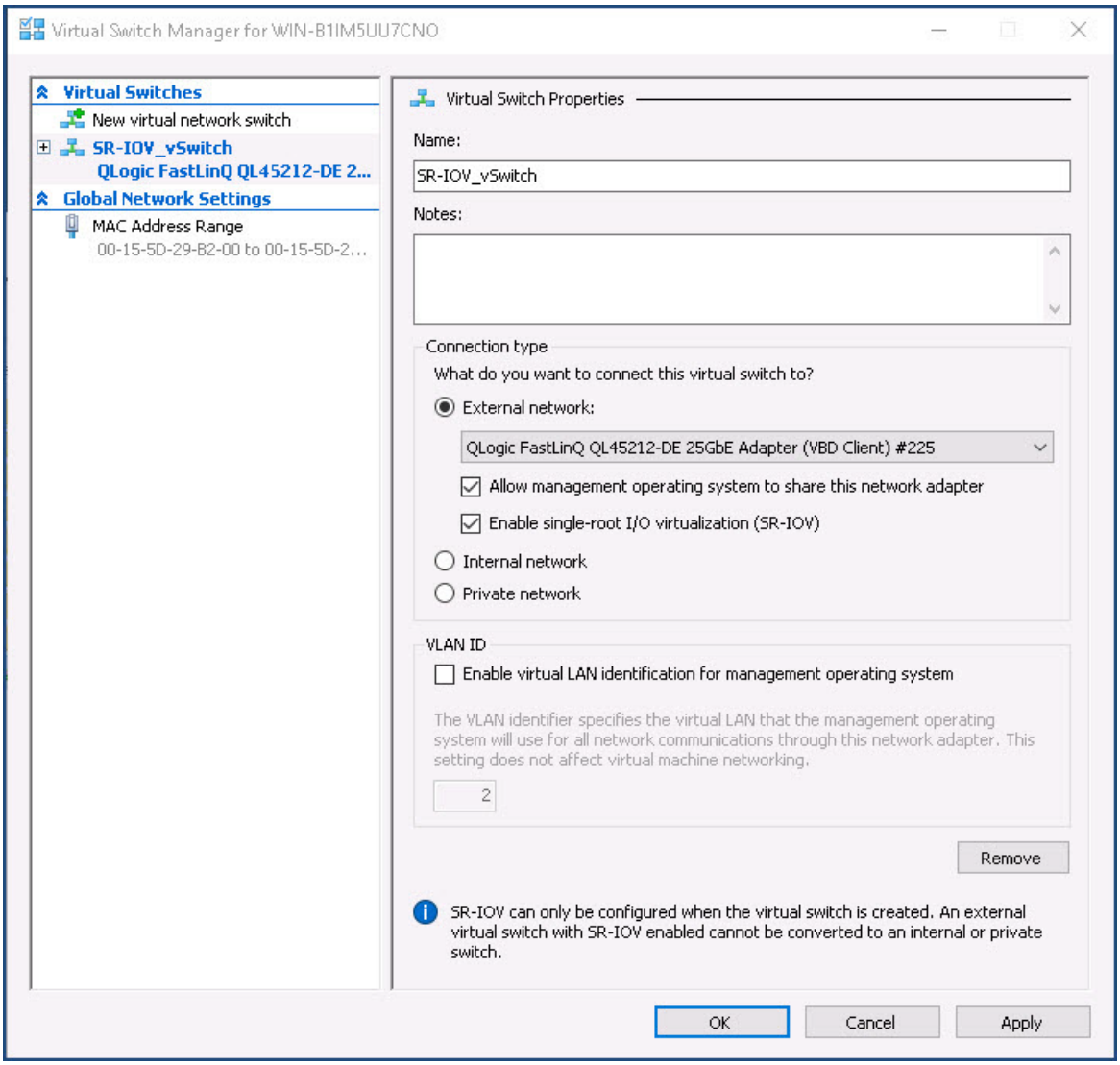


Figure 8-4. Virtual Switch Manager: Enabling SR-IOV

f. The Apply Networking Changes message box advises you that **Pending changes may disrupt network connectivity**. To save your changes and continue, click **Yes**.

11. To get the virtual machine switch capability, issue the following Windows PowerShell command:

```
PS C:\Users\Administrator> Get-VMSwitch -Name SR-IOV_vSwitch | fl
```

Output of the `Get-VMSwitch` command includes the following SR-IOV capabilities:

```
IovVirtualFunctionCount              : 96
IovVirtualFunctionsInUse             : 1
```

12. To create a virtual machine (VM) and export the virtual function (VF) in the VM:

a. Create a virtual machine.

b. Add the VMNetworkadapter to the virtual machine.

c. Assign a virtual switch to the VMNetworkadapter.

d. In the Settings for VM <VM_Name> dialog box (Figure 8-5), Hardware Acceleration page, under **Single-root I/O virtualization**, select the **Enable SR-IOV** check box, and then click **OK**.

NOTE

After the virtual adapter connection is created, the SR-IOV setting can be enabled or disabled at any time (even while traffic is running).

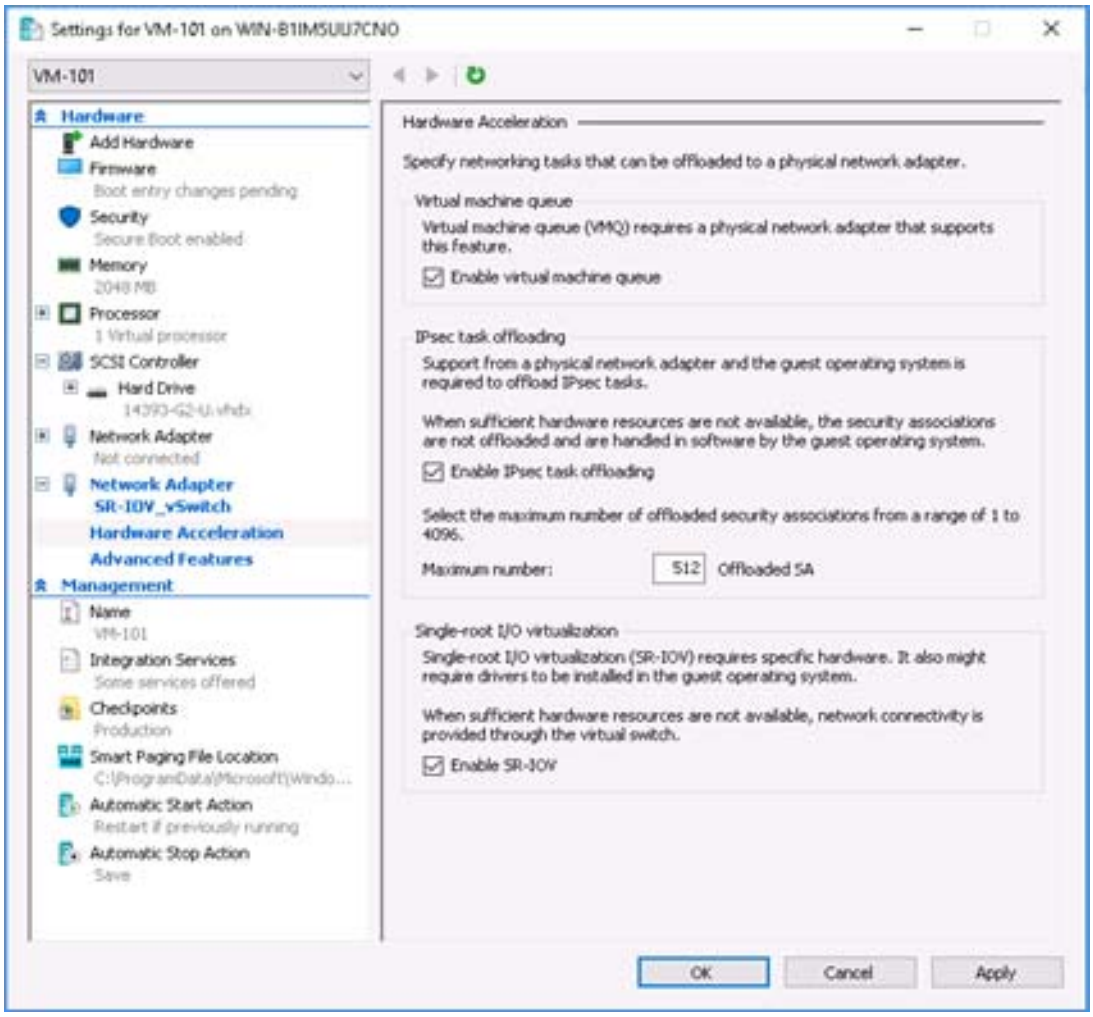


Figure 8-5. Settings for VM: Enabling SR-IOV

13. Install the QLogic drivers for the adapters detected in the VM. Use the latest drivers available from your vendor for your host OS (do not use inbox drivers).

NOTE

Be sure to use the same driver package on both the VM and the host system. For example, use the same qeVBD and qeND driver version on the Windows VM and in the Windows Hyper-V host.

After installing the drivers, the QLogic adapter is listed in the VM. Figure 8-6 shows an example.

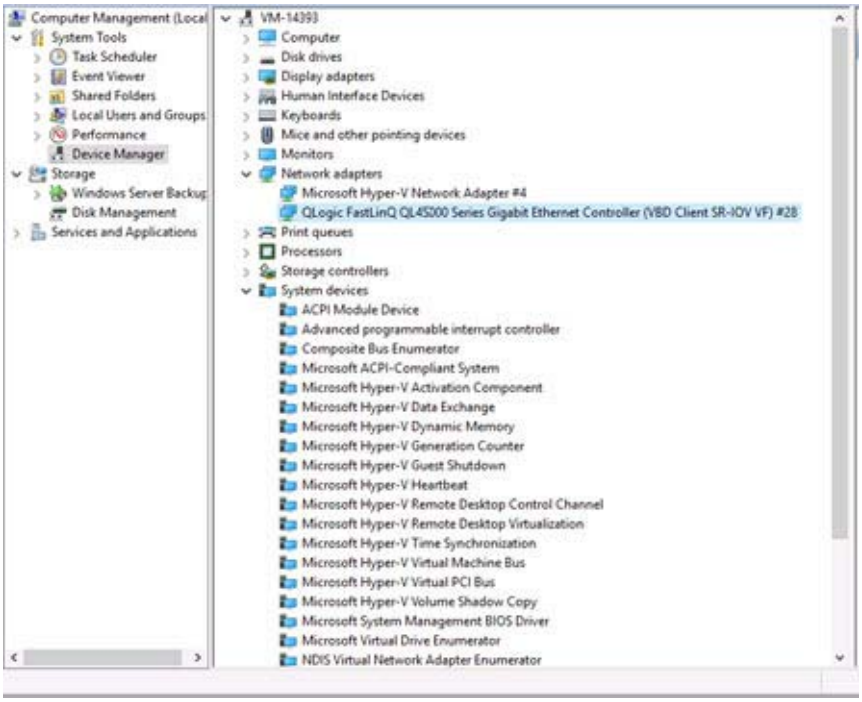


Figure 8-6. Device Manager: VM with QLogic Adapter

14. To view the SR-IOV VF details, issue the following Windows PowerShell command:

    ```
    PS C:\Users\Administrator> Get-NetadapterSriovVf
    ```

    Figure 8-7 shows example output.

```
PS C:\Users\Administrator>
PS C:\Users\Administrator> Get-NetAdapterSriovVf

Name                      FunctionID VPortID MacAddress        VmID                                 VmFriendlyName
----                      ---------- ------- ----------        ----                                 --------------
Ethernet 10               0          {2}     00-15-5D-29-B2-01 51F01C52-CDC6-4932-A95E-86D...       VM-101

PS C:\Users\Administrator> _
```

Figure 8-7. Windows PowerShell Command: Get-NetadapterSriovVf

Configuring SR-IOV on Linux

To configure SR-IOV on Linux:

1. Access the server BIOS System Setup, and then click **System BIOS Settings**.
2. On the System BIOS Settings page, click **Integrated Devices**.
3. On the System Integrated Devices page (see Figure 8-1 on page 87):
   a. Set the **SR-IOV Global Enable** option to **Enabled**.
   b. Click **Back**.
4. On the System BIOS Settings page, click **Processor Settings**.
5. On the Processor Settings (Figure 8-8) page:
   a. Set the **Virtualization Technology** option to **Enabled**.
   b. Click **Back**.

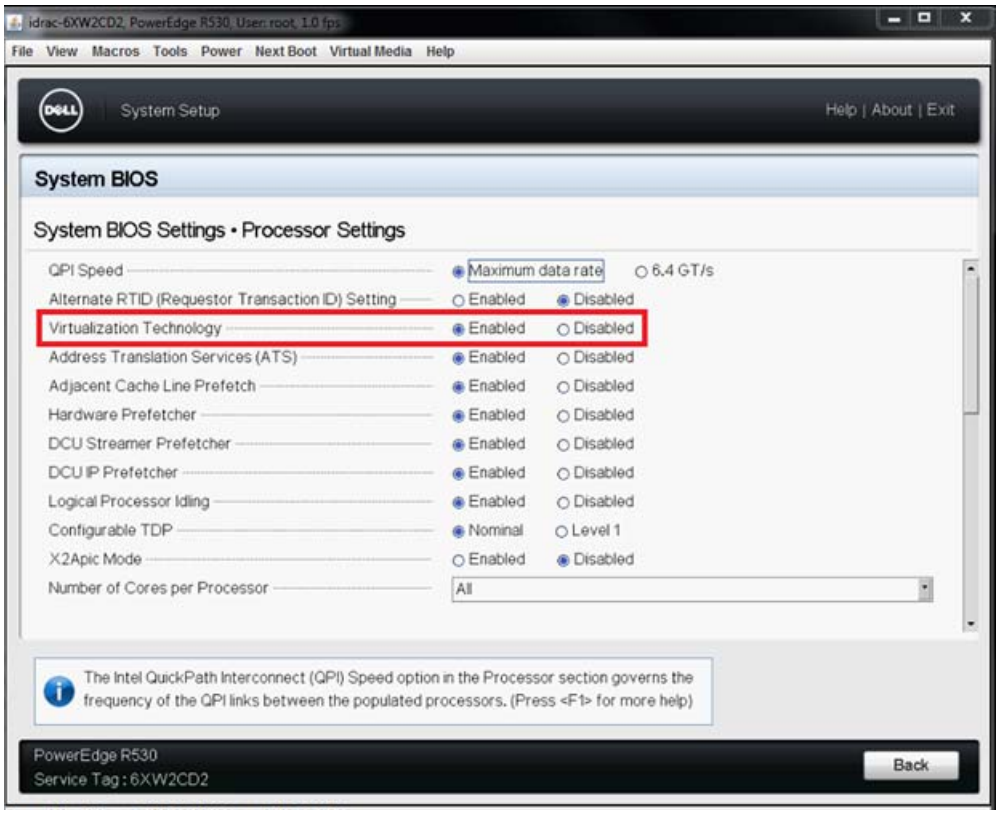


Figure 8-8. System Setup: Processor Settings for SR-IOV

6. On the System Setup page, select Device Settings.

7. On the Device Settings page, select **Port 1** for the QLogic adapter.
8. On the Device Level Configuration page (Figure 8-9):
   a. Set the **Virtualization Mode** to **SR-IOV**.
   b. Click **Back**.



Figure 8-9. System Setup for SR-IOV: Integrated Devices

9. On the Main Configuration Page, click **Finish**, save your settings, and then reboot the system.
10. To enable and verify virtualization:
    a. Open the `grub.conf` file and configure the `iommu` parameter as shown in Figure 8-10.
       - For Intel-based systems, add `intel_iommu=on`.
       - For AMD-based systems, add `amd_iommu=on`.

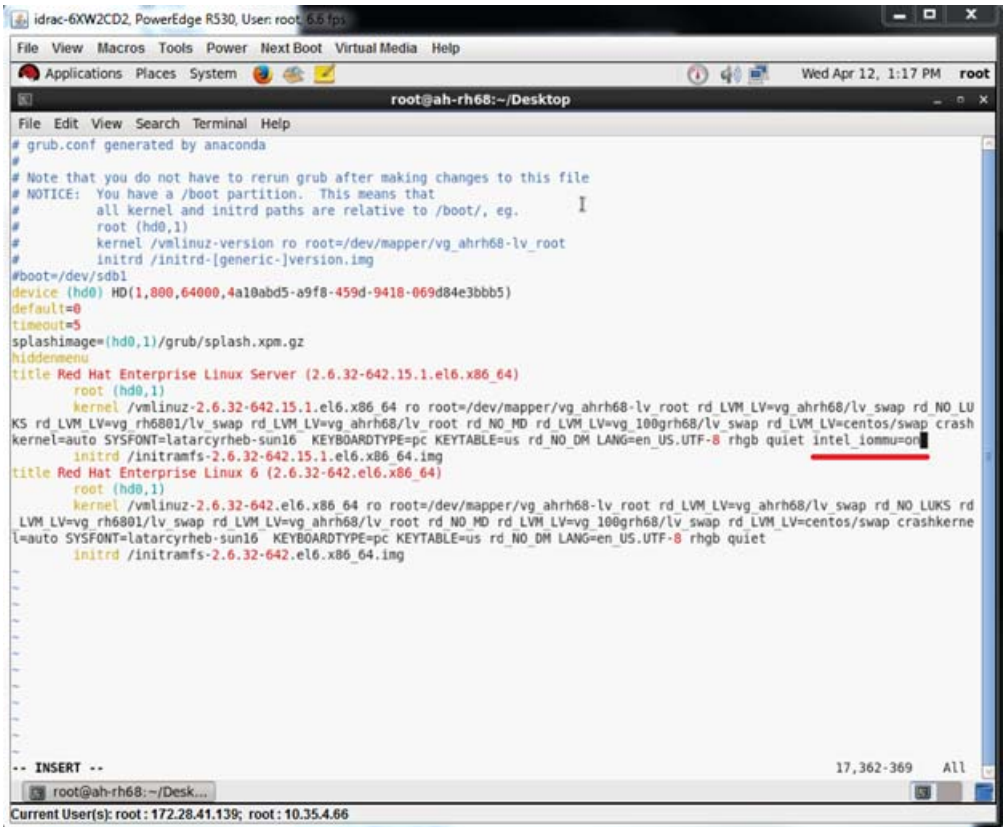


Figure 8-10. Editing the grub.conf File for SR-IOV

b. Save the `grub.conf` file and then reboot the system.

c. To verify that the changes are in effect, issue the following command:

`dmesg | grep -I iommu`

A successful input–output memory management unit (IOMMU) command output should show, for example:

`Intel-IOMMU: enabled`

d. To view VF details (number of VFs and total VFs), issue the `find /sys/|grep -I sriov` command.

11. For a specific port, enable a quantity of VFs.

a. Issue the following command to enable, for example, 8 VF on PCI instance 04:00.0 (bus 4, device 0, function 0):

`[root@ah-rh68 ~]# echo 8 >` **`/sys/devices/pci0000:00/0000:00:02.0/0000:04:00.0/sriov_numvfs`**

b. Review the command output (Figure 8-11) to confirm that actual VFs were created on bus 4, device 2 (from the 0000:00:02.0 parameter), functions 0 through 7. Note that the actual device ID is different on the PFs (8070 in this example) versus the VFs (9090 in this example).

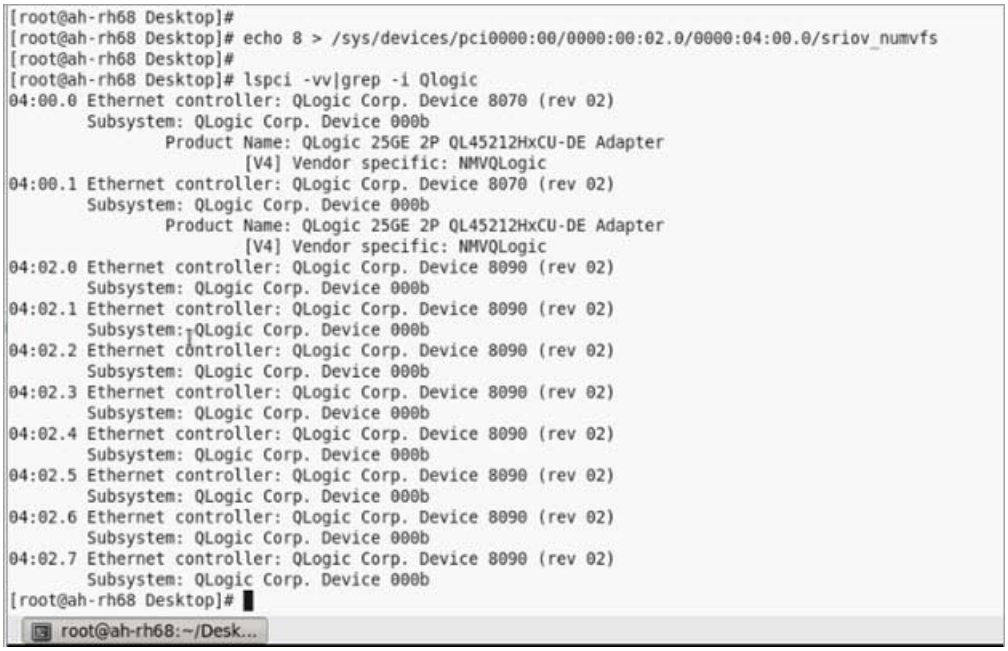


Figure 8-11. Command Output for sriov_numvfs

12. To view a list of all PF and VF interfaces, issue the following command:

    # **ip link show/ifconfig -a**

    Figure 8-12 shows example output.

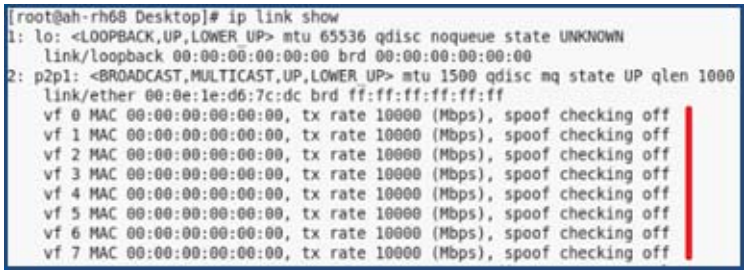


Figure 8-12. Command Output for ip link show Command

13. Assign and verify MAC addresses:

    a. To assign a MAC address to the VF, issue the following command:

       ip link set <pf device> vf <vf index> mac <mac address>

b. Ensure that the VF interface is up and running with the assigned MAC address.

14. Power off the VM and attach the VF. (Some OSs support hot-plugging of VFs to the VM.)

    a. In the Virtual Machine dialog box (Figure 8-13), click **Add Hardware**.

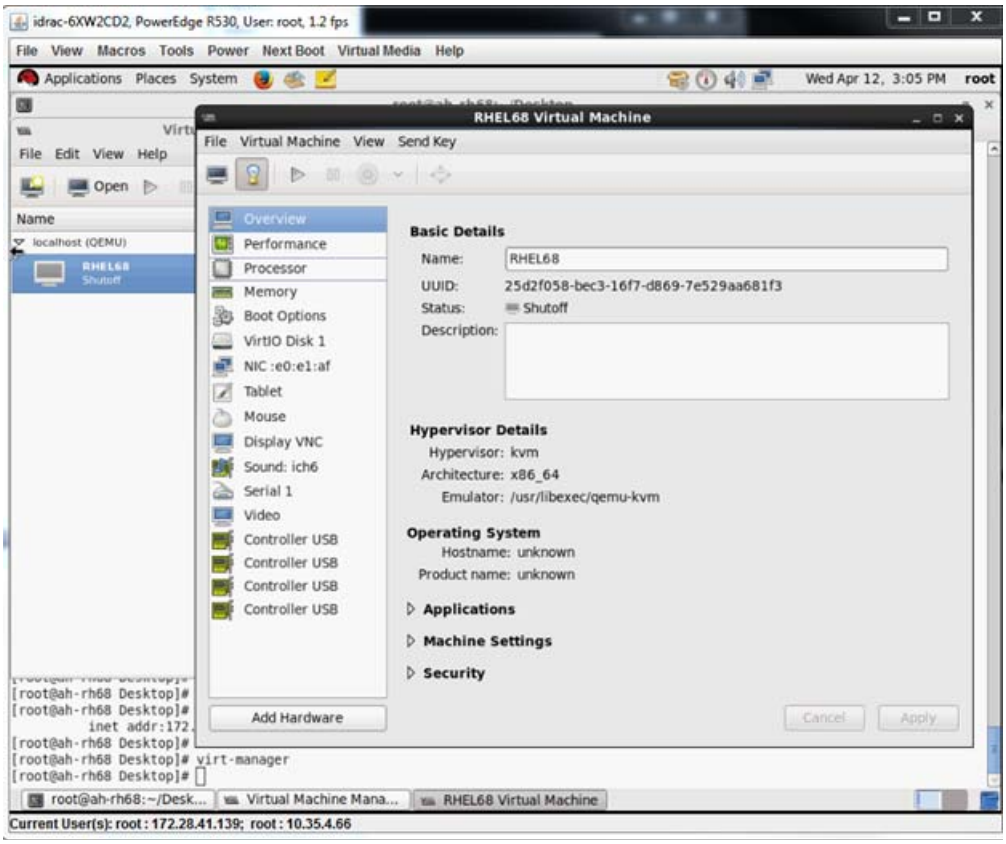


Figure 8-13. RHEL68 Virtual Machine

    b. In the left pane of the Add New Virtual Hardware dialog box (Figure 8-14), click **PCI Host Device**.

    c. In the right pane, select a host device.

    d. Click **Finish**.

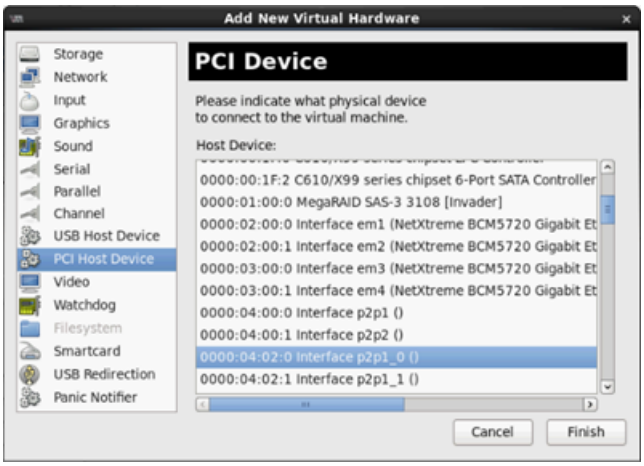


Figure 8-14. Add New Virtual Hardware

15. Power on the VM and then issue the following command:

    `check lspci -vv|grep -I ether`

16. Install the drivers for the adapters detected in the VM. Use the latest drivers available from your vendor for your host OS (do not use inbox drivers). The same driver version must be installed on the host and the VM.
17. As needed, add more VFs in the VM.

Configuring SR-IOV on VMware

To configure SR-IOV on VMware:

1. Access the server BIOS System Setup, and then click **System BIOS Settings**.
2. On the System BIOS Settings page, click **Integrated Devices**.
3. On the Integrated Devices page (see Figure 8-1 on page 87):
   a. Set the **SR-IOV Global Enable** option to **Enabled**.
   b. Click **Back**.
4. On the System Setup window, click **Device Settings**.
5. On the Device Settings page, select a port for the 25G QL45212 Adapter.
6. On the Device Level Configuration (see Figure 8-2 on page 87):
   a. Set the **Virtualization Mode** to **SR-IOV**.
   b. Click **Back**.
7. On the Main Configuration Page, click **Finish**.

8. Save the configuration settings and reboot the system.
9. To enable the needed quantity of VFs per port (in this example, 16 on each port of a dual-port adapter), issue the following command:

   "esxcfg-module -s "max_vfs=16,16" qedentv"

   NOTE

   Each Ethernet function of the QL45212 Adapter must have it's own entry.

10. Reboot the host.
11. To verify that the changes are complete at the module level, issue the following command:

    "esxcfg-module -g qedentv"

    ```
    [root@localhost:~] esxcfg-module -g qedentv
    qedentv enabled = 1 options = 'max_vfs=16,16'
    ```

12. To verify if actual VFs were created, issue the `lspci` command as follows:

```
[root@localhost:~] lspci | grep -i QLogic | grep -i 'ethernet\|network' | more
0000:05:00.0 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx 10/25
GbE Ethernet Adapter [vmnic6]
0000:05:00.1 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx 10/25
GbE Ethernet Adapter [vmnic7]
0000:05:02.0 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.0_VF_0]
0000:05:02.1 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.0_VF_1]
0000:05:02.2 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.0_VF_2]
0000:05:02.3 Network controller: QLogic Corp. QLogic FastLinQ QL41xQL45xxxxx
Series 10/25 GbE Controller (SR-IOV VF) [PF_0.5.0_VF_3]
.
.
.
0000:05:03.7 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.0_VF_15]
0000:05:0e.0 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.1_VF_0]
0000:05:0e.1 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.1_VF_1]
0000:05:0e.2 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.1_VF_2]
```

```
0000:05:0e.3 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.1_VF_3]
.
.
.
0000:05:0f.6 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.1_VF_14]
0000:05:0f.7 Network controller: QLogic Corp. QLogic FastLinQ QL45xxx Series
10/25 GbE Controller (SR-IOV VF) [PF_0.5.1_VF_15]
```

13. To validate the VFs per port, issue the `esxcli` command as follows:

```
[root@localhost:~] esxcli network sriovnic vf list -n vmnic6
VF ID  Active  PCI Address  Owner World ID
-----  ------  -----------  --------------
    0    true  005:02.0     60591
    1    true  005:02.1     60591
    2   false  005:02.2      -
    3   false  005:02.3      -
    4   false  005:02.4      -
    5   false  005:02.5      -
    6   false  005:02.6      -
    7   false  005:02.7      -
    8   false  005:03.0      -
    9   false  005:03.1      -
   10   false  005:03.2      -
   11   false  005:03.3      -
   12   false  005:03.4      -
   13   false  005:03.5      -
   14   false  005:03.6      -
   15   false  005:03.7      -
```

14. Attach VFs to the VM as follows:
    a. Power off the VM and attach the VF. (Some OSs support hot-plugging of VFs to the VM.)
    b. Add a host to a VMware vCenter Server Virtual Appliance (vCSA).
    c. Click **Edit Settings** of the VM.
15. Complete the Edit Settings dialog box (Figure 8-15) as follows:
    a. In the **New Device** box, select **Network**, and then click **Add**.
    b. For **Adapter Type**, select **SR-IOV Passthrough**.

c. For **Physical Function**, select the QLogic VF.

d. To save your configuration changes and close this dialog box, click **OK**.

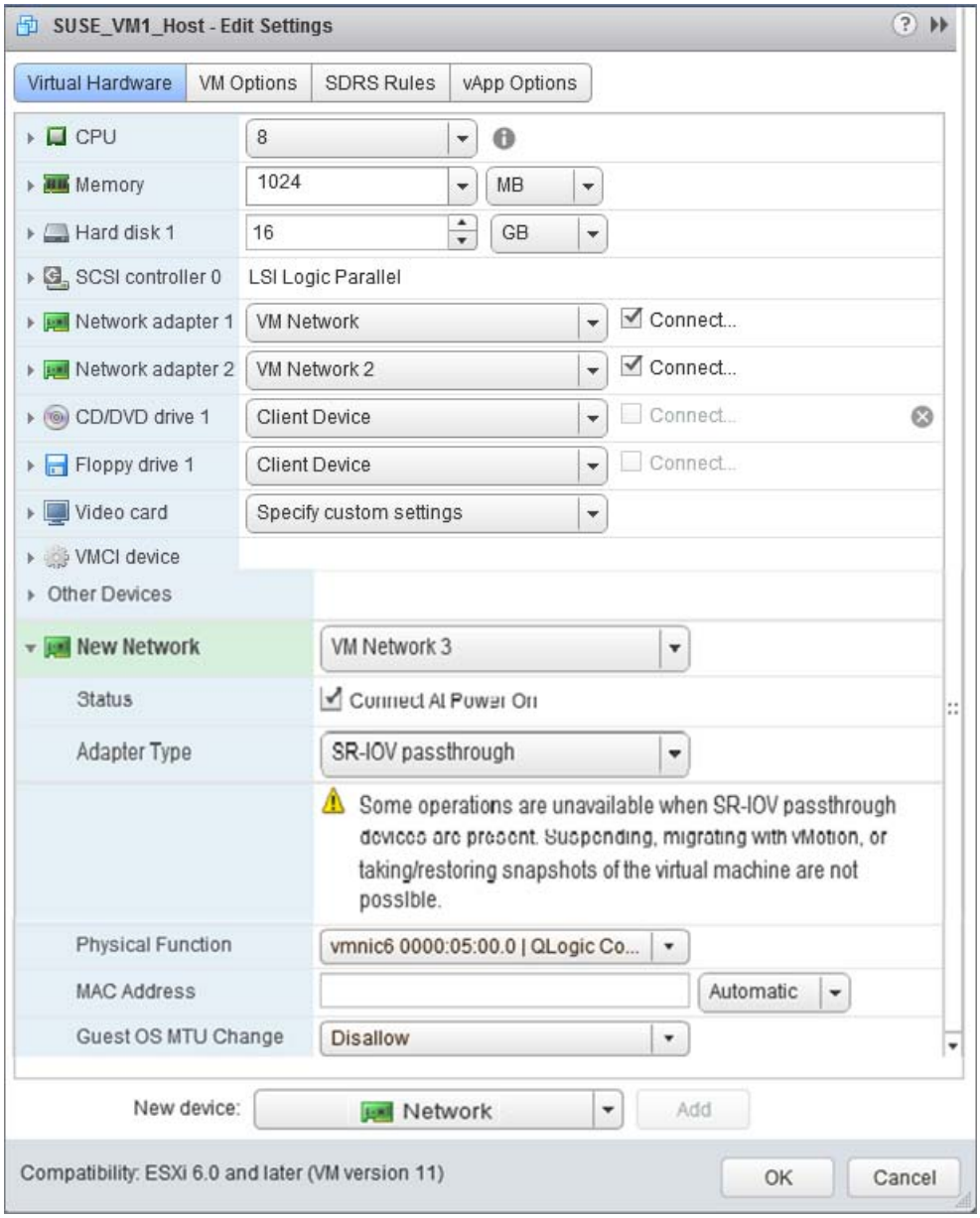


Figure 8-15. VMware Host Edit Settings

16. Power on the VM, and then issue the `ifconfig -a` command to verify that the added network interface is listed.

17. Install the QLogic drivers for the adapters detected in the VM. Use the latest drivers available from your vendor for your host OS (do not use inbox drivers). The same driver version must be installed on the host and the VM.
18. As needed, add more VFs in the VM.

9 Windows Server 2016

This chapter provides the following information for Windows Server 2016:

- Configuring RoCE Interfaces with Hyper-V
- RoCE over Switch Embedded Teaming
- Configuring QoS for RoCE
- Configuring VMMQ
- Configuring VXLAN
- Configuring Storage Spaces Direct
- Deploying and Managing a Nano Server

Configuring RoCE Interfaces with Hyper-V

In Windows Server 2016, Hyper-V with Network Direct Kernel Provider Interface (NDKPI) Mode-2, host virtual network adapters (host virtual NICs) support RDMA.

NOTE

DCBX is required for RoCE over Hyper-V. To configure DCBX, either:

- Configure through the HII (see “Preparing the Adapter” on page 58).
- Configure using QoS (see “Configuring QoS for RoCE” on page 110).

RoCE configuration procedures in this section include:

- Creating a Hyper-V Virtual Switch with an RDMA Virtual NIC
- Adding a VLAN ID to Host Virtual NIC
- Verifying If RoCE is Enabled
- Adding Host Virtual NICs (Virtual Ports)
- Mapping the SMB Drive and Running RoCE Traffic

Creating a Hyper-V Virtual Switch with an RDMA Virtual NIC

Follow the procedures in this section to create a Hyper-V virtual switch and then enable RDMA in the host VNIC.

To create a Hyper-V virtual switch with an RDMA virtual NIC:

1. Launch Hyper-V Manager.
2. Click **Virtual Switch Manager** (see Figure 9-1).

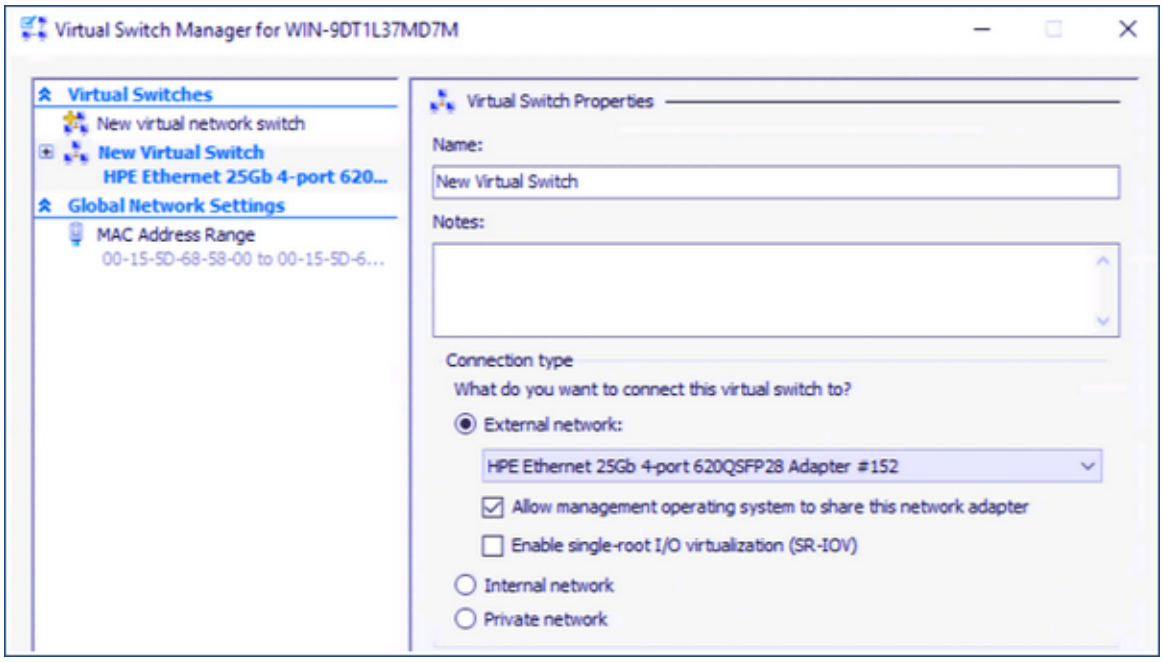


Figure 9-1. Enabling RDMA in Host Virtual NIC

3. Create a virtual switch.
4. Select the **Allow management operating system to share this network adapter** check box.

In Windows Server 2016, a new parameter—Network Direct (RDMA)—is added in the Host virtual NIC.

To enable RDMA in a host virtual NIC:

1. Open the Hyper-V Virtual Ethernet Adapter Properties window.
2. Click the **Advanced** tab.
3. On the Advanced page (Figure 9-2):
   a. Under **Property**, select **Network Direct (RDMA)**.
   b. Under **Value**, select **Enabled**.

c. Click **OK**.

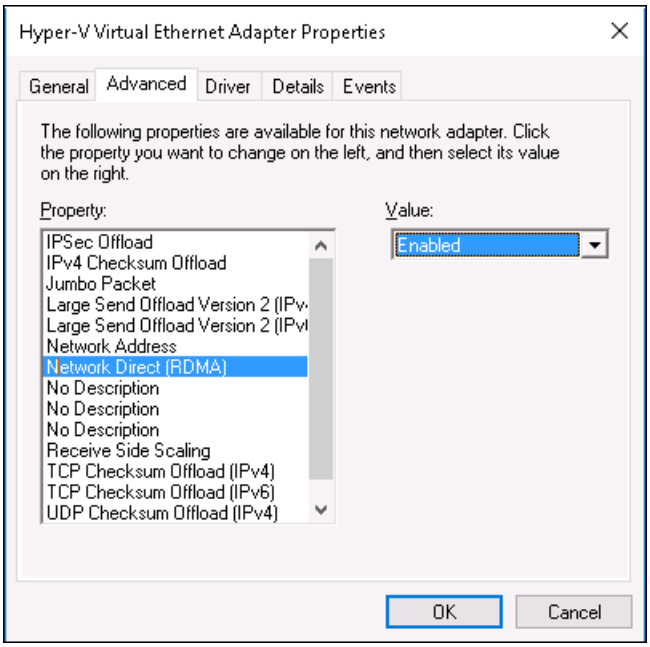


Figure 9-2. Hyper-V Virtual Ethernet Adapter Properties

4. To enable RDMA, issue the following Windows PowerShell command:

```
PS C:\Users\Administrator> Enable-NetAdapterRdma "vEthernet
(New Virtual Switch)"
PS C:\Users\Administrator>
```

Adding a VLAN ID to Host Virtual NIC

To add VLAN ID to a host virtual NIC:

1. To find the host virtual NIC name, issue the following Windows PowerShell command:

```
PS C:\Users\Administrator> Get-VMNetworkAdapter -ManagementOS
```

Figure 9-3 shows the command output.

```
PS C:\Users\Administrator> Get-VMNetworkAdapter -ManagementOS

Name                IsManagementOs VMName SwitchName         MacAddress   Status IPAddresses
----                -------------- ------ ----------         ----------   ------ -----------
New Virtual Switch  True                  New Virtual Switch 000E1EC41F0B {Ok}
```

Figure 9-3. Windows PowerShell Command: Get-VMNetworkAdapter

2. To set the VLAN ID to the host virtual NIC, issue the following Windows PowerShell command:

```
PS C:\Users\Administrator> Set-VMNetworkAdaptervlan
-VMNetworkAdapterName "New Virtual Switch" -VlanId 5 -Access
-Management05
```

NOTE

Note the following about adding a VLAN ID to a host virtual NIC:

- A VLAN ID must be assigned to a host virtual NIC. The same VLAN ID must be assigned to all the interfaces, and on the switch.
- Make sure that the VLAN ID is not assigned to the physical Interface when using a host virtual NIC for RoCE.
- If you are creating more than one host virtual NIC, you can assign a different VLAN to each host virtual NIC.

Verifying If RoCE is Enabled

To verify if the RoCE is enabled:

- Issue the following Windows PowerShell command:

  `Get-NetAdapterRdma`

  Command output lists the RDMA supported adapters as shown in Figure 9-4.

```
PS C:\Users\Administrator> Get-NetAdapterRdma

Name                      InterfaceDescription                Enabled
----                      --------------------                -------
vEthernet (New Virtual... Hyper-V Virtual Ethernet Adapter    True
```

Figure 9-4. Windows PowerShell Command: Get-NetAdapterRdma

Adding Host Virtual NICs (Virtual Ports)

To add host virtual NICs:

1. To add a host virtual NIC, issue the following command:

   `Add-VMNetworkAdapter -SwitchName "New Virtual Switch" -Name SMB - ManagementOS`

2. Enable RDMA on host virtual NICs as shown in “To enable RDMA in a host virtual NIC:” on page 104.

3. To assign a VLAN ID to the virtual port, issue the following command:

   `Set-VMNetworkAdapterVlan -VMNetworkAdapterName SMB -VlanId 5 -Access -ManagementOS`

Mapping the SMB Drive and Running RoCE Traffic

To map the SMB drive and run the RoCE traffic:

1. Launch the Performance Monitor (Perfmon).
2. Complete the Add Counters dialog box (Figure 9-5) as follows:
   a. Under **Available counters**, select **RDMA Activity.**
   b. Under **Instances of selected object**, select the adapter.
   c. Click **Add**.

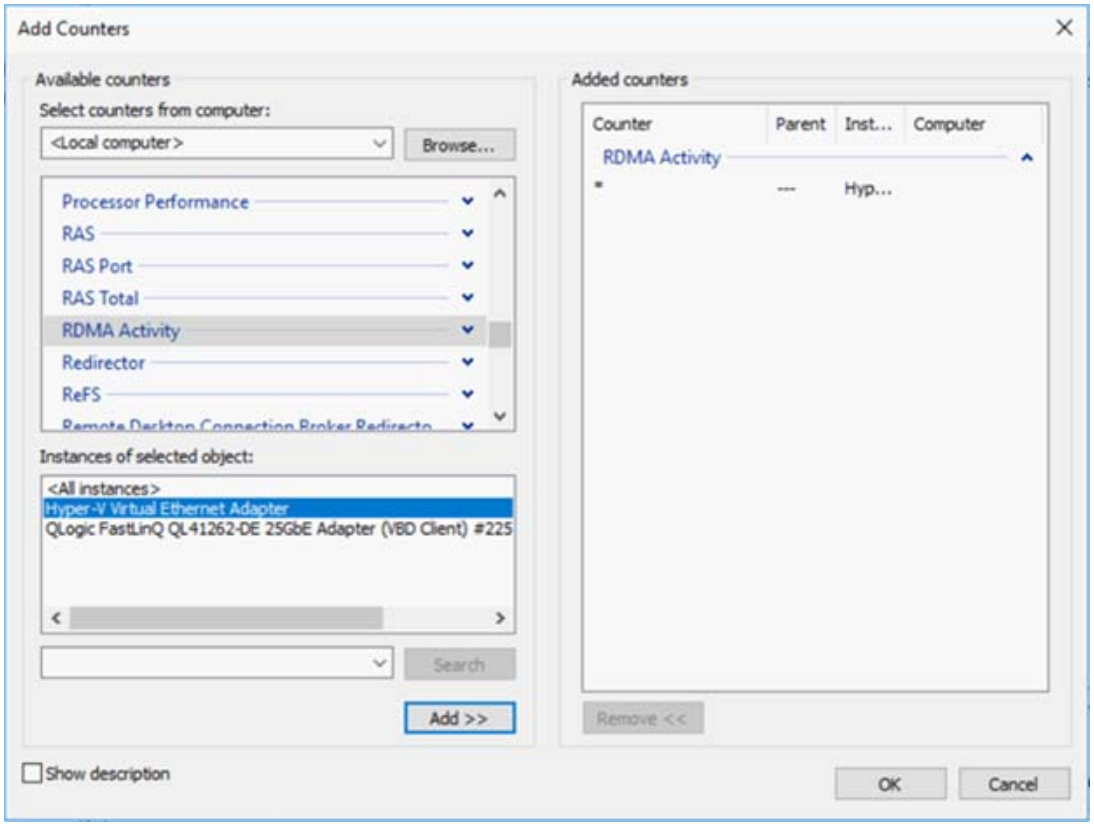


Figure 9-5. Add Counters Dialog Box

If the RoCE traffic is running, counters appear as shown in Figure 9-6.

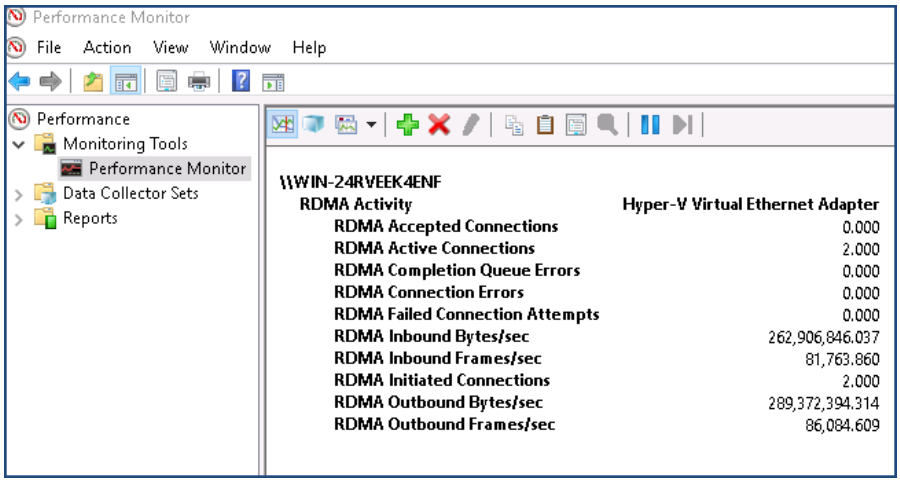


Figure 9-6. Performance Monitor Shows RoCE Traffic

RoCE over Switch Embedded Teaming

Switch Embedded Teaming (SET) is Microsoft's alternative NIC teaming solution available to use in environments that include Hyper-V and the Software Defined Networking (SDN) stack in Windows Server 2016 Technical Preview. SET integrates limited NIC Teaming functionality into the Hyper-V Virtual Switch.

Use SET to group between one and eight physical Ethernet network adapters into one or more software-based virtual network adapters. These adapters provide fast performance and fault tolerance if a network adapter failure occurs. To be placed on a team, SET member network adapters must all be installed in the same physical Hyper-V host.

RoCE over SET procedures included in this section:

- Creating a Hyper-V Virtual Switch with SET and RDMA Virtual NICs
- Enabling RDMA on SET
- Assigning a VLAN ID on SET
- Running RDMA Traffic on SET

Creating a Hyper-V Virtual Switch with SET and RDMA Virtual NICs

To create a Hyper-V virtual switch with SET and RDMA virtual NICs:

- To create a SET, issue the following Windows PowerShell command:

  ```
  PS C:\Users\Administrator> New-VMSwitch -Name SET
  -NetAdapterName "Ethernet 2","Ethernet 3"
  -EnableEmbeddedTeaming $true
  ```

  Figure 9-7 shows command output.

```
PS C:\Users\Administrator> New-VMSwitch -Name SET -NetAdapterName "Ethernet 2","Ethernet 3" -EnableEmbeddedTeaming $true

Name SwitchType NetAdapterInterfaceDescription
---- ---------- ------------------------------
SET  External   Teamed-Interface
```

Figure 9-7. Windows PowerShell Command: New-VMSwitch

Enabling RDMA on SET

To enable RDMA on SET:

1. To view the SET on the adapter, issue the following Windows PowerShell command:

   ```
   PS C:\Users\Administrator> Get-NetAdapter "vEthernet (SET)"
   ```

   Figure 9-8 shows command output.

```
PS C:\Users\Administrator> Get-NetAdapter "vEthernet (SET)"

Name                      InterfaceDescription                ifIndex Status       MacAddress             LinkSpeed
----                      --------------------                ------- ------       ----------             ---------
vEthernet (SET)           Hyper-V Virtual Ethernet Adapter         46 Up           00-0E-1E-C4-04-F8       50 Gbps
```

Figure 9-8. Windows PowerShell Command: Get-NetAdapter

2. To enable RDMA on SET, issue the following Windows PowerShell command:

   ```
   PS C:\Users\Administrator> Enable-NetAdapterRdma "vEthernet
   (SET)"
   ```

Assigning a VLAN ID on SET

To assign a VLAN ID on SET:

- To assign a VLAN ID on SET, issue the following Windows PowerShell command:

  ```
  PS C:\Users\Administrator> Set-VMNetworkAdapterVlan
  -VMNetworkAdapterName "SET" -VlanId 5 -Access -ManagementOS
  ```

NOTE

Note the following when adding a VLAN ID to a host virtual NIC:

- Make sure that the VLAN ID is not assigned to the physical Interface when using host virtual NIC for RoCE.
- If you are creating more than one host virtual NIC, a different VLAN can be assigned to each host virtual NIC.

Running RDMA Traffic on SET

For information about running RDMA traffic on SET, go to:

https://technet.microsoft.com/en-us/library/mt403349.aspx

Configuring QoS for RoCE

The two methods of configuring quality of service (QoS) include:

- Configuring QoS by Disabling DCBX on the Adapter
- Configuring QoS by Enabling DCBX on the Adapter

Configuring QoS by Disabling DCBX on the Adapter

All configuration must be completed on all of the systems in use before configuring quality of service by disabling DCBX on the adapter. The priority-based flow control (PFC), enhanced transition services (ETS), and traffic classes configuration must be the same on the switch and server.

To configure QoS by disabling DCBX:

1. Disable DCBX on the adapter.
2. Using HII, set the **RoCE Priority** to **0**.
3. To install the DCB role in the host, issue the following Windows PowerShell command:

   ```
   PS C:\Users\Administrators> Install-WindowsFeature
   Data-Center-Bridging
   ```

4. To set the **DCBX Willing** mode to **False**, issue the following Windows PowerShell command:

   ```
   PS C:\Users\Administrators> set-NetQosDcbxSetting -Willing 0
   ```

5. Enable QoS in the miniport as follows:
   a. Open the miniport window, and then click the **Advanced** tab.
   b. On the adapter's Advanced Properties page (Figure 9-9) under **Property**, select **Quality of Service**, and then set the value to **Enabled**.
   c. Click **OK**.

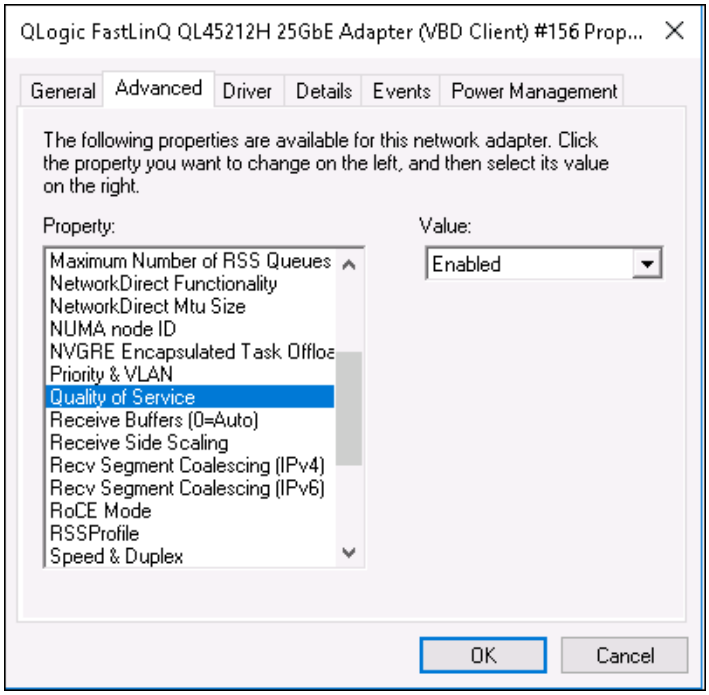


Figure 9-9. Advanced Properties: Enable QoS

6. Assign the VLAN ID to the interface as follows:
   a. Open the miniport window, and then click the **Advanced** tab.
   b. On the adapter's Advanced Properties page (Figure 9-10) under **Property**, select **VLAN ID**, and then set the value.
   c. Click **OK**.

NOTE

The preceding step is required for priority flow control (PFC).

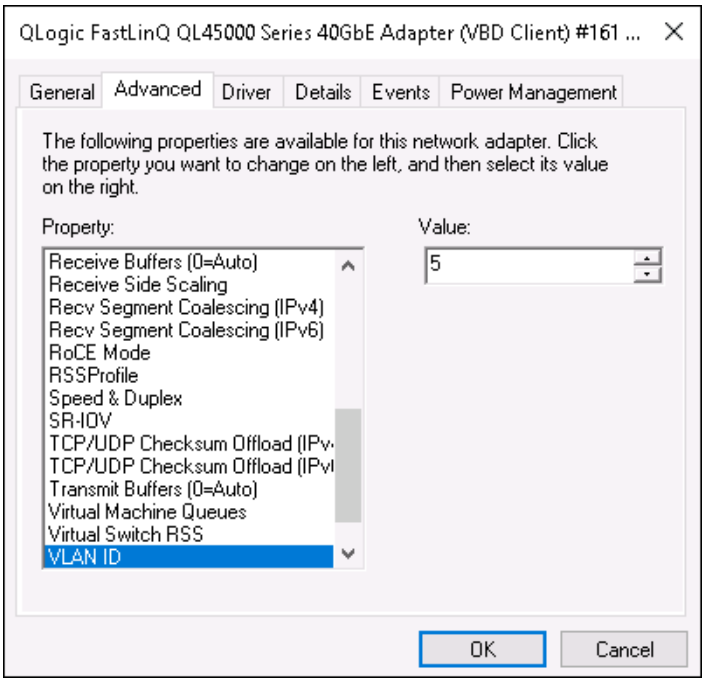


Figure 9-10. Advanced Properties: Setting VLAN ID

7. To enable priority flow control for RoCE on a specific priority, issue the following command:

```
PS C:\Users\Administrators> Enable-NetQoSFlowControl
-Priority 4
```

NOTE

If configuring RoCE over Hyper-V, do not assign a VLAN ID to the physical interface.

8. To disable priority flow control on any other priority, issue the following commands:

```
PS C:\Users\Administrator> Disable-NetQosFlowControl 0,1,2,3,5,6,7
PS C:\Users\Administrator> Get-NetQosFlowControl
Priority   Enabled    PolicySet        IfIndex IfAlias
--------   -------    ---------        ------- -------
0          False      Global
1          False      Global
2          False      Global
3          False      Global
4          True       Global
5          False      Global
6          False      Global
7          False      Global
```

9. To configure QoS and assign relevant priority to each type of traffic, issue the following commands (where Priority 4 is tagged for RoCE and Priority 0 is tagged for TCP):

```
PS C:\Users\Administrators> New-NetQosPolicy "SMB"
-NetDirectPortMatchCondition 445 -PriorityValue8021Action 4 -PolicyStore
ActiveStore

PS C:\Users\Administrators> New-NetQosPolicy "TCP" -IPProtocolMatchCondition
TCP -PriorityValue8021Action 0 -Policystore ActiveStore

PS C:\Users\Administrator> Get-NetQosPolicy -PolicyStore activestore

Name           : tcp
Owner          : PowerShell / WMI
NetworkProfile : All
Precedence     : 127
JobObject      :
IPProtocol     : TCP
PriorityValue  : 0

Name           : smb
Owner          : PowerShell / WMI
NetworkProfile : All
Precedence     : 127
JobObject      :
NetDirectPort  : 445
PriorityValue  : 4
```

10. To configure ETS for all traffic classes defined in the previous step, issue the following commands:

```
PS C:\Users\Administrators> New-NetQosTrafficClass -name "RDMA class"
-priority 4 -bandwidthPercentage 50 -Algorithm ETS

PS C:\Users\Administrators> New-NetQosTrafficClass -name "TCP class" -priority
0 -bandwidthPercentage 30 -Algorithm ETS

PS C:\Users\Administrator> Get-NetQosTrafficClass

Name        Algorithm Bandwidth(%) Priority   PolicySet       IfIndex IfAlias
----        --------- ------------ --------   ---------       ------- -------
[Default]    ETS       20           2-3,5-7    Global
RDMA class   ETS       50           4          Global
TCP class    ETS       30           0          Global
```

11. To see the network adapter QoS from the preceding configuration, issue the following Windows PowerShell command:

```
PS C:\Users\Administrator> Get-NetAdapterQos

Name                        : SLOT 4 Port 1
Enabled                     : True
Capabilities                :                          Hardware     Current
                                                       --------     -------
                              MacSecBypass          : NotSupported NotSupported
                              DcbxSupport           : None         None
                              NumTCs(Max/ETS/PFC)   : 4/4/4        4/4/4

OperationalTrafficClasses   : TC TSA    Bandwidth Priorities
                              -- ---    --------- ----------
                               0 ETS    20%       2-3,5-7
                               1 ETS    50%       4
                               2 ETS    30%       0

OperationalFlowControl      : Priority 4 Enabled
OperationalClassifications  : Protocol  Port/Type Priority
                              --------  --------- --------
                              Default             0
                              NetDirect 445       4
```

12. Create a startup script to make the settings persistent across the system reboots.
13. Run RDMA traffic and verify as described in “RoCE Configuration” on page 55.

Configuring QoS by Enabling DCBX on the Adapter

All configuration must be completed on all of the systems in use. The PFC, ETS, and traffic classes configuration must be the same on the switch and server.

To configure QoS by enabling DCBX:

1. Enable DCBX (IEEE, CEE, or Dynamic).
2. Using HII, set the **RoCE Priority** to **0**.
3. To install the DCB role in the host, issue the following Windows PowerShell command:

   ```
   PS C:\Users\Administrators> Install-WindowsFeature
   Data-Center-Bridging
   ```

   NOTE

   For this configuration, set **DCBX Protocol** to **CEE**.

4. To set the **DCBX Willing** mode to **True**, issue the following command:

   ```
   PS C:\Users\Administrators> set-NetQosDcbxSetting -Willing 1
   ```

5. Enable QoS in the miniport as follows:
   a. On the adapter's Advanced Properties page (Figure 9-11) under **Property**, select **Quality of Service**, and then set the value to **Enabled**.
   b. Click **OK**.

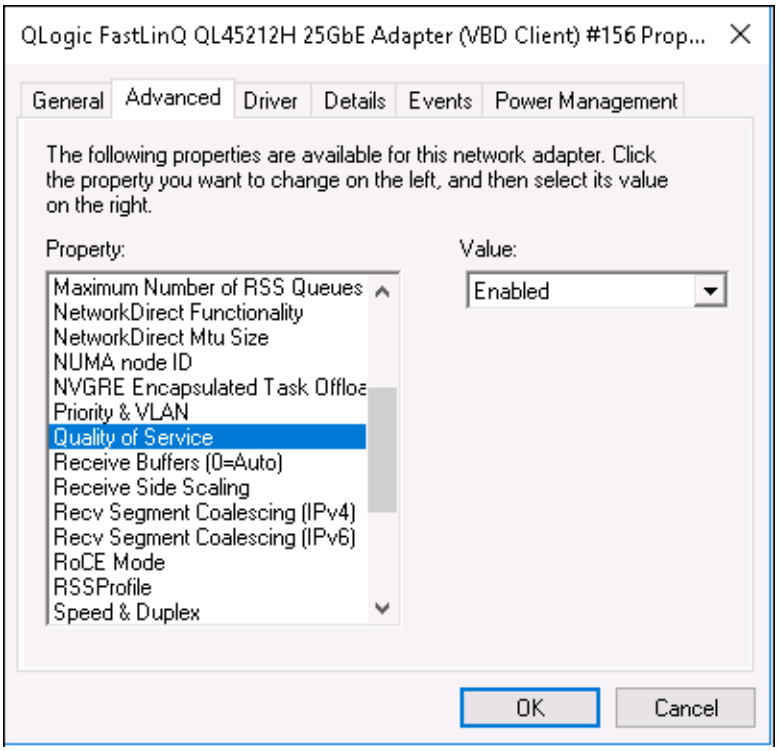


Figure 9-11. Advanced Properties: Enabling QoS

6. Assign the VLAN ID to the interface (required for PFC) as follows:
   a. Open the miniport window, and then click the **Advanced** tab.
   b. On the adapter's Advanced Properties page (Figure 9-12) under **Property**, select **VLAN ID**, and then set the value.

c. Click **OK**.

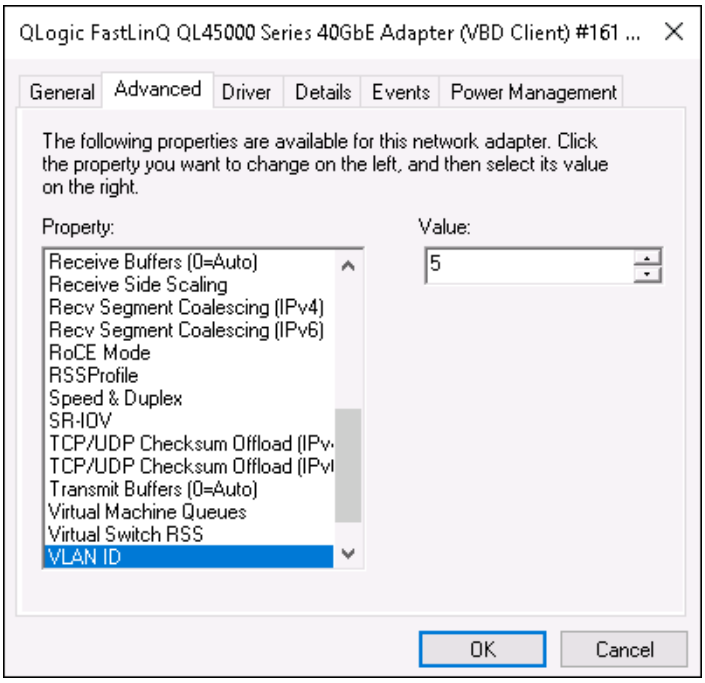


Figure 9-12. Advanced Properties: Setting VLAN ID

7. To configure the switch, issue the following Windows PowerShell command:

```
PS C:\Users\Administrators> Get-NetAdapterQoS

Name                        : Ethernet 5
Enabled                     : True
Capabilities                :                          Hardware     Current
                                                       --------     -------
                              MacSecBypass          : NotSupported NotSupported
                              DcbxSupport           : CEE          CEE
                              NumTCs(Max/ETS/PFC)   : 4/4/4        4/4/4

OperationalTrafficClasses   : TC TSA    Bandwidth Priorities
                              -- ---    --------- ----------
                               0 ETS    5%        0-3,5-7
                               1 ETS    95%       4

OperationalFlowControl      : Priority 4 Enabled
OperationalClassifications  : Protocol  Port/Type Priority
                              --------  --------- --------
```

```
                              NetDirect 445       4

RemoteTrafficClasses        : TC TSA    Bandwidth Priorities
                              -- ---    --------- ----------
                               0 ETS    5%        0-3,5-7
                               1 ETS    95%       4

RemoteFlowControl           : Priority 4 Enabled
RemoteClassifications       : Protocol  Port/Type Priority
                              --------  --------- --------
                              NetDirect 445       4
```

NOTE

The preceding example is taken when the adapter port is connected to an Arista 7060X switch. In this example, the switch PFC is enabled on Priority 4. RoCE App TLVs are defined. The two traffic classes are defined as TC0 and TC1, where TC1 is defined for RoCE. **DCBX Protocol** mode is set to **CEE**. For Arista switch configuration, refer to “Preparing the Ethernet Switch” on page 58“. When the adapter is in **Willing** mode, it accepts Remote Configuration and shows it as **Operational Parameters**.

Configuring VMMQ

Virtual machine multiqueue (VMMQ) configuration information includes:

- Enabling VMMQ on the Adapter
- Setting the VMMQ Max QPs Default and Non-Default VPort
- Creating a Virtual Machine Switch with or Without SR-IOV
- Enabling VMMQ on the Virtual Machine Switch
- Getting the Virtual Machine Switch Capability
- Creating a VM and Enabling VMMQ on VMNetworkadapters in the VM
- Default and Maximum VMMQ Virtual NIC
- Enabling and Disabling VMMQ on a Management NIC
- Monitoring Traffic Statistics

Enabling VMMQ on the Adapter

To enable VMMQ on the adapter:

1. Open the miniport window, and then click the **Advanced** tab.
2. On the Advanced Properties page (Figure 9-13) under **Property**, select **Virtual Switch RSS**, and then set the value to **Enabled**.
3. Click **OK**.

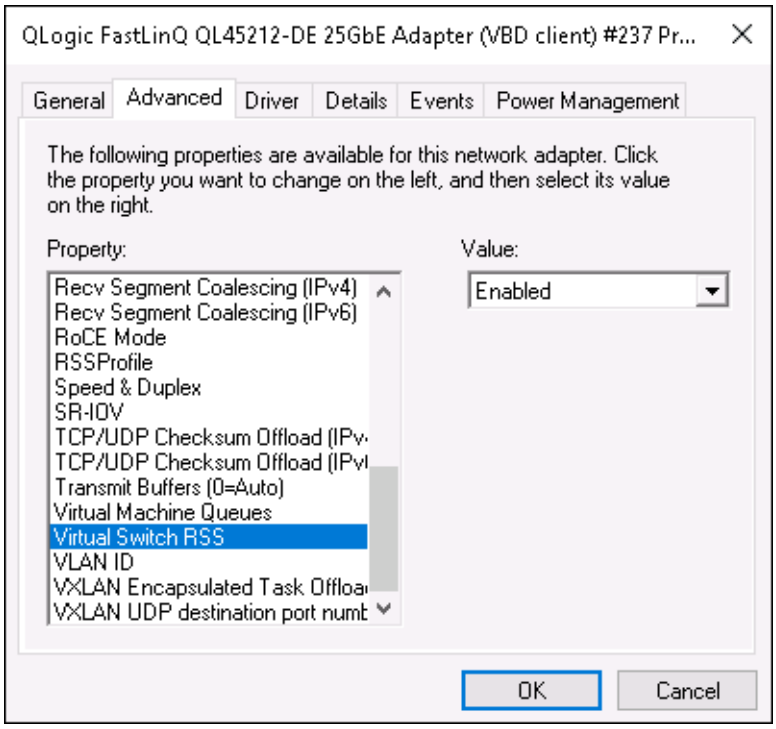


Figure 9-13. Advanced Properties: Enabling Virtual Switch RSS

Setting the VMMQ Max QPs Default and Non-Default VPort

To set the VMMQ maximum QPs default and non-default VPort:

1. Open the miniport window, and click the **Advanced** tab.
2. On the Advanced Properties page under **Property**, select one of the following:
   - ❑ **VMMQ Max QPs Default VPort**
   - ❑ **VMMQ Max QPs - Non-Default VPort**
3. If applicable, adjust the **Value** for the selected property.
4. Click **OK**.

Creating a Virtual Machine Switch with or Without SR-IOV

To create a virtual machine switch with or without SR-IOV:

1. Launch the Hyper-V Manager.
2. Select **Virtual Switch Manager** (see Figure 9-14).
3. In the **Name** box, type a name for the virtual switch.
4. Under **Connection type**:
   a. Click **External network**.
   b. Select the **Allow management operating system to share this network adapter** check box.

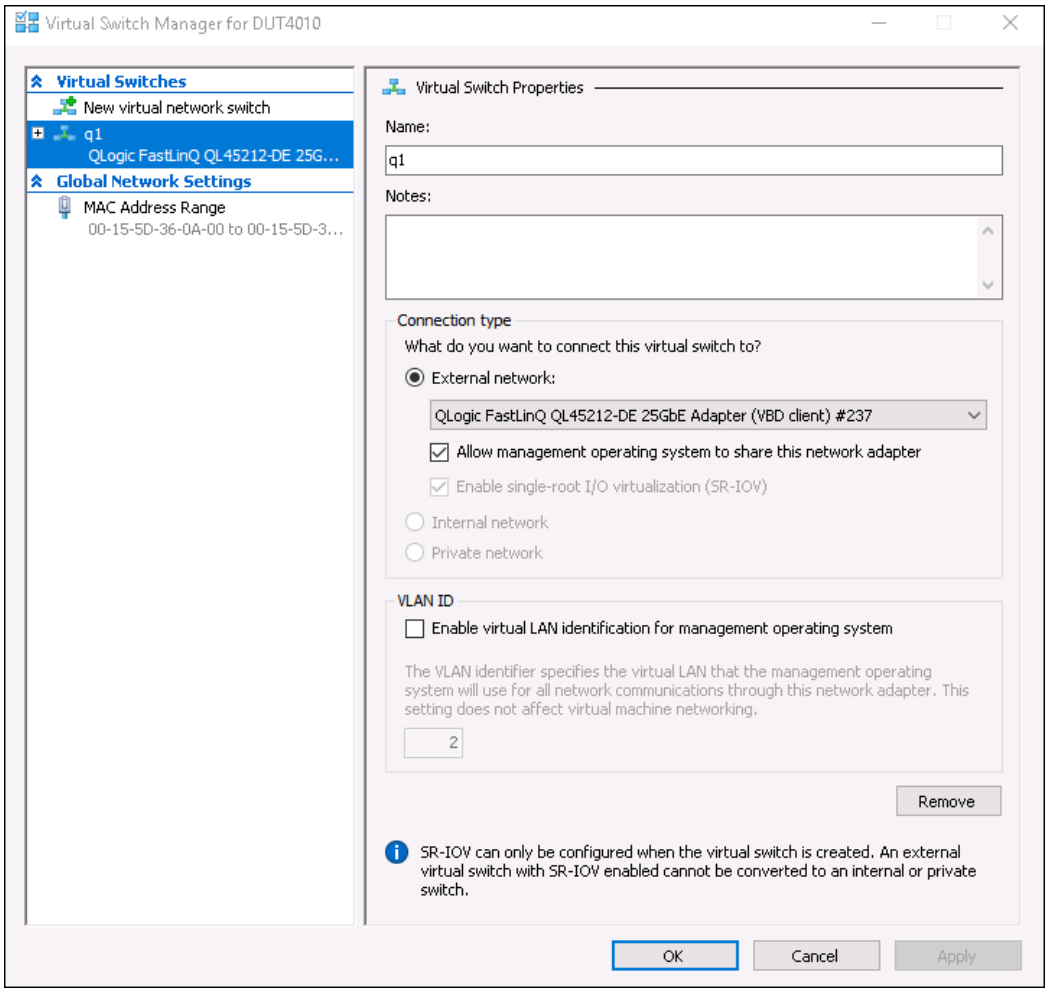


Figure 9-14. Virtual Switch Manager

5. Click **OK**.

Enabling VMMQ on the Virtual Machine Switch

To enable VMMQ on the virtual machine switch:

- Issue the following Windows PowerShell command:

```
PS C:\Users\Administrators> Set-VMSwitch -name q1
-defaultqueuevmmqenabled $true -defaultqueuevmmqqueuepairs 4
```

Getting the Virtual Machine Switch Capability

To get the virtual machine switch capability:

- Issue the following Windows PowerShell command:

```
PS C:\Users\Administrator> Get-VMSwitch -Name ql | fl
```

Figure 9-15 shows example output.

```
PS C:\Users\Administrator> Get-VMSwitch -Name q1 | fl *


Name                                : q1
Id                                  : 95344e61-d5d3-42cc-8246-e7d96d2f79d7
Notes                               :
Extensions                          : {Microsoft Windows Filtering Platform, Microsoft Azure VFP Switch Extension,
                                      Microsoft NDIS Capture}
BandwidthReservationMode            : None
PacketDirectEnabled                 : False
EmbeddedTeamingEnabled              : False
IovEnabled                          : True
SwitchType                          : External
AllowManagementOS                   : True
NetAdapterInterfaceDescription      : QLogic FastLinQ QL45212-DE 25GbE Adapter (VBD client) #237
NetAdapterInterfaceDescriptions     : {QLogic FastLinQ QL45212-DE 25GbE Adapter (VBD client) #237}
IovSupport                          : True
IovSupportReasons                   :
AvailableIPSecSA                    : 0
NumberIPSecSAAllocated              : 0
AvailableVMQueues                   : 16
NumberVmqAllocated                  : 16
IovQueuePairCount                   : 255
IovQueuePairsInUse                  : 132
IovVirtualFunctionCount             : 64
IovVirtualFunctionsInUse            : 64
PacketDirectInUse                   : False
DefaultQueueVrssEnabledRequested    : True
DefaultQueueVrssEnabled             : True
DefaultQueueVmmqEnabledRequested    : True
DefaultQueueVmmqEnabled             : True
DefaultQueueVmmqQueuePairsRequested : 4
DefaultQueueVmmqQueuePairs          : 4
BandwidthPercentage                 : 0
DefaultFlowMinimumBandwidthAbsolute : 0
DefaultFlowMinimumBandwidthWeight   : 0
CimSession                          : CimSession: .
ComputerName                        : DUT4010
IsDeleted                           : False
```

Figure 9-15. Windows PowerShell Command: Get-VMSwitch

Creating a VM and Enabling VMMQ on VMNetworkadapters in the VM

To create a virtual machine (VM) and enable VMMQ on VMNetworksadapters in the VM:

1. Create a VM.
2. Add the VMNetworkadapter to the VM.
3. Assign a virtual switch to the VMNetworkadapter.
4. To enable VMMQ on the VM, issue the following Windows PowerShell command:

```
PS C:\Users\Administrators> set-vmnetworkadapter -vmname vm1
-VMNetworkAdapterName "network adapter" -vmmqenabled $true
-vmmqqueuepairs 4
```

NOTE

For an SR-IOV capable virtual switch: If the VM switch and hardware acceleration is SR-IOV-enabled, you must create 10 VMs with 8 virtual NICs each to utilize VMMQ. This requirement is because the SR-IOV has precedence over VMMQ.

Example output of 64 virtual functions and 16 VMMQs is shown here:

```
PS C:\Users\Administrator> get-netadaptervport

Name         ID     MacAddress               VID   ProcMask   FID       State       ITR       QPairs
----         --     ----------               ---   --------   ---       -----       ---       ------
Ethernet 3   0      00-15-5D-36-0A-FB              0:0        PF        Activated   Unknown   4
Ethernet 3   1      00-0E-1E-C4-C0-A4              0:8        PF        Activated   Adaptive  4
Ethernet 3   2                                     0:0        0         Activated   Unknown   1
Ethernet 3   3                                     0:0        1         Activated   Unknown   1
Ethernet 3   4                                     0:0        2         Activated   Unknown   1
Ethernet 3   5                                     0:0        3         Activated   Unknown   1
Ethernet 3   6                                     0:0        4         Activated   Unknown   1
Ethernet 3   7                                     0:0        5         Activated   Unknown   1
Ethernet 3   8                                     0:0        6         Activated   Unknown   1
Ethernet 3   9                                     0:0        7         Activated   Unknown   1
Ethernet 3   10                                    0:0        8         Activated   Unknown   1
Ethernet 3   11                                    0:0        9         Activated   Unknown   1
.
.
.
Ethernet 3   64                                    0:0        62        Activated   Unknown   1
Ethernet 3   65                                    0:0        63        Activated   Unknown   1
Ethernet 3   66     00-15-5D-36-0A-04              0:16       PF        Activated   Adaptive  4
Ethernet 3   67     00-15-5D-36-0A-05              1:0        PF        Activated   Adaptive  4
Ethernet 3   68     00-15-5D-36-0A-06              0:0        PF        Activated   Adaptive  4
```

```
Name         ID     MacAddress                VID   ProcMask   FID       State       ITR       QPairs
----         --     ----------                ---   --------   ---       -----       ---       ------
Ethernet 3   69     00-15-5D-36-0A-07               0:8        PF        Activated   Adaptive  4
Ethernet 3   70     00-15-5D-36-0A-08               0:16       PF        Activated   Adaptive  4
Ethernet 3   71     00-15-5D-36-0A-09               1:0        PF        Activated   Adaptive  4
Ethernet 3   72     00-15-5D-36-0A-0A               0:0        PF        Activated   Adaptive  4
Ethernet 3   73     00-15-5D-36-0A-0B               0:8        PF        Activated   Adaptive  4
Ethernet 3   74     00-15-5D-36-0A-F4               0:16       PF        Activated   Adaptive  4
Ethernet 3   75     00-15-5D-36-0A-F5               1:0        PF        Activated   Adaptive  4
Ethernet 3   76     00-15-5D-36-0A-F6               0:0        PF        Activated   Adaptive  4
Ethernet 3   77     00-15-5D-36-0A-F7               0:8        PF        Activated   Adaptive  4
Ethernet 3   78     00-15-5D-36-0A-F8               0:16       PF        Activated   Adaptive  4
Ethernet 3   79     00-15-5D-36-0A-F9               1:0        PF        Activated   Adaptive  4
Ethernet 3   80     00-15-5D-36-0A-FA               0:0        PF        Activated   Adaptive  4

PS C:\Users\Administrator> get-netadaptervmq

Name         InterfaceDescription              Enabled  BaseVmqProcessor  MaxProcessors NumberOfReceive
                                                                                        Queues
----         --------------------              -------  ----------------  ------------- ---------------
Ethernet 4   QLogic FastLinQ QL45212             False    0:0                 16            1
```

Default and Maximum VMMQ Virtual NIC

According to the current implementation, a maximum quantity of 4 VMMQs is available per virtual NIC; that is, up to 16 virtual NICs.

Four default queues are available as previously set using Windows PowerShell commands. The maximum default queue can currently be set to 8. To verify the maximum default queue, use the VMswitch capability.

Enabling and Disabling VMMQ on a Management NIC

To enable or disable VMMQ on a management NIC:

- To enable VMMQ on a management NIC, issue the following command:

  ```
  PS C:\Users\Administrator> Set-VMNetworkAdapter –ManagementOS
  –vmmqEnabled $true
  ```

  The MOS VNIC has four VMMQs.

- To disable VMMQ on a management NIC, issue the following command:

  ```
  PS C:\Users\Administrator> Set-VMNetworkAdapter –ManagementOS
  –vmmqEnabled $false
  ```

  A VMMQ will also be available for the multicast open shortest path first (MOSPF).

Monitoring Traffic Statistics

To monitor virtual function traffic in a virtual machine, issue the following Windows PowerShell command:

```
PS C:\Users\Administrator> Use get-netadapterstatistics | fl
```

Configuring VXLAN

VXLAN configuration information includes:

- Enabling VXLAN Offload on the Adapter
- Deploying a Software Defined Network

Enabling VXLAN Offload on the Adapter

To enable VXLAN offload on the adapter:

1. Open the miniport window, and then click the **Advanced** tab.
2. On the Advanced Properties page (Figure 9-16) under **Property**, select **VXLAN Encapsulated Task Offload**.

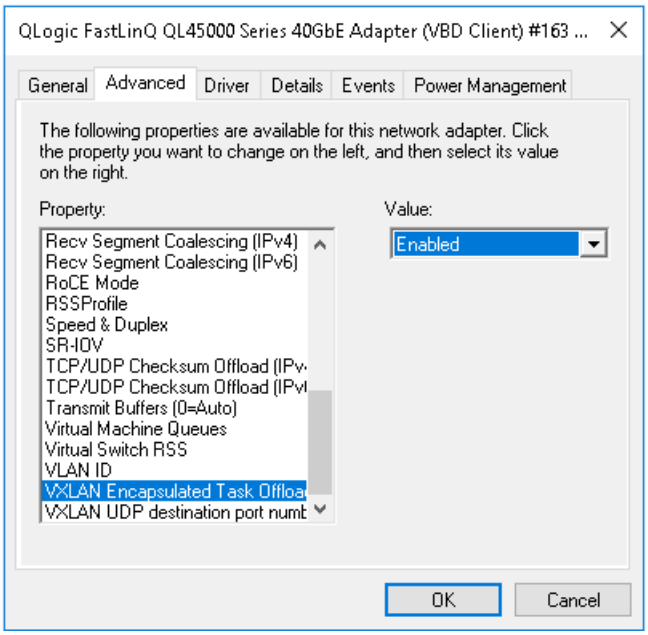


Figure 9-16. Advanced Properties: Enabling VXLAN

3. Set the **Value** to **Enabled**.
4. Click **OK**.

Deploying a Software Defined Network

To take advantage of VXLAN encapsulation task offload on virtual machines, you must deploy a Software Defined Networking (SDN) stack that utilizes a Microsoft Network Controller.

For more details, refer to the following Microsoft TechNet link on Software Defined Networking:

https://technet.microsoft.com/en-us/windows-server-docs/networking/sdn/software-defined-networking--sdn-

Configuring Storage Spaces Direct

Windows Server 2016 introduces Storage Spaces Direct, which allows you to build highly available and scalable storage systems with local storage.

For more information, refer to the following Microsoft TechnNet link:

https://technet.microsoft.com/en-us/windows-server-docs/storage/storage-spaces/storage-spaces-direct-windows-server-2016

Configuring the Hardware

Figure 9-17 shows an example of hardware configuration.

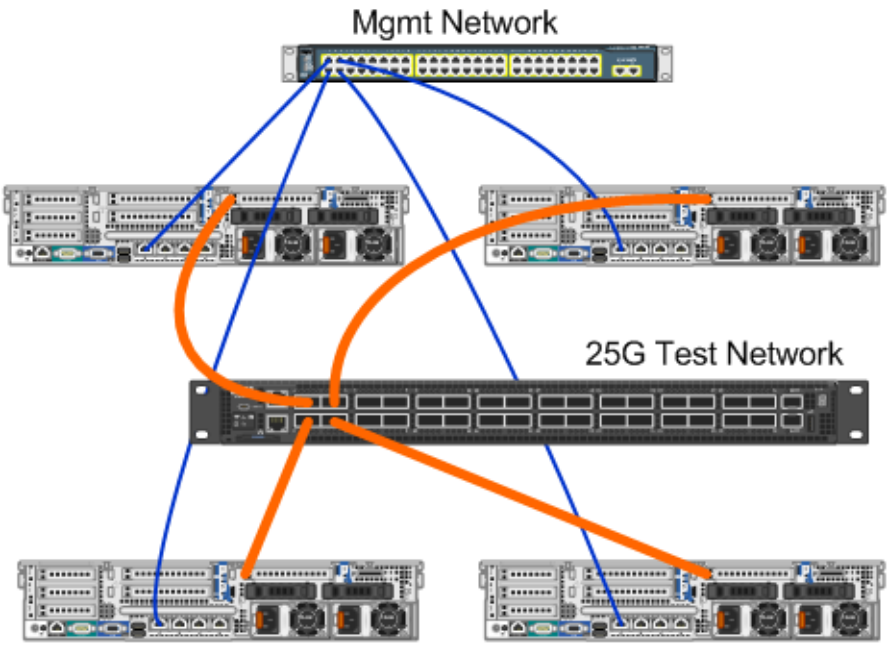


Figure 9-17. Example Hardware Configuration

NOTE

The disks used in this example are 4 × 400G NVMe™, and 12 × 200G SSD disks.

Deploying a Hyper-Converged System

This section includes instructions to install and configure the components of a Hyper-Converged system using the Windows Server 2016. The act of deploying a Hyper-Converged system can be divided into the following three high-level phases:

- Deploying the Operating System
- Configuring the Network
- Configuring Storage Spaces Direct

Deploying the Operating System

To deploy the operating systems:

1. Install the operating system.
2. Install the Windows server roles (Hyper-V).
3. Install the following features:
   - ❑ Failover
   - ❑ Cluster
   - ❑ Data center bridging (DCB)
4. Connect the nodes to domain and adding domain accounts.

Configuring the Network

To deploy Storage Spaces Direct, the Hyper-V switch must be deployed with RDMA-enabled host virtual NICs.

NOTE

The following procedure assumes that there are four RDMA NIC ports.

To configure the network on each server:

1. Configure the physical network switch as follows:
   a. Connect all adapter NICs to the switch port.

      NOTE

      If your test adapter has more than one NIC port, you must connect both ports to the same switch.

   b. Enable the switch port and make sure that the switch port supports switch-independent teaming mode, and is also part of multiple VLAN networks.

      Example Dell switch configuration:

```
no ip address
mtu 9416
portmode hybrid
switchport
dcb-map roce_S2D
protocol lldp
dcbx version cee
no shutdown
```

2. Enable **Network Quality of Service**.

NOTE

Network Quality of Service is used to ensure that the Software Defined Storage system has enough bandwidth to communicate between the nodes to ensure resiliency and performance. To configure QoS on the adapter, see “Configuring QoS for RoCE” on page 110.

3. Create a Hyper-V virtual switch with SET and RDMA virtual NIC as follows:

   a. To identify the network adapters, issue the following command:

```
Get-NetAdapter | FT
Name,InterfaceDescription,Status,LinkSpeed
```

   b. To create the virtual switch connected to all of the physical network adapters, and then enable the switch embedded teaming, issue the following command:

```
New-VMSwitch -Name SETswitch -NetAdapterName
"<port1>","<port2>","<port3>","<port4>" –
EnableEmbeddedTeaming $true
```

   c. To add host virtual NICs to the virtual switch, issue the following commands:

```
Add-VMNetworkAdapter –SwitchName SETswitch –Name SMB_1 –
managementOS
Add-VMNetworkAdapter –SwitchName SETswitch –Name SMB_2 –
managementOS
```

NOTE

The preceding commands configure the virtual NIC from the virtual switch that you just configured for the management operating system to use.

   d. To configure the host virtual NIC to use a VLAN, issue the following commands:

```
Set-VMNetworkAdapterVlan -VMNetworkAdapterName "SMB_1"
-VlanId 5 -Access -ManagementOS
Set-VMNetworkAdapterVlan -VMNetworkAdapterName "SMB_2"
-VlanId 5 -Access -ManagementOS
```

NOTE

These commands can be on the same or different VLANs.

e. To verify that the VLAN ID is set, issue the following command:

```
Get-VMNetworkAdapterVlan -ManagementOS
```

f. To disable and enable each host virtual NIC adapter so that the VLAN is active, issue the following command:

```
Disable-NetAdapter "vEthernet (SMB_1)"
Enable-NetAdapter "vEthernet (SMB_1)"
Disable-NetAdapter "vEthernet (SMB_2)"
Enable-NetAdapter "vEthernet (SMB_2)"
```

g. To enable RDMA on the host virtual NIC adapters, issue the following command:

```
Enable-NetAdapterRdma "SMB1","SMB2"
```

h. To verify RDMA capabilities, issue the following command:

```
Get-SmbClientNetworkInterface | where RdmaCapable -EQ
$true
```

Configuring Storage Spaces Direct

Configuring Storage Spaces Direct in Windows Server 2016 includes the following steps:

- Step 1. Running Cluster Validation Tool
- Step 2. Creating a Cluster
- Step 3. Configuring a Cluster Witness
- Step 4. Cleaning Disks Used for Storage Spaces Direct
- Step 5. Enabling Storage Spaces Direct
- Step 6. Creating Virtual Disks
- Step 7. Creating or Deploying Virtual Machines

Step 1. Running Cluster Validation Tool

Run the cluster validation tool to make sure server nodes are configured correctly to create a cluster using Storage Spaces Direct.

Issue the following Windows PowerShell command to validate a set of servers for use as Storage Spaces Direct cluster:

```
Test-Cluster -Node <MachineName1, MachineName2, MachineName3,
MachineName4> -Include "Storage Spaces Direct", Inventory,
Network, "System Configuration"
```

Step 2. Creating a Cluster

Create a cluster with the four nodes (which was validated for cluster creation) in Step 1. Running Cluster Validation Tool.

To create a cluster, issue the following Windows PowerShell command:

```
New-Cluster -Name <ClusterName> -Node <MachineName1, MachineName2,
MachineName3, MachineName4> -NoStorage
```

The `–NoStorage` parameter is required. If it is not included, the disks are automatically added to the cluster, and you must remove them before enabling Storage Spaces Direct. Otherwise, they will not be included in the Storage Spaces Direct storage pool.

Step 3. Configuring a Cluster Witness

You should configure a witness for the cluster, so that this four-node system can withstand two nodes failing or being offline. With these systems, you can configure file share witness or cloud witness.

For more information, go to:

https://blogs.msdn.microsoft.com/clustering/2014/03/31/configuring-a-file-share-witness-on-a-scale-out-file-server/

Step 4. Cleaning Disks Used for Storage Spaces Direct

The disks intended to be used for Storage Spaces Direct must be empty, and without partitions or other data. If a disk has partitions or other data, it will not be included in the Storage Spaces Direct system.

The following Windows PowerShell command can be placed in a Windows PowerShell script (`.PS1`) file and executed from the management system in an open Windows PowerShell (or Windows PowerShell ISE) console with Administrator privileges.

NOTE

Running this script helps identify the disks on each node that can be used for Storage Spaces Direct and removes all data and partitions from those disks.

```
icm (Get-Cluster -Name HCNanoUSClu3 | Get-ClusterNode) {
Update-StorageProviderCache
```

```
Get-StoragePool |? IsPrimordial -eq $false | Set-StoragePool
-IsReadOnly:$false -ErrorAction SilentlyContinue

Get-StoragePool |? IsPrimordial -eq $false | Get-VirtualDisk |
Remove-VirtualDisk -Confirm:$false -ErrorAction SilentlyContinue

Get-StoragePool |? IsPrimordial -eq $false | Remove-StoragePool
-Confirm:$false -ErrorAction SilentlyContinue

Get-PhysicalDisk | Reset-PhysicalDisk -ErrorAction
SilentlyContinue

Get-Disk |? Number -ne $null |? IsBoot -ne $true |? IsSystem -ne
$true |? PartitionStyle -ne RAW |% {
$_ | Set-Disk -isoffline:$false
$_ | Set-Disk -isreadonly:$false
$_ | Clear-Disk -RemoveData -RemoveOEM -Confirm:$false
$_ | Set-Disk -isreadonly:$true
$_ | Set-Disk -isoffline:$true
}
Get-Disk |? Number -ne $null |? IsBoot -ne $true |? IsSystem -ne
$true |? PartitionStyle -eq RAW | Group -NoElement -Property
FriendlyName

} | Sort -Property PsComputerName,Count
```

Step 5. Enabling Storage Spaces Direct

After creating the cluster, issue the `Enable-ClusterStorageSpacesDirect` Windows PowerShell cmdlet. The cmdlet places the storage system into the Storage Spaces Direct mode and automatically does the following:

- Creates a single large pool that has a name such as *S2D on Cluster1*.
- Configures Storage Spaces Direct cache. If there is more than one media type available for Storage Spaces Direct use, it configures the most efficient type as cache devices (in most cases, read and write).
- Creates two tiers—**Capacity** and **Performance**—as default tiers. The cmdlet analyzes the devices and configures each tier with the mix of device types and resiliency.

Step 6. Creating Virtual Disks

If the Storage Spaces Direct was enabled, it creates a single pool using all of the disks. It also names the pool (for example *S2D on Cluster1*), with the name of the cluster that is specified in the name.

The following Windows PowerShell command creates a virtual disk with both mirror and parity resiliency on the storage pool:

```
New-Volume -StoragePoolFriendlyName "S2D*" -FriendlyName
<VirtualDiskName> -FileSystem CSVFS_ReFS -StorageTierfriendlyNames
Capacity,Performance -StorageTierSizes <Size of capacity tier in
size units, example: 800GB>, <Size of Performance tier in size
units, example: 80GB> -CimSession <ClusterName>
```

Step 7. Creating or Deploying Virtual Machines

You can provision the virtual machines onto the nodes of the hyper-converged S2D cluster. Store the virtual machine's files on the system's CSV namespace (for example, `c:\ClusterStorage\Volume1`), similar to clustered virtual machines on failover clusters.

Deploying and Managing a Nano Server

Windows Server 2016 offers Nano Server as a new installation option. Nano Server is a remotely administered server operating system optimized for private clouds and data centers. It is similar to Windows Server in Server Core mode, but is significantly smaller, has no local logon capability, and supports only 64-bit applications, tools, and agents. The Nano Server takes less disk space, sets up faster, and requires fewer updates and restarts than Windows Server. When it does restart, it restarts much faster.

Roles and Features

Table 9-1 shows the roles and features that are available in this release of Nano Server, along with the Windows PowerShell options that will install the packages for them. Some packages are installed directly with their own Windows PowerShell options (such as `-Compute`). Others are installed as extensions to the `-Packages` option, which you can combine in a comma-separated list.

Table 9-1. Roles and Features of Nano Server

Role or Feature	Options
Hyper-V role	`-Compute`
Failover Clustering	`-Clustering`
Hyper-V guest drivers for hosting the Nano Server as a virtual machine	`-GuestDrivers`
Basic drivers for a variety of network adapters and storage controllers. This is the same set of drivers included in a Server Core installation of Windows Server 2016 Technical Preview.	`-OEMDrivers`

Table 9-1. Roles and Features of Nano Server (Continued)

Role or Feature	Options
File Server role and other storage components	`-Storage`
Windows Defender Antimalware, including a default signature file	`-Defender`
Reverse forwarders for application compatibility; for example, common application frameworks such as Ruby, Node.js, and others.	`-ReverseForwarders`
DNS Server Role	`-Packages Microsoft-NanoServer-DNS-Package`
Desired State Configuration (DSC)	`-Packages Microsoft-NanoServer-DSC-Package`
Internet Information Server (IIS)	`-Packages Microsoft-NanoServer-IIS-Package`
Host Support for Windows Containers	`-Containers`

System Center Virtual Machine Manager Agent	`-Packages Microsoft-Windows-Server-SCVMM-Package`
	`-Packages Microsoft-Windows-Server-SCVMM-Compute-Package`
	Note: Use this package only if you are monitoring Hyper-V. If you install this package, do not use the `-Compute` option for the Hyper-V role; instead use the `-Packages` option to install `-Packages Microsoft-NanoServer-Compute-Package`, `Microsoft-Windows-Server-SCVMM-Compute-Package`.

Network Performance Diagnostics Service (NPDS)	`-Packages Microsoft-NanoServer-NPDS-Package`
Data Center Bridging	`-Packages Microsoft-NanoServer-DCB-Package`

The next sections describe how to configure a Nano Server image with the required packages, and how to add additional device drivers specific to QLogic devices. They also explain how to use the Nano Server Recovery Console, how to manage a Nano Server remotely, and how to run Ntttcp traffic from a Nano Server.

Deploying a Nano Server on a Physical Server

Follow these steps to create a Nano Server virtual hard disk (VHD) that will run on a physical server using the preinstalled device drivers.

To deploy the Nano Server:

1. Download the Windows Server 2016 OS image.
2. Mount the ISO.
3. Copy the following files from the `NanoServer` folder to a folder on your hard drive:
   - ❑ `NanoServerImageGenerator.psm1`
   - ❑ `Convert-WindowsImage.ps1`
4. Start Windows PowerShell as an administrator.
5. Change directory to the folder where you pasted the files from Step 3.
6. Import the NanoServerImageGenerator script by issuing the following command:

   `Import-Module .\NanoServerImageGenerator.psm1 -Verbose`

7. To create a VHD that sets a computer name and includes the OEM drivers and Hyper-V, issue the following Windows PowerShell command:

NOTE

This command will prompt you for an administrator password for the new VHD.

```
New-NanoServerImage –DeploymentType Host –Edition
<Standard/Datacenter> -MediaPath <path to root of media>
-BasePath
.\Base -TargetPath .\NanoServerPhysical\NanoServer.vhd
-ComputerName
<computer name> –Compute -Storage -Cluster -OEMDrivers –
Compute
-DriversPath “<Path to Qlogic Driver sets>”
```

Example:

```
New-NanoServerImage –DeploymentType Host –Edition Datacenter
-MediaPath C:\tmp\TP4_iso\Bld_10586_iso
-BasePath ".\Base" -TargetPath
"C:\Nano\PhysicalSystem\Nano_phy_vhd.vhd" -ComputerName
"Nano-server1" –Compute -Storage -Cluster -OEMDrivers
-DriversPath
```

```
"C:\Nano\Drivers"
```

In the preceding example, `C:\Nano\Drivers` is the path for QLogic drivers. This command takes about 10 to 15 minutes to create a VHD file. A sample output for this command is shown here:

```
Windows(R) Image to Virtual Hard Disk Converter for Windows(R) 10
Copyright (C) Microsoft Corporation.  All rights reserved.
Version 10.0.14300.1000.amd64fre.rs1_release_svc.160324-1723
INFO   : Looking for the requested Windows image in the WIM file
INFO   : Image 1 selected (ServerDatacenterNano)...
INFO   : Creating sparse disk...
INFO   : Mounting VHD...
INFO   : Initializing disk...
INFO   : Creating single partition...
INFO   : Formatting windows volume...
INFO   : Windows path (I:) has been assigned.
INFO   : System volume location: I:
INFO   : Applying image to VHD. This could take a while...
INFO   : Image was applied successfully.
INFO   : Making image bootable...
INFO   : Fixing the Device ID in the BCD store on VHD...
INFO   : Drive is bootable.  Cleaning up...
INFO   : Dismounting VHD...
INFO   : Closing Windows image...
INFO   : Done.
Done. The log is at:
C:\Users\ADMINI~1\AppData\Local\Temp\2\NanoServerImageGenerator.log
```

8. Log in as an administrator on the physical server where you want to run the Nano Server VHD.
9. To copy the VHD to the physical server and configure it to boot from the new VHD:
    a. Go to **Computer Management > Storage > Disk Management**.
    b. Right-click **Disk Management** and select **Attach VHD**.
    c. Provide the VHD file path.
    d. Click **OK**.

e. Run **`bcdboot d:\windows`**.

> **NOTE**
>
> In this example, the VHD is attached under `D:\`.

f. Right-click **Disk Management** and select **Detach VHD**.

10. Reboot the physical server into the Nano Server VHD.
11. Log in to the Recovery Console using the administrator and password you supplied while running the script in Step 7.
12. Obtain the IP address of the Nano Server computer.
13. Use Windows PowerShell remoting (or other remote management) tool to connect and remotely manage the server.

Deploying a Nano Server in a Virtual Machine

To create a Nano Server virtual hard drive (VHD) to run in a virtual machine:

1. Download the Windows Server 2016 OS image.
2. Go to the `NanoServer` folder from the downloaded file in Step 1.
3. Copy the following files from the `NanoServer` folder to a folder on your hard drive:
   - ❑ `NanoServerImageGenerator.psm1`
   - ❑ `Convert-WindowsImage.ps1`
4. Start Windows PowerShell as an administrator.
5. Change directory to the folder where you pasted the files from Step 3.
6. Import the NanoServerImageGenerator script by issuing the following command:

   `Import-Module .\NanoServerImageGenerator.psm1 -Verbose`

7. Issue the following Windows PowerShell command to create a VHD that sets a computer name and includes the Hyper-V guest drivers:

> **NOTE**
>
> This following command will prompt you for an administrator password for the new VHD.

```
New-NanoServerImage –DeploymentType Guest –Edition
<Standard/Datacenter> -MediaPath <path to root of media>
-BasePath
```

```
.\Base -TargetPath .\NanoServerPhysical\NanoServer.vhd
-ComputerName
<computer name> –GuestDrivers
```

Example:

```
New-NanoServerImage –DeploymentType Guest –Edition Datacenter
-MediaPath C:\tmp\TP4_iso\Bld_10586_iso
-BasePath .\Base -TargetPath .\Nano1\VM_NanoServer.vhd
-ComputerName
Nano-VM1 –GuestDrivers
```

The preceding command takes about 10 to 15 minutes to create a VHD file. A sample output for this command follows:

```
PS C:\Nano> New-NanoServerImage –DeploymentType Guest –Edition
Datacenter -MediaPath
C:\tmp\TP4_iso\Bld_10586_iso -BasePath .\Base -TargetPath
.\Nano1\VM_NanoServer.vhd -ComputerName Nano-VM1 –GuestDrivers
cmdlet New-NanoServerImage at command pipeline position 1
Supply values for the following parameters:
Windows(R) Image to Virtual Hard Disk Converter for Windows(R) 10
Copyright (C) Microsoft Corporation.  All rights reserved.
Version 10.0.14300. 1000.amd64fre.rs1_release_svc.160324-1723
INFO   : Looking for the requested Windows image in the WIM file
INFO   : Image 1 selected (ServerTuva)...
INFO   : Creating sparse disk...
INFO   : Attaching VHD...
INFO   : Initializing disk...
INFO   : Creating single partition...
INFO   : Formatting windows volume...
INFO   : Windows path (G:) has been assigned.
INFO   : System volume location: G:
INFO   : Applying image to VHD. This could take a while...
INFO   : Image was applied successfully.
INFO   : Making image bootable...
INFO   : Fixing the Device ID in the BCD store on VHD...
INFO   : Drive is bootable.  Cleaning up...
INFO   : Closing VHD...
INFO   : Deleting pre-existing VHD : Base.vhd...
INFO   : Closing Windows image...
INFO   : Done.
```

```
Done. The log is at:
C:\Users\ADMINI~1\AppData\Local\Temp\2\NanoServerImageGenerator.log
```

8. Create a new virtual machine in Hyper-V Manager, and use the VHD created in Step 7.
9. Boot the virtual machine.
10. Connect to the virtual machine in Hyper-V Manager.
11. Log in to the Recovery Console using the administrator and password you supplied while running the script in Step 7.
12. Obtain the IP address of the Nano Server computer.
13. Use Windows PowerShell remoting (or other remote management) tool to connect and remotely manage the server.

Managing a Nano Server Remotely

Options for managing Nano Server remotely include Windows PowerShell, Windows Management Instrumentation (WMI), Windows Remote Management, and Emergency Management Services (EMS). This section describes how to access Nano Server using Windows PowerShell remoting.

Managing a Nano Server with Windows PowerShell Remoting

To manage Nano Server with Windows PowerShell remoting:

1. Add the IP address of the Nano Server to your management computer's list of trusted hosts.

NOTE

Use the recovery console to find the server IP address.

2. Add the account you are using to the Nano Server's administrators.
3. (Optional) Enable **CredSSP** if applicable.

Adding the Nano Server to a List of Trusted Hosts

At an elevated Windows PowerShell prompt, add the Nano Server to the list of trusted hosts by issuing the following command:

```
Set-Item WSMan:\localhost\Client\TrustedHosts "<IP address of Nano Server>"
```

Examples:

```
Set-Item WSMan:\localhost\Client\TrustedHosts "172.28.41.152"
Set-Item WSMan:\localhost\Client\TrustedHosts "*"
```

NOTE

The preceding command sets all host servers as trusted hosts.

Starting the Remote Windows PowerShell Session

At an elevated local Windows PowerShell session, start the remote Windows PowerShell session by issuing the following commands:

```
$ip = "<IP address of Nano Server>"
$user = "$ip\Administrator"
Enter-PSSession -ComputerName $ip -Credential $user
```

You can now run Windows PowerShell commands on the Nano Server as usual. However, not all Windows PowerShell commands are available in this release of Nano Server. To see which commands are available, issue the command `Get-Command -CommandType Cmdlet`. To stop the remote session, issue the command `Exit-PSSession`.

For more details on Nano Server, go to:

https://technet.microsoft.com/en-us/library/mt126167.aspx

Managing QLogic Adapters on a Windows Nano Server

To manage QLogic adapters in Nano Server environments, refer to the Windows QConvergeConsole GUI and Windows QLogic Control Suite CLI management tools and associated documentation, available from the Cavium Web site.

RoCE Configuration

To manage the Nano Server with Windows PowerShell remoting:

1. Connect to the Nano Server through Windows PowerShell Remoting from another machine. For example:

```
PS C:\Windows\system32> $1p="172.28.41.152"
PS C:\Windows\system32> $user="172.28.41.152\Administrator"
PS C:\Windows\system32> Enter-PSSession -ComputerName $ip
-Credential $user
```

NOTE

In the preceding example, the Nano Server IP address is `172.28.41.152` and the user name is `Administrator`.

If the Nano Server connects successfully, the following is returned:

```
[172.28.41.152]: PS C:\Users\Administrator\Documents>
```

2. To determine if the drivers are installed and the link is up, issue the following Windows PowerShell command:

   ```
   [172.28.41.152]: PS C:\Users\Administrator\Documents>
   Get-NetAdapter
   ```

   Figure 9-18 shows example output.

```
[172.28.41.152]: PS C:\Users\Administrator\Documents> Get-NetAdapter

Name                      InterfaceDescription         ifIndex Status       MacAddress             LinkSpeed
----                      --------------------         ------- ------       ----------             ---------
SLOT 1 Port 1              Ethernet 25Gb 2-port...#137       9 Up           00-0E-1E-C4-04-F8         25 Gbps
SLOT 1 Port 2              Ethernet 25Gb 2-port...#140      12 Up           00-0E-1E-C4-04-F9         25 Gbps
SLOT 2 Port 1              Ethernet 25Gb 2-port...#139       5 Up           00-0E-1E-C4-04-FA         25 Gbps
SLOT 2 Port 2              Ethernet 25Gb 2-port...#138      10 Up           00-0E-1E-C4-04-FB         25 Gbps
```

Figure 9-18. Windows PowerShell Command: Get-NetAdapter

3. To verify whether the RDMA is enabled on the adapter, issue the following Windows PowerShell command:

   ```
   [172.28.41.152]: PS C:\Users\Administrator\Documents>
   Get-NetAdapterRdma
   ```

   Figure 9-19 shows example output.

```
[172.28.41.152]: PS C:\Users\Administrator\Documents> Get-NetAdapterRdma

Name                      InterfaceDescription     Enabled
----                      --------------------     -------
SLOT 1 Port 1              Ethernet 25Gb 2-port ... True
SLOT 1 Port 2              Ethernet 25Gb 2-port ... True
SLOT 2 Port 1              Ethernet 25Gb 2-port ... True
SLOT 2 Port 2              Ethernet 25Gb 2-port ... True
```

Figure 9-19. Windows PowerShell Command: Get-NetAdapterRdma

4. To assign an IP address and VLAN ID to all interfaces of the adapter, issue the following Windows PowerShell commands:

   ```
   [172.28.41.152]: PS C:\> Set-NetAdapterAdvancedProperty
   -InterfaceAlias "slot 1 port 1" -RegistryKeyword vlanid
   -RegistryValue 5
   [172.28.41.152]: PS C:\> netsh interface ip set address
   name="SLOT 1 Port 1" static 192.168.10.10 255.255.255.0
   ```

5. To create SMBShare on the Nano Server, issue the following Windows PowerShell command:

   ```
   [172.28.41.152]: PS C:\Users\Administrator\Documents>
   New-Item -Path c:\ -Type Directory -Name smbshare -Verbose
   ```

Figure 9-20 shows example output.

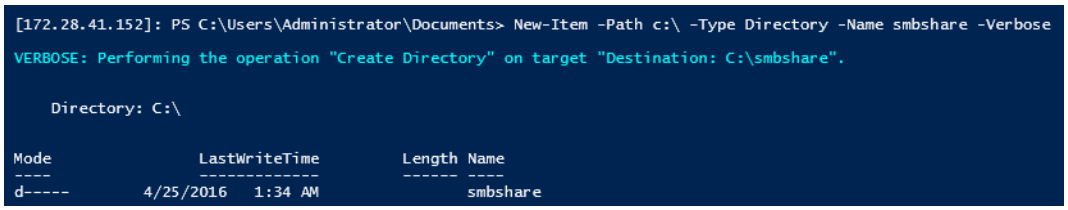


Figure 9-20. Windows PowerShell Command: New-Item

```
[172.28.41.152]: PS C:\> New-SMBShare -Name "smbshare" -Path
c:\smbshare -FullAccess Everyone
```

Figure 9-21 shows example output.

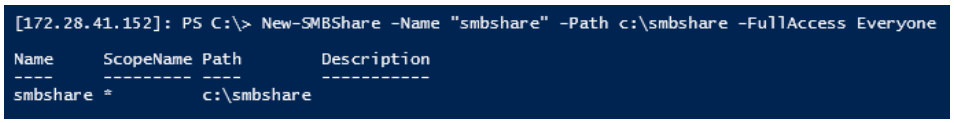


Figure 9-21. Windows PowerShell Command: New-SMBShare

6. To map the SMBShare as a network drive in the client machine, issue the following Windows PowerShell command:

NOTE

The IP address of an interface on the Nano Server is 192.168.10.10.

```
PS C:\Windows\system32> net use z: \\192.168.10.10\smbshare
This command completed successfully.
```

7. To perform read/write on SMBShare and check RDMA statistics on the Nano Sever, issue the following Windows PowerShell command:

```
[172.28.41.152]: PS C:\>
(Get-NetAdapterStatistics).RdmaStatistics
```

Figure 9-22 shows the command output.

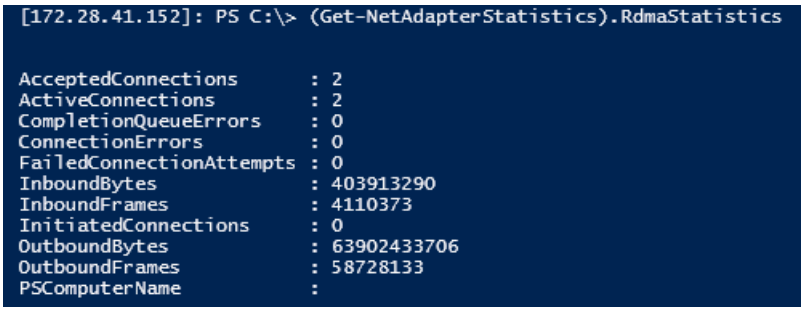


Figure 9-22. Windows PowerShell Command: Get-NetAdapterStatistics

10 Troubleshooting

This chapter provides the following troubleshooting information:

- Troubleshooting Checklist
- Verifying that Current Drivers Are Loaded
- Testing Network Connectivity
- Microsoft Virtualization with Hyper-V
- Linux-specific Issues
- Miscellaneous Issues
- Collecting Debug Data

Troubleshooting Checklist

CAUTION

Before you open the server cabinet to add or remove the adapter, review the "Safety Precautions" on page 7.

The following checklist provides recommended actions to resolve problems that may arise while installing the QL45212 Adapter or running it in your system.

- Inspect all cables and connections. Verify that the cable connections at the network adapter and the switch are attached properly.
- Verify the adapter installation by reviewing "Installing the Adapter" on page 8. Ensure that the adapter is properly seated in the slot. Check for specific hardware problems, such as obvious damage to board components or the PCI edge connector.
- Verify the configuration settings and change them if they are in conflict with another device.
- Verify that your server is using the latest BIOS.
- Try inserting the adapter in another slot. If the new position works, the original slot in your system may be defective.

- Replace the failed adapter with one that is known to work properly. If the second adapter works in the slot where the first one failed, the original adapter is probably defective.
- Install the adapter in another functioning system, and then run the tests again. If the adapter passes the tests in the new system, the original system may be defective.
- Remove all other adapters from the system, and then run the tests again. If the adapter passes the tests, the other adapters may be causing contention.

Verifying that Current Drivers Are Loaded

Ensure that the current drivers are loaded for your Windows, Linux, or VMware system.

Verifying Drivers in Windows

See the Device Manager to view vital information about the adapter, link status, and network connectivity.

Verifying Drivers in Linux

To verify that the qed.ko driver is loaded properly, issue the following command:

```
# lsmod | grep -i <module name>
```

If the driver is loaded, the output of this command shows the size of the driver in bytes. The following example shows the drivers loaded for the qed module:

```
# lsmod | grep -i qed
qed                   199238  1
qede                 1417947  0
```

If you reboot after loading a new driver, you can issue the following command to verify that the currently loaded driver is the correct version:

```
modinfo qede
```

Or, you can issue the following command:

```
[root@test1]# ethtool -i eth2
driver: qede
version: 8.4.7.0
firmware-version: mfw 8.4.7.0 storm 8.4.7.0
bus-info: 0000:04:00.2
```

If you loaded a new driver, but have not yet rebooted, the `modinfo` command will not show the updated driver information. Instead, issue the following `dmesg` command to view the logs. In this example, the last entry identifies the driver that will be active upon reboot.

```
# dmesg | grep -i "QLogic" | grep -i "qede"

[   10.097526] QLogic FastLinQ 4xxxx Ethernet Driver qede x.x.x.x
[   23.093526] QLogic FastLinQ 4xxxx Ethernet Driver qede x.x.x.x
[   34.975396] QLogic FastLinQ 4xxxx Ethernet Driver qede x.x.x.x
[   34.975896] QLogic FastLinQ 4xxxx Ethernet Driver qede x.x.x.x
[ 3334.975896] QLogic FastLinQ 4xxxx Ethernet Driver qede x.x.x.x
```

Verifying Drivers in VMware

To verify that the VMware ESXi drivers are loaded, issue the following command:

```
# esxcli software vib list
```

Testing Network Connectivity

This section provides procedures for testing network connectivity in Windows and Linux environments.

NOTE

When using forced link speeds, verify that both the adapter and the switch are forced to the same speed.

Testing Network Connectivity for Windows

Test network connectivity using the `ping` command.

To determine if the network connection is working:

1. Click **Start**, and then click **Run**.
2. In the **Open** box, type **cmd**, and then click **OK**.
3. To view the network connection to be tested, issue the following command:

   ```
   ipconfig /all
   ```

4. Issue the following command, and then press ENTER.

   ```
   ping <ip_address>
   ```

The displayed ping statistics indicate whether or not the network connection is working.

Testing Network Connectivity for Linux

To verify that the Ethernet interface is up and running:

1. To check the status of the Ethernet interface, issue the `ifconfig` command.
2. To check the statistics on the Ethernet interface, issue the `netstat -i` command.

To verify that the connection has been established:

1. Ping an IP host on the network. From the command line, issue the following command:

   `ping <ip_address>`

2. Press ENTER.

The displayed ping statistics indicate whether or not the network connection is working.

The adapter link speed can be forced to 10Gbps or 25Gbps using either the operating system GUI tool or the ethtool command, `ethtool –s ethX speed` ssss. Set both ports on the QL45212 Adapter to the same speed, both 10Gbps or both 25Gbps.

Microsoft Virtualization with Hyper-V

Microsoft Virtualization is a hypervisor virtualization system for Windows Server 2012 R2. For more information on Hyper-V, go to:

https://technet.microsoft.com/en-us/library/Dn282278.aspx

Linux-specific Issues

Problem: Errors appear when compiling driver source code.

Solution: Some installations of Linux distributions do not install the development tools and kernel sources by default. Before compiling driver source code, ensure that the development tools for the Linux distribution that you are using are installed.

Miscellaneous Issues

Problem: The QL45212 Adapter has shut down, and an error message appears indicating that the fan on the adapter has failed.

Solution: The QL45212 Adapter may intentionally shut down to prevent permanent damage. Contact QLogic Technical Support for assistance.

Collecting Debug Data

Use the commands in Table 10-1 to collect debug data.

Table 10-1. Collecting Debug Data Commands

Debug Data	Description
`demesg-T`	Kernel logs
`ethtool-d`	Register dump
`sys_info.sh`	System information; available in the driver bundle

A Adapter LEDS

Table A-1 lists the LED indicators for the state of the adapter port link and activity.

Table A-1. Adapter Port Link and Activity LEDs

Port LED	LED Appearance	Network State
Link LED	Off Continuously illuminated	No link (cable disconnected) Link
Activity LED	Off Blinking	No port activity Port activity

B Cables and Optical Modules

This appendix provides the following information for the supported cables and optical modules:

- Supported Specifications
- Tested Cables and Optical Modules

Supported Specifications

The QL45212 Adapters support a variety of cables and optical modules that comply with SFF8024. Specific form factor compliance is as follows:

- SFPs:
  - SFF8472 (for memory map)
  - SFF8419 or SFF8431 (low speed signals and power)
- Quad small form factor pluggable (QSFP):
  - SFF8636 (for memory map)
  - SFF8679 or SFF8436 (low speed signals and power)
- Optical modules electrical input/output, active copper cables (ACC), and active optical cables (AOC):
  - 10G—SFF8431 limiting interface
  - 25G—IEEE802.3by Annex 109B (25GAUI)

Tested Cables and Optical Modules

QLogic does not guarantee that every cable or optical module that satisfies the compliance requirements will operate with the QL45212 Adapters. The QL45212 Adapter supports 3m DAC and 30m AOC, but does not support 25G SR optics. QLogic has tested the QL45212 Adapter for interoperability with the cables listed in Table B-1.

Table B-1. Tested Cables and Optical Modules

Speed/Form Factor	Manufac-turer	Part Number	Type	Cable Length[a]	Gauge
Cables					
10G DAC[b]	Dell	407-BBBK	SFP+10G-to-SFP+10G	1	30
		407-BBBI	SFP+10G-to-SFP+10G	3	26
		407-BBBP	SFP+10G-to-SFP+10G	5	26
25G DAC	Amphenol®	NDCCGF0001	SFP28-25G-to-SFP28-25G	1	30
		NDCCGF0003	SFP28-25G-to-SFP28-25G	3	30
		NDCCGJ0003	SFP28-25G-to-SFP28-25G	3	26
40G DAC Splitter (4 × 10G)	Dell	470-AAVO	QSFP+40G-to-4xSFP+10G	1	26
		470-AAXG	QSFP+40G-to-4xSFP+10G	3	26
		470-AAXH	QSFP+40G-to-4xSFP+10G	5	26
100G DAC Splitter (4 × 25G)	Amphenol	NDAQGJ-0001	QSFP28-100G-to-4xSFP28-25G	1	26
		NDAQGF-0002	QSFP28-100G-to-4xSFP28-25G	2	30
		NDAQGF-0003	QSFP28-100G-to-4xSFP28-25G	3	30
	Dell	026FN3 Rev A00	QSFP28-100G-to-4XSFP28-25G	1	26
		0YFNDD Rev A00	QSFP28-100G-to-4XSFP28-25G	2	26
		07R9N9 Rev A00	QSFP28-100G-to-4XSFP28-25G	3	26

Table B-1. Tested Cables and Optical Modules (Continued)

Speed/Form Factor	Manufac-turer	Part Number	Type	Cable Length[a]	Gauge
Optical Solutions					
10G Optical Transceivers	Avago	AFBR-703SMZ	SFP+ SR	N/A	N/A
		AFBR-701SDZ	SFP+ LR	N/A	N/A
	Finisar	FTLX8571D3BCL-QL	SFP+ SR	N/A	N/A
		FTLX1471D3BCL-QL	SFP+ LR	N/A	N/A
10G AOC [c]	Dell	470-ABLV	SFP+ AOC	2	N/A
		470-ABLZ	SFP+ AOC	3	N/A
		470-ABLT	SFP+ AOC	5	N/A
		470-ABML	SFP+ AOC	7	N/A
		470-ABLU	SFP+ AOC	10	N/A
		470-ABMD	SFP+ AOC	15	N/A
		470-ABMJ	SFP+ AOC	15	N/A
25G AOC	InnoLight	TF-PY003-N00	SFP28 AOC	3	N/A
		TF-PY020-N00	SFP28 AOC	20	N/A

[a] Cable length is indicated in meters.

[b] DAC is direct attach cable.

[c] AOC is active optical cable.

Tested Switches

Table B-2 lists the switches that have been tested for interoperability with the QL45212 Adapters. This list is based on switches that are available at the time of product release, and is subject to change over time as new switches enter the market or are discontinued.

Table B-2. Switches Tested for Interoperability

Manufacturer	Ethernet Switch Model
Arista	7060X
Cisco	Nexus 3132
	Nexus 5548 and 5596T
	Nexus 6000
Dell EMC	Z9100
HPE	FlexFabric 5950
Mellanox	SN2700

C Dell Z9100 Switch Configuration

The QL45212 Adapters supports connections with the Dell Z9100 Ethernet Switch. However, until the auto-negotiation process is standardized, the switch must be explicitly configured to connect to the adapter at 25Gbps.

To configure a Dell Z9100 switch port to connect to the QL45212 Adapter at 25Gbps:

1. Establish a serial port connection between your management workstation and the switch.
2. Open a command line session, and then log in to the switch as follows:

   ```
   Login: admin
   Password: admin
   ```

3. Enable configuration of the switch port:

   ```
   Dell> enable
   Password: xxxxxx
   Dell# config
   ```

4. Identify the module and port to be configured. The following example uses module 1, port 5:

   ```
   Dell(conf)#stack-unit 1 port 5 ?
   portmode                Set portmode for a module
   Dell(conf)#stack-unit 1 port 5 portmode ?
   dual                    Enable dual mode
   quad                    Enable quad mode
   single                  Enable single mode
   Dell(conf)#stack-unit 1 port 5 portmode quad ?
   speed                   Each port speed in quad mode
   Dell(conf)#stack-unit 1 port 5 portmode quad speed ?
   10G                     Quad port mode with 10G speed
   25G                     Quad port mode with 25G speed
   ```

```
Dell(conf)#stack-unit 1 port 5 portmode quad speed 25G
```

For information about changing the adapter link speed, see “Testing Network Connectivity” on page 145.

5. Verify that the port is operating at 25Gbps:

```
Dell# Dell#show running-config | grep "port 5"
stack-unit 1 port 5 portmode quad speed 25G
```

6. To disable auto-negotiation on switch port 5, follow these steps:

   a. Identify the switch port interface (module 1, port 5, interface 1) and confirm the auto-negotiation status:

```
Dell(conf)#interface tw 1/5/1

Dell(conf-if-tf-1/5/1)#intf-type cr4 ?
autoneg                 Enable autoneg
```

   b. Disable auto-negotiation:

```
Dell(conf-if-tf-1/5/1)#no intf-type cr4 autoneg
```

   c. Verify that auto-negotiation is disabled.

```
Dell(conf-if-tf-1/5/1)#do show run interface tw 1/5/1
!
interface twentyFiveGigE 1/5/1
no ip address
mtu 9416
switchport
flowcontrol rx on tx on
no shutdown
no intf-type cr4 autoneg
```

For more information about configuring the Dell Z9100 switch, refer to the *Dell Z9100 Switch Configuration Guide* on the Dell Support Web site:

support.dell.com

Glossary

ACPI

The *Advanced Configuration and Power Interface (ACPI)* specification provides an open standard for unified operating system-centric device configuration and power management. The ACPI defines platform-independent interfaces for hardware discovery, configuration, power management, and monitoring. The specification is central to operating system-directed configuration and Power Management (OSPM), a term used to describe a system implementing ACPI, which therefore removes device management responsibilities from legacy firmware interfaces.

adapter

The board that interfaces between the host system and the target devices. Adapter is synonymous with Host Bus Adapter, host adapter, and board.

adapter port

A port on the adapter board.

Advanced Configuration and Power Interface

See ACPI.

alternative routing-ID interpretation

See ARI.

ARI

Alternative routing-ID interpretation. The PCI-SIG ARI provides a mechanism for a single PCIe device to support up to 256 physical functions (PFs). Current OSs allow up to 16 PFs (per device) to be used with ARI enabled in the system BIOS, and up to 8 PFs (per device) to be used when ARI is disabled.

bandwidth

A measure of the volume of data that can be transmitted at a specific transmission rate. A 1Gbps or 2Gbps Fibre Channel port can transmit or receive at nominal rates of 1 or 2Gbps, depending on the device to which it is connected. This corresponds to actual bandwidth values of 106MB and 212MB, respectively.

BAR

Base address register. Used to hold memory addresses used by a device, or offsets for port addresses. Typically, memory address BARs must be located in physical RAM while I/O space BARs can reside at any memory address (even beyond physical memory).

base address register

See BAR.

basic input output system

See BIOS.

BIOS

Basic input output system. Typically in Flash PROM, the program (or utility) that serves as an interface between the hardware and the operating system and allows booting from the adapter at startup.

data center bridging

See DCB.

data center bridging exchange

See DCBX.

DCB

Data center bridging. Provides enhancements to existing 802.1 bridge specifications to satisfy the requirements of protocols and applications in the data center. Because existing high-performance data centers typically comprise multiple application-specific networks that run on different link layer technologies (Fibre Channel for storage and Ethernet for network management and LAN connectivity), DCB enables 802.1 bridges to be used for the deployment of a converged network where all applications can be run over a single physical infrastructure.

DCBX

Data center bridging exchange. A protocol used by DCB devices to exchange configuration information with directly connected peers. The protocol may also be used for misconfiguration detection and for configuration of the peer.

device

A target, typically a disk drive. Hardware such as a disk drive, tape drive, printer, or keyboard that is installed in or connected to a system. In Fibre Channel, a *target device*.

DHCP

Dynamic host configuration protocol. Enables computers on an IP network to extract their configuration from servers that have information about the computer only after it is requested.

driver

The software that interfaces between the file system and a physical data storage device or network media.

dynamic host configuration protocol

See DHCP.

eCore

A layer between the OS and the hardware and firmware. It is device-specific and OS-agnostic. When eCore code requires OS services (for example, for memory allocation, PCI configuration space access, and so on) it calls an abstract OS function that is implemented in OS-specific layers. eCore flows may be driven by the hardware (for example, by an interrupt) or by the OS-specific portion of the driver (for example, loading and unloading the load and unload).

EEE

Energy-efficient Ethernet. A set of enhancements to the twisted-pair and backplane Ethernet family of computer networking standards that allows for less power consumption during periods of low data activity. The intention was to reduce power consumption by 50 percent or more, while retaining full compatibility with existing equipment. The Institute of Electrical and Electronics Engineers (IEEE), through the IEEE 802.3az task force, developed the standard.

EFI

Extensible firmware interface. A specification that defines a software interface between an operating system and platform firmware. EFI is a replacement for the older BIOS firmware interface present in all IBM PC-compatible personal computers.

energy-efficient Ethernet

See EEE.

enhanced transmission selection

See ETS.

Ethernet

The most widely used LAN technology that transmits information between computers, typically at speeds of 10 and 100 million bits per second (Mbps).

ETS

Enhanced transmission selection. A standard that specifies the enhancement of transmission selection to support the allocation of bandwidth among traffic classes. When the offered load in a traffic class does not use its allocated bandwidth, enhanced transmission selection allows other traffic classes to use the available bandwidth. The bandwidth-allocation priorities coexist with strict priorities. ETS includes managed objects to support bandwidth allocation. For more information, refer to:

http://ieee802.org/1/pages/802.1az.html

extensible firmware interface

See EFI.

FCoE

Fibre Channel over Ethernet. A new technology defined by the T11 standards body that allows traditional Fibre Channel storage networking traffic to travel over an Ethernet link by encapsulating Fibre Channel frames inside Layer 2 Ethernet frames. For more information, visit www.fcoe.com.

Fibre Channel over Ethernet

See FCoE.

file transfer protocol

See FTP.

FTP

File transfer protocol. A standard network protocol used to transfer files from one host to another host over a TCP-based network, such as the Internet. FTP is required for out-of-band firmware uploads that will complete faster than in-band firmware uploads.

human interface infrastructure

See HII.

HII

Human interface infrastructure. A specification (part of UEFI 2.1) for managing user input, localized strings, fonts, and forms, that allows OEMs to develop graphical interfaces for preboot configuration.

IEEE

Institute of Electrical and Electronics Engineers. An international nonprofit organization for the advancement of technology related to electricity.

Internet Protocol

See IP.

Internet small computer system interface

See iSCSI.

IP

Internet protocol. A method by which data is sent from one computer to another over the Internet. IP specifies the format of packets, also called *datagrams*, and the addressing scheme.

IQN

iSCSI qualified name. iSCSI node name based on the initiator manufacturer and a unique device name section.

iSCSI

Internet small computer system interface. Protocol that encapsulates data into IP packets to send over Ethernet connections.

iSCSI qualified name

See IQN.

jumbo frames

Large IP frames used in high-performance networks to increase performance over long distances. Jumbo frames generally means 9,000 bytes for Gigabit Ethernet, but can refer to anything over the IP MTU, which is 1,500 bytes on an Ethernet.

large send offload

See LSO.

Layer 2

Refers to the data link layer of the multilayered communication model, Open Systems Interconnection (OSI). The function of the data link layer is to move data across the physical links in a network, where a switch redirects data messages at the Layer 2 level using the destination MAC address to determine the message destination.

Link Layer Discovery Protocol

See LLDP.

LLDP

A vendor-neutral Layer 2 protocol that allows a network device to advertise its identity and capabilities on the local network. This protocol supersedes proprietary protocols like Cisco Discovery Protocol, Extreme Discovery Protocol, and Nortel Discovery Protocol (also known as SONMP).

Information gathered with LLDP is stored in the device and can be queried using SNMP. The topology of a LLDP-enabled network can be discovered by crawling the hosts and querying this database.

LSO

Large send offload. LSO Ethernet adapter feature that allows the TCP\IP network stack to build a large (up to 64KB) TCP message before sending it to the adapter. The adapter hardware segments the message into smaller data packets (frames) that can be sent over the wire: up to 1,500 bytes for standard Ethernet frames and up to 9,000 bytes for jumbo Ethernet frames. The segmentation process frees up the server CPU from having to segment large TCP messages into smaller packets that will fit inside the supported frame size.

maximum transmission unit

See MTU.

message signaled interrupts

See MSI, MSI-X.

MSI, MSI-X

Message signaled interrupts. One of two PCI-defined extensions to support message signaled interrupts (MSIs), in PCI 2.2 and later and PCI Express. MSIs are an alternative way of generating an interrupt through special messages that allow emulation of a pin assertion or deassertion.

MSI-X (defined in PCI 3.0) allows a device to allocate any number of interrupts between 1 and 2,048 and gives each interrupt separate data and address registers. Optional features in MSI (64-bit addressing and interrupt masking) are mandatory with MSI-X.

MTU

Maximum transmission unit. Refers to the size (in bytes) of the largest packet (IP datagram) that a specified layer of a communications protocol can transfer.

network interface card

See NIC.

NIC

Network interface card. Computer card installed to enable a dedicated network connection.

NIC partitioning

See NPAR.

non-volatile random access memory

See NVRAM.

NPAR

NIC partitioning. The division of a single NIC port into multiple physical functions or partitions, each with a user-configurable bandwidth and personality (interface type). Personalities include NIC, FCoE, and iSCSI.

NVRAM

Non-volatile random access memory. A type of memory that retains data (configuration settings) even when power is removed. You can manually configure NVRAM settings or restore them from a file.

OFED™

OpenFabrics Enterprise Distribution. An open source software for RDMA and kernel bypass applications.

PCI™

Peripheral component interface. A 32-bit local bus specification introduced by Intel®.

PCI Express (PCIe)

A third-generation I/O standard that allows enhanced Ethernet network performance beyond that of the older peripheral component interconnect (PCI) and PCI extended (PCI-X) desktop and server slots.

QoS

Quality of service. Refers to the methods used to prevent bottlenecks and ensure business continuity when transmitting data over virtual ports by setting priorities and allocating bandwidth.

quality of service

See QoS.

PF

Physical function.

RDMA

Remote direct memory access. The ability for one node to write directly to the memory of another (with address and size semantics) over a network. This capability is an important feature of VI networks.

reduced instruction set computer

See RISC.

remote direct memory access

See RDMA.

RISC

Reduced instruction set computer. A computer microprocessor that performs fewer types of computer instructions, thereby operating at higher speeds.

RDMA over Converged Ethernet

See RoCE.

RoCE

RDMA over Converged Ethernet. A network protocol that allows remote direct memory access (RDMA) over a converged or a non-converged Ethernet network. RoCE is a link layer protocol that allows communication between any two hosts in the same Ethernet broadcast domain.

SCSI

Small computer system interface. A high-speed interface used to connect devices, such as hard drives, CD drives, printers, and scanners, to a computer. The SCSI can connect many devices using a single controller. Each device is accessed by an individual identification number on the SCSI controller bus.

SerDes

Serializer/deserializer. A pair of functional blocks commonly used in high-speed communications to compensate for limited input/output. These blocks convert data between serial data and parallel interfaces in each direction.

serializer/deserializer

See SerDes.

single root input/output virtualization

See SR-IOV.

small computer system interface

See SCSI.

SR-IOV

Single root input/output virtualization. A specification by the PCI SIG that enables a single PCIe device to appear as multiple, separate physical PCIe devices. SR-IOV permits isolation of PCIe resources for performance, interoperability, and manageability.

target

The storage-device endpoint of a SCSI session. Initiators request data from targets. Targets are typically disk-drives, tape-drives, or other media devices. Typically a SCSI peripheral device is the target but an adapter may, in some cases, be a target. A target can contain many LUNs.

A target is a device that responds to a requested by an initiator (the host system). Peripherals are targets, but for some commands (for example, a SCSI COPY command), the peripheral may act as an initiator.

TCP

Transmission control protocol. A set of rules to send data in packets over the Internet protocol.

TCP/IP

Transmission control protocol/Internet protocol. Basic communication language of the Internet.

TLV

Type-length-value. Optional information that may be encoded as an element inside of the protocol. The type and length fields are fixed in size (typically 1–4 bytes), and the value field is of variable size. These fields are used as follows:

- Type—A numeric code that indicates the kind of field that this part of the message represents.
- Length—The size of the value field (typically in bytes).
- Value—Variable-sized set of bytes that contains data for this part of the message.

transmission control protocol

See TCP.

transmission control protocol/Internet protocol

See TCP/IP.

type-length-value

See TLV.

UDP

User datagram protocol. A connectionless transport protocol without any guarantee of packet sequence or delivery. It functions directly on top of IP.

UEFI

Unified extensible firmware interface. A specification detailing an interface that helps hand off control of the system for the preboot environment (that is, after the system is powered on, but before the operating system starts) to an operating system, such as Windows or Linux. UEFI provides a clean interface between operating systems and platform firmware at boot time, and supports an architecture-independent mechanism for initializing add-in cards.

unified extensible firmware interface

See UEFI.

user datagram protocol

See UDP.

VF

Virtual function.

VI

Virtual interface. An initiative for remote direct memory access across Fibre Channel and other communication protocols. Used in clustering and messaging.

virtual interface

See VI.

virtual logical area network

See VLAN.

virtual machine

See VM.

VLAN

Virtual logical area network (LAN). A group of hosts with a common set of requirements that communicate as if they were attached to the same wire, regardless of their physical location. Although a VLAN has the same attributes as a physical LAN, it allows for end stations to be grouped together even if they are not located on the same LAN segment. VLANs enable network reconfiguration through software, instead of physically relocating devices.

VM

Virtual machine. A software implementation of a machine (computer) that executes programs like a real machine.

wake on LAN

See WoL.

WoL

Wake on LAN. An Ethernet computer networking standard that allows a computer to be remotely switched on or awakened by a network message sent usually by a simple program executed on another computer on the network.

Corporate Headquarters Cavium, Inc. 2315 N. First Street San Jose, CA 95131 408-943-7100

International Offices UK | Ireland | Germany | France | India | Japan | China | Hong Kong | Singapore | Taiwan | Israel

Copyright © 2015–2018 Cavium, Inc. All rights reserved worldwide. QLogic Corporation is a wholly owned subsidiary of Cavium, Inc. QLogic, FastLinQ, QConvergeConsole, and QLogic Control Suite are registered trademarks or trademarks of Cavium, Inc. All other brand and product names are trademarks or registered trademarks of their respective owners.

This document is provided for informational purposes only and may contain errors. Cavium reserves the right, without notice, to make changes to this document or in product design or specifications. Cavium disclaims any warranty of any kind, expressed or implied, and does not guarantee that any results or performance described in the document will be achieved by you. All statements regarding Cavium's future direction and intent are subject to change or withdrawal without notice and represent goals and objectives only.