



**用户指南**  
**智能以太网适配器**  
QL45212

文档修订历史	
修订版 A, 2015 年 1 月 11 日	
修订版 B, 2016 年 4 月 19 日	
修订版 C, 2017 年 1 月 27 日	
修订版 D, 2017 年 8 月 24 日	
修订版 E, 2018 年 1 月 31 日	
修订版 F, 2018 年 4 月 19 日	
更改	受影响的章节
<p>更新了文档规范的实例。</p> <p>删除了过时的 Qlogic 许可协议和保证书章节。</p> <p>在表 3-5 中, 添加了脚注: “发布此用户指南后, 可能会提供其它的 ESXi 驱动程序。有关更多信息, 请参阅发行说明。”</p> <p>在表 6-1 中</p> <ul style="list-style-type: none"> <li>■ 更新了 Windows Server 和 VMware ESXi 的 OED 值</li> <li>■ 为 VMware ESXi 6.7 添加了一行。</li> <li>■ 删除了脚注: “此版本不包括经认证的 RoCE 驱动程序。未经认证的驱动程序作为早期预览提供。”</li> </ul> <p>添加了一个新程序, 以用于查看 Windows 上 RoCE 的 Cavium RDMA 计数器。</p> <p>添加了为 VMware ESXi 6.7 配置 iSER 的信息。</p> <p>在于 <b>Vmware 上配置 SR-IOV</b> 的程序中, 重新整理了一些步骤:</p> <ul style="list-style-type: none"> <li>■ 已将步骤“验证各端口的 VFs, 发出 esxcli 命令 .....”移至步骤“填写 Edit Setting (编辑设置) 对话框 .....”之后。</li> <li>■ 已将步骤“打开 Vm.....”移至步骤“为检测到的适配器安装 Qlogic 驱动程序 .....”之后。</li> </ul> <p>在表 B-2 中, 添加了更多的测试交换机。</p>	<p>第 xii 页上的“说明文件惯例”</p> <p>前言</p> <p>第 26 页上的“VMware 驱动程序和驱动程序包”</p> <p>第 52 页上的“支持的操作系统和 OFED”</p> <p>第 59 页上的“查看 RDMA 计数器”</p> <p>第 86 页上的“在 ESXi 6.7 上配置 iSER”</p> <p>第 103 页上的“在 VMware 上配置 SR-IOV”</p> <p>第 158 页上的“经测试的交换机”</p>

# 目录

## 前言

支持的产品 .....	xi
目标读者 .....	xi
本指南的内容 .....	xi
说明文件惯例 .....	xii
法律声明 .....	xiv
激光安全 - FDA 公告 .....	xiv
机构认证 .....	xiv
EMI 和 EMC 要求 .....	xiv
KCC: A 类 .....	xv
VCCI: A 类 .....	xvi
产品安全符合性 .....	xvi

## 1

### 产品概览

功能说明 .....	1
功能 .....	1
Dell QL45212 特定功能 .....	3
自适应中断频率 .....	3
带有嵌入式 RISC 处理器的 ASIC .....	3
适配器规格 .....	3
物理特征 .....	3
标准规格 .....	3

## 2

### 硬件安装

系统要求 .....	5
安全预防措施 .....	6
预安装核查表 .....	6
安装适配器 .....	7

## 3

### 驱动程序安装

安装 Linux 驱动程序软件 .....	8
-----------------------	---

	安装不含 RDMA 的 Linux 驱动程序.....	10
	移除 Linux 驱动程序.....	10
	使用源代码 RPM 软件包安装 Linux 驱动程序.....	12
	使用 kmp/kmod RPM 软件包安装 Linux 驱动程序.....	13
	使用 TAR 文件安装 Linux 驱动程序.....	13
	安装包含 RDMA 的 Linux 驱动程序.....	14
	Linux 驱动程序可选参数.....	15
	Linux 驱动程序操作默认设置.....	15
	Linux 驱动程序消息.....	16
	统计信息.....	16
	安装 Windows 驱动程序软件.....	16
	安装 Windows 驱动程序.....	16
	在 GUI 中运行 DUP.....	16
	DUP 安装选项.....	22
	DUP 安装示例.....	23
	移除 Windows 驱动程序.....	23
	管理适配器属性.....	23
	设置电源管理选项.....	25
	安装 VMware 驱动程序软件.....	25
	VMware 驱动程序和驱动程序包.....	26
	安装 VMware 驱动程序.....	27
	VMware 驱动程序可选参数.....	28
	VMware 驱动程序参数默认值.....	30
	移除 VMware 驱动程序.....	30
<b>4</b>	<b>升级固件</b>	
	通过双击运行 DUP.....	31
	从命令行运行 DUP.....	34
	使用 .bin 文件运行 DUP.....	35
<b>5</b>	<b>适配器预引导配置</b>	
	使用入门.....	38
	显示固件映像属性.....	41
	配置设备级参数.....	42
	配置 NIC 参数.....	43
	配置数据中心桥接.....	46
	配置分区.....	48
<b>6</b>	<b>RoCE 配置</b>	
	支持的操作系统和 OFED.....	52
	规划 RoCE.....	53

准备适配器 .....	54
准备以太网交换机.....	54
配置 Cisco Nexus 6000 以太网交换机.....	54
配置 Dell Z9100 以太网交换机 .....	56
在 Windows Server 的适配器上配置 RoCE .....	56
查看 RDMA 计数器 .....	59
在 Linux 的适配器上配置 RoCE .....	63
RHEL 的 RoCE 配置 .....	64
SLES 的 RoCE 配置 .....	64
验证 Linux 上的 RoCE 配置.....	65
VLAN 接口和 GID 索引值 .....	67
Linux 的 RoCE v2 配置 .....	68
识别 RoCE v2 GID 索引或地址.....	69
通过 sys 和 class 参数	
验证 RoCE v1 或 RoCE v2 GID 索引和地址 .....	69
通过 perfest 应用程序验证 RoCE v1 或 RoCE v2 功能.....	70
在 VMware ESX 的适配器上配置 RoCE.....	73
配置 RDMA 接口 .....	74
配置 MTU.....	75
RoCE 模式和统计信息.....	75
配置半虚拟 RDMA 设备 (PVRDMA) .....	76
<b>7</b>	
<b>iSER 配置</b>	
在您开始之前 .....	80
针对 RHEL 来配置 iSER.....	81
针对 SLES 12 来配置 iSER .....	84
优化 Linux 性能.....	85
将 CPU 配置为最高性能模式.....	85
配置内核 sysctl 设置 .....	85
配置 IRQ 关联设置.....	86
配置块设备暂存 .....	86
在 ESXi 6.7 上配置 iSER .....	86
准备工作 .....	86
为 ESXi 6.7 配置 iSER.....	87
<b>8</b>	
<b>SR-IOV 配置</b>	
在 Windows 上配置 SR-IOV .....	90
在 Linux 上配置 SR-IOV.....	97
在 VMware 上配置 SR-IOV .....	103

9	Windows Server 2016	
	使用 Hyper-V 配置 RoCE 接口	108
	创建带 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机	109
	将 VLAN ID 添加到主机虚拟 NIC	110
	验证 RoCE 是否启用	111
	添加主机虚拟 NIC（虚拟端口）	112
	映射 SMB 驱动器和运行 RoCE 流量	112
	Switch Embedded Teaming 上的 RoCE	114
	创建带 SET 和 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机	114
	在 SET 上启用 RDMA	114
	在 SET 上分配 VLAN ID	115
	在 SET 上运行 RDMA 流量	115
	为 RoCE 配置 QoS	115
	通过在适配器上禁用 DCBX 配置 QoS	116
	通过在适配器上启用 DCBX 配置 QoS	121
	配置 VMMQ	124
	在适配器上启用 VMMQ	125
	设置 VMMQ 最大 QP 默认和非默认 VPort	125
	创建带或不带 SR-IOV 的虚拟机交换机	126
	在虚拟机交换机上启用 VMMQ	128
	获取虚拟机交换机功能	128
	创建 VM 并在 VM 中的 VMNetworkadapter 上启用 VMMQ	129
	默认和最大 VMMQ 虚拟 NIC	130
	在管理 NIC 上启用和禁用 VMMQ	130
	监测流量统计信息	130
	配置 VXLAN	131
	在适配器上启用 VXLAN 卸载	131
	部署软件定义网络	132
	配置 Storage Spaces Direct	132
	配置硬件	133
	部署超聚合系统	133
	部署操作系统	133
	配置网络	134
	配置 Storage Spaces Direct	136
	部署和管理 Nano Server	138
	角色和功能	139
	在物理服务器上部署 Nano Server	140
	在虚拟机中部署 Nano Server	142

	远程管理 Nano Server .....	145
	使用 Windows PowerShell 远程处理管理 Nano Server .....	145
	将 Nano Server 添加到受信任的主机列表 .....	145
	启动远程 Windows PowerShell 会话 .....	145
	在 Windows Nano Server 上管理 QLogic 适配器 .....	146
	RoCE 配置 .....	146
10	<b>故障排除</b>	
	故障排除核查表 .....	150
	验证是否已加载最新驱动程序 .....	151
	在 Windows 中验证驱动程序 .....	151
	在 Linux 中验证驱动程序 .....	151
	在 Vmware 中验证驱动程序 .....	152
	测试网络连接 .....	152
	测试 Windows 的网络连接 .....	152
	测试 Linux 的网络连接 .....	153
	使用 Hyper-V 的 Microsoft Virtualization .....	153
	Linux 特定问题 .....	153
	其他问题 .....	153
	收集调试数据 .....	154
A	<b>适配器 LED</b>	
B	<b>电缆和光学模块</b>	
	支持的规格 .....	156
	测试的电缆和光学模块 .....	156
	经测试的交换机 .....	158
C	<b>Dell Z9100 交换机配置</b>	
	<b>词汇表</b>	

## 图片列表

图		页
3-1	Dell 更新包窗口	17
3-2	QLogic InstallShield 向导: Welcome (欢迎) 窗口	17
3-3	QLogic InstallShield 向导: License Agreement (许可协议) 窗口	18
3-4	InstallShield 向导: Setup Type (设置类型) 窗口	19
3-5	InstallShield 向导: Custom Setup (自定义设置) 窗口	20
3-6	InstallShield 向导: Ready to Install the Program (准备安装程序) 窗口	20
3-7	InstallShield 向导: Completed (完成) 窗口	21
3-8	Dell 更新包窗口	22
3-9	设置高级适配器属性	24
3-10	电源管理选项	25
4-1	Dell 更新包: 初始屏幕	31
4-2	Dell 更新包: 继续更新	32
4-3	Dell 更新包: 加载新固件	33
4-4	Dell 更新包: 安装结果	33
4-5	Dell 更新包: 完成安装	34
4-6	DUP 命令行选项	35
5-1	系统设置	38
5-2	系统设置: 设备设置	38
5-3	主要配置页面	39
5-4	主要配置页面, 将分区模式设置为 NPAR	40
5-5	主要配置页面, 设置 NPAR 分区模式	40
5-6	固件映像属性	42
5-7	系统设置: 设备级配置页面	42
5-8	NIC 配置	44
5-9	系统设置: 数据中心桥接 (DCB) 设置	46
5-10	NIC 分区配置, 全局带宽分配	48
5-11	全局带宽分配, 分区 1 配置	49
5-12	全局带宽分配页面: NPAReP 模式已启用	49
5-13	分区 1 配置	50
5-14	分区 2 配置	51
6-1	配置 RoCE 属性	57
6-2	Add Counters (添加计数器) 对话框	59
6-3	性能监测器: Cavium FastLinQ 计数器	61
6-4	交换机设置, 服务器	72
6-5	交换机设置, 客户端	72
6-6	配置 RDMA_CM 应用程序: 服务器	73
6-7	配置 RDMA_CM 应用程序: 客户端	73
6-8	配置新的分布式交换机	77
6-9	为 PVRDMA 分配 vmknic	78
6-10	设置防火墙规则	79
7-1	RDMA Ping 成功	82
7-2	利用 iSER 门户实例来测试 iSER 门户实例	82
7-3	iface 传输确认	83



7-4	检查新的 iSCSI 设备.....	83
8-1	SR-IOV 的系统设置：集成设备.....	91
8-2	SR-IOV 的系统设置：设备级配置.....	91
8-3	适配器属性，高级：启用 SR-IOV.....	92
8-4	虚拟交换机管理器：启用 SR-IOV.....	93
8-5	VM 的设置：启用 SR-IOV.....	95
8-6	设备管理器：带有 QLogic 适配器的 VM.....	96
8-7	Windows PowerShell 命令：Get-NetadapterSriovVf.....	96
8-8	系统设置：SR-IOV 的处理器设置.....	98
8-9	SR-IOV 的系统设置：集成设备.....	99
8-10	为 SR-IOV 编辑 grub.conf 文件.....	100
8-11	sriov_numvfs 的命令输出.....	101
8-12	ip link show 命令的命令输出.....	101
8-13	RHEL68 虚拟机.....	102
8-14	添加新虚拟硬件.....	103
8-15	VMware 主机编辑设置.....	106
9-1	在主机虚拟 NIC 中启用 RDMA.....	109
9-2	Hyper-V 虚拟以太网适配器属性.....	110
9-3	Windows PowerShell 命令：Get-VMNetworkAdapter.....	111
9-4	Windows PowerShell 命令：Get-NetAdapterRdma.....	111
9-5	添加计数器对话框.....	113
9-6	显示 RoCE 流量的性能监视器.....	113
9-7	Windows PowerShell 命令：New-VMSwitch.....	114
9-8	Windows PowerShell 命令：Get-NetAdapter.....	115
9-9	高级属性：启用 QoS.....	117
9-10	高级属性：设置 VLAN ID.....	118
9-11	高级属性：启用 QoS.....	122
9-12	高级属性：设置 VLAN ID.....	123
9-13	高级属性：启用虚拟交换机 RSS.....	125
9-14	虚拟交换机管理器.....	127
9-15	Windows PowerShell 命令：Get-VMSwitch.....	128
9-16	高级属性：启用 VXLAN.....	131
9-17	示例性硬件配置.....	133
9-18	Windows PowerShell 命令：Get-NetAdapter.....	147
9-19	Windows PowerShell 命令：Get-NetAdapterRdma.....	147
9-20	Windows PowerShell 命令：New-Item.....	148
9-21	Windows PowerShell 命令：New-SMBShare.....	148
9-22	Windows PowerShell 命令：Get-NetAdapterStatistics.....	149

## 表格列表

表		页
2-1	主机硬件要求 .....	5
2-2	最低主机操作系统要求 .....	5
3-1	QLogic QL45212 Adapters Linux 驱动程序 .....	8
3-2	qede 驱动程序可选参数 .....	15
3-3	Linux 驱动程序操作默认设置 .....	15
3-4	VMware 驱动程序 .....	26
3-5	按版本的 ESXi 驱动程序包 .....	26
3-6	VMware 驱动程序可选参数 .....	28
3-7	VMware 驱动程序参数默认值 .....	30
5-1	适配器属性 .....	41
6-1	RoCE v1、RoCE v2、iWARP 和 OFED 的操作系统支持 .....	52
6-2	RoCE 的高级属性 .....	56
6-3	Cavium FastLinQ RDMA 错误计数器 .....	61
9-1	Nano Server 的角色和功能 .....	139
10-1	收集调试数据命令 .....	154
A-1	适配器端口链路和活动 LED .....	155
B-1	测试的电缆和光学模块 .....	157
B-2	经测试互操作性的交换机 .....	158

# 前言

本前言列出了支持的产品，指定了目标读者，说明了本指南中使用的排版惯例，并介绍了法律声明。

## 支持的产品

本用户指南描述了 cavium™ QL45212 25Gb 智能以太网适配器（QL45212HLCU-DE 标准高度支架）的安装、配置和管理。

## 目标读者

本指南面向负责对安装在 Dell® PowerEdge® 服务器（在 Windows®、Linux® 或 VMware® 环境中）上的适配器进行配置和管理的系统管理员和其他技术人员。

## 本指南的内容

前言之后，本指南的其余部分划分为以下章节和附录：



- [第 1 章 产品概览](#) 提供产品的功能说明、功能列表以及适配器规格。
- [第 2 章 硬件安装](#) 介绍如何安装适配器，包括系统要求列表和预安装核查表。
- [第 3 章 驱动程序安装](#) 介绍在 Windows、Linux 和 VMware 中安装适配器驱动程序。
- [第 4 章 升级固件](#) 介绍如何使用 Dell Update Package (DUP) 升级适配器固件。
- [第 5 章 适配器预引导配置](#) 介绍如何使用人机界面基础设施 (HII) 应用程序执行预引导适配器配置任务。
- [第 6 章 RoCE 配置](#) 介绍如何配置适配器、以太网交换机和主机以使用基于聚合以太网的 RDMA (RoCE)。
- [第 7 章 iSER 配置](#) 描述如何针对 Linux RHEL 和 SLES 来配置 RDMA 的 iSCSI 扩展 (iSER)。
- [第 8 章 SR-IOV 配置](#) 提供用于在 Windows、Linux 和 VMware 系统中配置单根输入 / 输出虚拟化 (SR-IOV) 的过程。
- [第 9 章 Windows Server 2016](#) 介绍 Windows Server 2016 功能。
- [第 10 章 故障排除](#) 介绍各种故障排除方法和资源。

- [附录 A 适配器 LED](#) 列出适配器 LED 及其意义。
- [附录 B 电缆和光学模块](#) 列出 QL45212 Adapters 支持的电缆和光学模块。
- [附录 C Dell Z9100 交换机配置](#) 介绍如何配置 Dell Z9100 交换机端口以建立 25Gbps 连接。

本指南的最后部分是术语表。

## 说明文件惯例

本指南使用以下说明文件惯例：

- **注** 提供额外的信息。
- **小心** 不带警报符号，表示存在可能导致设备损坏或数据丢失的危险。
-  **小心** 带警报符号，表示存在可能造成轻度或中度伤害的危险。
-  **警告** 表示存在可能造成严重伤害或死亡的危险。
- **蓝色** 字体的文字表示至本指南中的插图、表格或章节的超链接（跳转），至网站的链接以 下划线蓝色 文字显示。例如：
  - [表 9-2](#) 列出与用户界面和远程代理有关的问题。
  - 请参阅 [第 6 页上的“安装核查表”](#)。
  - 有关更多信息，请访问 [www.cavium.com](http://www.cavium.com)。
- **黑体** 文字表示用户界面元素，如菜单项、按钮、复选框或列标题。例如：
  - 单击 **开始** 按钮，指向 **程序**，指向 **附件**，然后单击 **命令提示符**。
  - 在 **通知选项** 下，选中 **警报** 复选框。
- **Courier** 字体文本表示文件名、目录路径或命令行文字。例如：
  - 要从文件结构的任何地方返回根目录：  
键入 `cd /root` 并按 ENTER 键。
  - 发出以下命令：`sh ./install.bin.`
- 键盘的键名和击键用大写字母表示：
  - 按 CTRL+P。
  - 按向上箭头键。

- *斜体文字*表示术语、强调、变量或说明文件标题。例如：
  - 哪些是*快捷键*?
  - 要输入日期，键入 *mm/dd/yyyy*（其中 *mm* 是月，*dd* 是日，*yyyy* 是年）。
- 引号中的主题标题表示本手册中的联机帮助的相关主题；在本说明文件中，联机帮助又称为*帮助系统*。
- 命令行界面 (CLI) 命令语法惯例包括以下内容：
  - 纯文本表示必须按照所示键入的项目。例如：
    - `qaucli -pr nic -ei`
  - `< >`（尖括号）表示必须指定其值的变量。例如：
    - `<serial_number>`

---

### 注

仅适用于 CLI 命令，变量名称始终使用尖括号而不是*斜体*表示。

---

- [ ]（方括号）表示可选的参数。例如：
  - [`<file_name>`] 表示指定文件名，或省略以选择默认文件名。
- |（竖线）表示互斥的选项；只能选择一个选项。例如：
  - `on|off`
  - `1|2|3|4`
- ...（省略号）表示前面的项可重复。例如：
  - `x...` 表示 `x` 的一个或多个实例。
  - [`x...`] 表示 `x` 的零个或多个实例。
- 命令示例输出中的垂直省略号指示重复输出数据的哪些部分被特意省略。
- ( )（圆括号）和 { }（大括号）用来避免逻辑模糊不清。例如：
  - `a|b c` 模糊不清
  - `{(a|b) c}` 表示 `a` 或 `b`，后跟 `c`
  - `{a|(b c)}` 表示 `a` 或 `b c`

## 法律声明

本节中涵盖的法律声明包括激光安全（FDA 公告）、机构认证和产品安全合规性。

### 激光安全 - FDA 公告

本产品符合 DHHS 规则 21CFR I 章 J 节的规定。本产品的设计和和生产符合 IEC60825-1 中有关激光产品安全标签的规定。

I 类激光产品

1 类 激光产品	<b>小心</b> — 在打开时存在 1 类激光辐射 请勿直视光学仪器
Appareil laser de classe 1	<b>Attention</b> —Radiation laser de classe 1 Ne pas regarder directement avec des instruments optiques
Produkt der Laser Klasse 1	<b>Vorsicht</b> —Laserstrahlung der Klasse 1 bei geöffneter Abdeckung Direktes Ansehen mit optischen Instrumenten vermeiden
Luokan 1 Laserlaite	<b>Varoitus</b> —Luokan 1 lasersäteilyä, kun laite on auki Älä katso suoraan laitteeseen käyttämällä optisia instrumenttejä

### 机构认证

以下章节概述对 QL45212 Adapters 执行的 EMC 和 EMI 检验规格，验证其是否符合辐射排放、抗辐射干扰性以及产品安全标准。

#### EMI 和 EMC 要求

##### FCC 第 15 部分符合性：A 级

**FCC 符合性信息声明：**本设备符合 FCC 规则第 15 部分的规定。操作受限于以下两个条件：(1) 此设备不得造成有害干扰；(2) 此设备必须能够承受接收到的任何干扰，包括可能导致不良操作的干扰。

##### ICES-003 符合性：A 级

本 A 级数字装置符合加拿大 ICES-003 的规定。Cet appareil numérisé de la classe A est conforme à la norme NMB-003 du Canada.

##### CE 标记 2014/30/EU、2014/35/EU EMC 指令符合性：

EN55032:2012/ CISPR 32:2015 A 类

EN55024:2010

EN61000-3-2: 谐波电流发射

EN61000-3-3: 电压波动和闪动

抗扰性标准

EN61000-4-2: ESD  
EN61000-4-3: 射频电磁场  
EN61000-4-4: 快速瞬变 / 爆发  
EN61000-4-5: 快速电涌常见 / 差动  
EN61000-4-6: 射频传导敏感度  
EN61000-4-8: 电力频率电磁场  
EN61000-4-11: 电压骤降和中断

**VCCI: 2015-04 ; A 类**

**AS/NZS ; CISPR 32: 2015 A 类**

**CNS 13438: 2006 A 类**

**KCC: A 类**

经过韩国 RRA A 级认证



产品名称 / 型号: 聚合网络适配器和智能以太网适配器  
证书持有者: QLogic Corporation  
制造日期: 请参阅产品上的日期代码  
制造商 / 原产国: QLogic Corporation/ 美国

A 级设备  
(商用信息 / 电讯设备)

由于本设备已就其商用性执行了 EMC 注册, 因此要求销售方和 / 或购买方对此引起充分注意, 倘若发生错误的销售或购买行为, 要求立即将其更换成家用型设备。

韩国语言格式 - A 级

**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: A类

基于 Voluntary Control Council for Interference(VCCI) 的标准, 该设备为 A 类产品。如果在家庭环境中使用该设备, 可能会发生无线电干扰, 在这种情况下, 用户可能需要采取纠正措施。

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

## 产品安全符合性

### UL、cUL 产品安全:

UL 60950-1 (第二版) A1 + A2 2014-10-14

CSA C22.2 No.60950-1-07 (第二版) A1 +A2 2014-10

仅与所列出的 ITE 或等同物一起使用。

符合 21 CFR 1040.10 和 1040.11、2014/30/EU、2014/35/EU。

### 2006/95/EC 低电压指令:

TUV EN60950-1:2006+A11+A1+A12+A2 第二版

TUV IEC 60950-1: 2005 第二版 Am1: 2009 + Am2: 2013 CB

CB 被认证为 IEC 60950-1 第二版



# 1 产品概览

本章提供 QL45212 Adapters 的以下信息：

- [功能说明](#)
- [功能](#)
- [第 3 页上的“适配器规格”](#)

## 功能说明

Cavium QL45212 适配器是一种 25Gb 智能以太网适配器，该适配器设计用于执行 Dell PowerEdge 系统的加速数据联网。QL45212 适配器包括 25Gb 以太网 MAC（具有全双工能力）。

## 功能

QL45212 适配器具有下列特性：

- NIC 分区 (NPAR)
- 单芯片解决方案：
  - 25Gb MAC
  - 用于直连式铜缆 (DAC) 收发器连接的 SerDes 接口
  - PCIe® 3.0 x8
  - 具备零拷贝功能的硬件
- 性能特点：
  - TCP、IP、UDP 校验和卸载
  - TCP 分段卸载 (TSO)
    - 大段卸载 (LSO)
    - 普通分段卸载 (GSO)
    - 大量接收卸载 (LRO)
    - 接收分段结合 (RSC)

- Microsoft® 动态虚拟机队列 (VMQ)、虚拟机多队列 (VMMQ) 和 Linux 多队列
- 自适应中断：
  - 发送方 / 接收方缩放 (TSS/RSS)
  - 使用通用路由封装 (NVGRE) 和虚拟 LAN (VXLAN) L2/L3 GRE 隧道流量进行网络虚拟化无状态卸载<sup>1</sup>
- 可管理性：
  - 系统管理总线 (SMB) 控制器
  - 符合 *高级配置与电源接口* (ACPI) 1.1a 标准 (多电源模式)
  - 网络控制器边带接口 (NC-SI) 支持
- 高级网络特性：
  - 巨型帧 (多达 9,600 个字节)。操作系统和链路伙伴必须支持巨型帧。
  - 虚拟 LAN (VLAN)
  - 流控制 (IEEE Std 802.3x)
- 逻辑链路控制 (IEEE 标准 802.2)
- 高速芯片搭载精简指令集计算机 (RISC) 处理器
- 集成式 96KB 帧缓冲区存储区 (并非适用于所有型号)
- 1,024 分类过滤器 (并非适用于所有型号)
- 通过 128 位散列硬件功能支持多播地址
- 串行 NVRAM 闪存
- *PCI 电源管理接口* (v1.1)
- 64 位基本地址寄存器 (BAR) 支持
- EM64T 处理器支持
- iSCSI 和 FCoE 引导支持<sup>2</sup>

---

<sup>1</sup> 此功能要求操作系统或监控程序支持以使用卸载。

<sup>2</sup> SR-IOV VF 的硬件支持限制有所不同。该限制在某些操作系统环境中可能较低；请参阅您的操作系统的相应章节。

## Dell QL45212 特定功能

Dell QL45212 智能以太网适配器特定的功能包括自适应中断频率和带有嵌入式 RISC 处理器的 ASIC。

### 自适应中断频率

适配器驱动程序根据流量条件智能调整主机中断频率，从而提高整个应用程序吞吐量。流量低时，适配器驱动程序为每个接收的数据包中断主机，以将等待时间降到最低。流量高时，适配器对多个连续的传入数据包发出一次主机中断，从而保持主机 CPU 周期。

### 带有嵌入式 RISC 处理器的 ASIC

QL45212 Adapter 的核心控制驻留在紧密集成的高性能 ASIC 中。ASIC 包括一个 RISC 处理器，该处理器能灵活地将新功能添加至网卡，并通过软件下载使其适应将来的网络要求。当主机操作系统得到增强以利用适配器上的这些功能时，RISC 处理器也使适配器驱动程序可充分利用适配器上的内置主机卸载功能。

## 适配器规格

QL45212 Adapter 规格包括适配器的物理特征和标准符合性引用。

### 物理特征

QL45212 Adapters 为标准的 PCIe 卡并附带全高或薄型支架以在标准 PCIe 插槽中使用。

### 标准规格

支持的标准规格包括：

- *PCI Express 基本规格*，修订版 3.1
- *PCI Express 卡机电规格*，修订版 3.0
- *PCI 总线电源管理接口规格*，修订版 1.2
- IEEE 规格：
  - 以太网的 802.3-2015 IEEE 标准*（流控制）
  - 802.1q (VLAN)*
  - 802.1AX*（链路聚合）
  - 802.1ad (QinQ)*
  - 802.1p*（优先级编码）
  - 1588-2002 PTPv1*（精确时间协议）

- 1588-2008 PTPv2
- IEEE 802.3az 节能以太网 (EEE)
- IPv4 (RFQ 791)
- IPv6 (RFC 2460)

# 2 硬件安装

本章提供以下硬件安装信息：

- 系统要求
- 第 6 页上的“安全预防措施”
- 第 6 页上的“预安装核查表”
- 第 7 页上的“安装适配器”

## 系统要求

安装 Cavium QL45212 Adapter 前，请确认您的系统满足表 2-1 和表 2-2 所示的硬件和操作系统要求。有关支持的操作系统的完整列表，请访问 Cavium 网站。

**表 2-1. 主机硬件要求**

硬件	要求
体系结构	可满足操作系统要求的 IA-32 或 EMT64
内存	8GB RAM（最少）
直连电缆 (DAC)	Amphenol® NDDCCGF-0001 Amphenol NDAQGF-0003 Amphenol NDCCGF-0005 Leoni® ParaLink® LA0SF064-SD-R

**表 2-2. 最低主机操作系统要求**

操作系统	要求
Windows Server	2012、2012 R2、2016（包括 Nano）
Linux	RHEL® 6.8、6.9、7.2、7.3、7.4 SLES® 11 SP4、SLES 12 SP2、SLES 12 SP3
VMware	用于 25G 适配器的 ESXi 6.0 u3 及更高版本

### 注

表 2-2 说明最低主机操作系统要求。有关支持的操作系统的完整列表，请访问 Cavium 网站。

---

## 安全预防措施

### 警告

安装适配器的系统的操作电压可能会有致命危险。打开系统外壳之前，请遵从以下预防措施以保护您自己并避免损坏系统组件。

- 除去手上和手腕上的任何金属物体或首饰。
  - 确保仅使用绝缘工具或非导电工具。
  - 在触摸内部组件之前，确认系统电源已关闭并已拔掉电源线。
  - 在不受静电干扰的环境中安装或卸下适配器。使用正确接地的腕带或其他人体防静电设备，强烈建议使用防静电地垫。
- 

## 预安装核查表

安装适配器之前，请完成以下操作：

1. 确认您的系统满足第 5 页上的“系统要求”中列出的硬件和软件要求。
2. 确认您的系统使用最新的 BIOS。

### 注

如果您从 Cavium 网站获取适配器软件，请确认适配器驱动程序文件的路径。

---

3. 如果系统正在运行，请将其关闭。
4. 系统关闭后，断开电源并拔下电源线。
5. 将适配器从其运输包装中取出并放在防静电表面上。
6. 检查适配器，特别是边缘连接器上是否有明显的损坏痕迹。切勿尝试安装损坏的适配器。

## 安装适配器

以下说明适用于在大多数系统中安装 Cavium QL45212 Adapters。有关执行这些任务的详细信息，请参阅随系统提供的手册。

### 要安装适配器：

1. 复查第 6 页上的“安全预防措施”和第 6 页上的“预安装核查表”。安装适配器前，确保系统电源已关闭而且电源线已从电源插座上拔下，并且遵守适当的电接地步骤。
2. 打开系统机箱，然后选择符合适配器大小的插槽，可以是 PCIe Gen 2 x8 或 PCIe Gen 3 x8。较窄的适配器可插入更宽的插槽中（x8 的适配器可插入 x16 的插槽中），但较宽的适配器不能插入更窄的插槽中（x8 的适配器不能插入 x4 的插槽中）。如果不知道如何识别 PCIe 插槽，请参考系统说明文件。
3. 从选择的插槽卸下空挡板。
4. 将适配器连接器边缘与系统中的 PCIe 连接器插槽对齐。
5. 在适配器卡的两个边角均匀施压以推进插卡，直至其牢固就位在插槽中。当适配器正确就位时，其端口连接器将与插槽开口处对齐，适配器面板将与系统机箱齐平。

---

### 小心

将插卡推进到位时不要过度用力，否则可能损坏系统或适配器。如果无法固定适配器，将其卸下，重新对齐，并再次尝试。

---

6. 使用适配器夹或螺丝固定适配器。
7. 合上系统机箱，并断开任何个人防静电设备。

# 3 驱动程序安装

本章提供有关驱动程序安装的以下信息：

- 安装 Linux 驱动程序软件
- 第 16 页上的“安装 Windows 驱动程序软件”
- 第 25 页上的“安装 VMware 驱动程序软件”

## 安装 Linux 驱动程序软件

本节描述如何安装具有或不具有远程直接内存访问 (RDMA) 的 Linux 驱动程序。本节还描述 Linux 驱动程可选参数、默认值、消息和统计信息。

- 安装不含 RDMA 的 Linux 驱动程序
- 安装包含 RDMA 的 Linux 驱动程序
- Linux 驱动程序可选参数
- Linux 驱动程序操作默认设置
- Linux 驱动程序消息
- 统计信息

Dell 支持页面上提供了 QL45212 Adapter Linux 驱动程序和支持说明文件：

[dell.support.com](http://dell.support.com)

表 3-1 介绍了 QL45212 Adapter Linux 驱动程序。

**表 3-1. QLogic QL45212 Adapters Linux 驱动程序**

Linux 驱动程序	说明
qed	qed 核心驱动程序模块直接控制固件、处理中断并为协议特定驱动程序集提供低级别 API。qed 与 qede、qedr、qedi 和 qedf 驱动程序接合。Linux 核心模块管理所有 PCI 设备资源（寄存器、主机接口队列，等等）。qed 核心模块需要使用 Linux 内核版本 2.6.32 或更高版本。测试集中于 x86_64 体系结构。



表 3-1. QLogic QL45212 Adapters Linux 驱动程序 (续)

Linux 驱动程序	说明
qede	适用于 QL45212 Adapter 的 Linux 以太网驱动程序。该驱动程序直接控制硬件，并负责代表 Linux 主机网络堆栈发送和接收以太网数据包。该驱动程序还代表其自身接收和处理设备中断（适用于 L2 网络）。qede 驱动程序需要使用 Linux 内核版本 2.6.32 或更高版本。测试集中于 x86_64 体系结构。
qedr	Linux 基于聚合以太网的 RDMA (RoCE) 驱动程序。此驱动程序在开放结构企业分布 (OFED™) 环境中与 qed 核心模块和 qede 以太网驱动程序配合使用。RDMA 用户空间应用程序还需要在服务器上安装 libqedr 用户库。
qedi	适用于 45xxx 适配器的 Linux iSCSI 卸载驱动程序。此驱动程序与 Open iSCSI 库一起使用。
qedf	适用于 45xxx 适配器的 Linux FCoE 卸载驱动程序。此驱动程序与 Open FCoE 库一起使用。

可以使用源 Red Hat® Package Manager (RPM) 包或 kmod RPM 包安装 Linux 驱动程序。RHEL RPM 包内容如下：

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

SLES 源和 kmp RPM 包内容如下：

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

以下内核模块 (kmod) RPM 会在运行 Xen 虚拟机监控程序的 SLES 主机上安装 Linux 驱动程序：

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

以下源 RPM 会在 RHEL 和 SLES 主机上安装 RDMA 库代码：

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

以下源代码 TAR BZip2 (BZ2) 压缩文件会在 RHEL 和 SLES 主机上安装 Linux 驱动程序：

- fastlinq-<version>.tar.bz2

### 注

对于通过 NFS、FTP 或 HTTP 执行（使用网络引导盘）的网络安装，可能需要使用其中包含 qede 驱动程序的驱动程序盘。可以通过修改 makefile 和 make 环境来编译 Linux 引导驱动程序。

---

## 安装不含 RDMA 的 Linux 驱动程序

要安装不含 RDMA 的 Linux 驱动程序：

1. 从 Dell 下载 QL45212 Adapter Linux 驱动程序：  
[dell.support.com](http://dell.support.com)
2. 如第 10 页上的“移除 Linux 驱动程序”中所述，移除现有 Linux 驱动程序。
3. 使用以下方法之一安装新 Linux 驱动程序：
  - 使用源代码 RPM 软件包安装 Linux 驱动程序
  - 使用 kmp/kmod RPM 软件包安装 Linux 驱动程序
  - 使用 TAR 文件安装 Linux 驱动程序

## 移除 Linux 驱动程序

移除 Linux 驱动程序有两种步骤：一种用于非 RDMA 环境，另一种用于 RDMA 环境。选择符合您的环境的步骤。

**要在非 RDMA 环境中移除 Linux 驱动程序，请卸载并移除驱动程序：**

按照与原始安装方法和操作系统相关的步骤进行操作。

- 如果以前是使用 RPM 软件包安装的 Linux 驱动程序，请发出以下命令：

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

- 如果以前是使用 TAR 文件安装的 Linux 驱动程序，请发出以下命令：

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

- ❑ 对于 RHEL:  

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

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

#### 要在非 RDMA 环境中移除 Linux 驱动程序:

1. 要获取当前安装的驱动程序的路径, 请发出以下命令:  

```
modinfo <driver name>
```
2. 卸载并移除 Linux 驱动程序。
  - ❑ 如果以前是使用 RPM 软件包安装的 Linux 驱动程序, 请发出以下命令:  

```
modprobe -r qede
depmod -a
rpm -e qlgc-fastlinq-kmp-default-<version>.<arch>
```
  - ❑ 如果以前是使用 TAR 文件安装的 Linux 驱动程序, 请发出以下命令:  

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

---

#### 注

如果存在 qedr, 请发出 `modprobe -r qedr` 命令代替。

---

3. 从 `qed.ko`、`qede.ko` 和 `qedr.ko` 文件所在的目录中删除它们。例如, 在 SLES 中发出以下命令:  

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

#### 要在 RDMA 环境中移除 Linux 驱动程序:

1. 要获取所安装的驱动程序的路径, 请发出以下命令:  

```
modinfo <driver name>
```

2. 卸载并移除 Linux 驱动程序。

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

3. 移除驱动程序模块文件:

- ❑ 如果以前是使用 RPM 软件包安装的驱动程序, 请发出以下命令:

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

- ❑ 如果以前是使用 TAR 文件安装的驱动程序, 则针对您的操作系统发出以下命令:

对于 RHEL:

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

对于 SLES:

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

## 使用源代码 RPM 软件包安装 Linux 驱动程序

要使用源代码 RPM 软件包安装 Linux 驱动程序:

1. 在命令提示符处发出以下命令:

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

2. 将目录更改至 RPM 路径并针对内核构建二进制 RPM。

对于 RHEL:

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

对于 SLES:

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

3. 安装新编译的 RPM:

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

**注**

如果报告了冲突，则可能需要对某些 Linux 分发版使用 `--force` 选项。

---

驱动程序将安装在以下路径中。

对于 SLES:

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

对于 RHEL:

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

4. 开启所有 ethX 接口，如下所示：  
`ifconfig <ethX> up`
5. 对于 SLES，使用 YaST 来配置以太网接口以在引导时自动启动（通过设置静态 IP 地址或启用接口上的 DHCP 实现）。

### 使用 kmp/kmod RPM 软件包安装 Linux 驱动程序

要安装 kmod RPM 软件包:

1. 在命令提示符处发出以下命令：  
`rpm -ivh qlgc-fastlinq-<version>.<arch>.rpm`
2. 重新加载驱动程序：  
`modprobe -r qede`  
`modprobe qede`

### 使用 TAR 文件安装 Linux 驱动程序

要使用 TAR 文件安装 Linux 驱动程序:

1. 创建目录并将 TAR 文件解压缩到该目录：  
`tar xjvf fastlinq-<version>.tar.bz2`
2. 切换到最近创建的目录，然后安装驱动程序：  
`cd fastlinq-<version>`  
`make clean; make install`  
  
qed 和 qede 驱动程序将安装在以下路径中。

对于 SLES:

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

对于 RHEL:

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

3. 加载驱动程序进行测试（如有必要，先卸载现有驱动程序）:

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

## 安装包含 RDMA 的 Linux 驱动程序

要在内建 OFED 环境中安装 Linux 驱动程序:

1. 从 Dell 下载 QL45212 Adapter Linux 驱动程序:  
[dell.support.com](http://dell.support.com)
2. 如第 63 页上的“在 Linux 的适配器上配置 RoCE”中所述，在适配器上配置 RoCE。
3. 如第 10 页上的“移除 Linux 驱动程序”中所述，移除现有 Linux 驱动程序。
4. 使用以下方法之一安装新 Linux 驱动程序:
  - 使用 kmp/kmod RPM 软件包安装 Linux 驱动程序
  - 使用 TAR 文件安装 Linux 驱动程序
5. 安装 libqedr 库以使用 RDMA 用户空间应用程序。libqedr RPM 仅可用于内建 OFED。发出以下命令:  

```
rpm -ivh qlgc-libqedr-<version>.<arch>.rpm
```
6. 要构建和安装 libqedr 用户空间库，请发出以下命令:  

```
'make libqedr_install'
```
7. 通过加载驱动程序来测试它们，如下所示:  

```
modprobe qedr  
make install_libqedr
```

## Linux 驱动程序可选参数

表 3-2 介绍 qede 驱动程序的可选参数。

表 3-2. qede 驱动程序可选参数

参数	说明
debug	控制驱动程序详细级别，与 <code>ethtool -s &lt;dev&gt; msglvl</code> 类似。
int_mode	控制除 MSI-X 以外的中断模式。
gro_enable	启用或禁用硬件通用接收卸载 (GRO) 功能。此功能与内核的软件 GRO 类似，但只能通过设备硬件执行。
err_flags_override	在发生硬件错误情况下用于禁用或强制采取措施的位图： <ul style="list-style-type: none"><li>■ 位 #31 - 此位掩码的启用位</li><li>■ 位 #0 - 阻止重新断言硬件关注</li><li>■ 位 #1 - 收集调试数据</li><li>■ 位 #2 - 触发恢复过程</li><li>■ 位 #3 - 调用 WARN 以获取导致错误的流的调用跟踪</li></ul>

## Linux 驱动程序操作默认设置

表 3-3 列出了 qed 和 qede Linux 驱动程序操作默认设置。

表 3-3. Linux 驱动程序操作默认设置

操作	qed 驱动程序默认值	qede 驱动程序默认值
Speed	自动协商并广告速度	自动协商并广告速度
MSI/MSI-X	Enabled (已启用)	Enabled (已启用)
Flow Control	—	自动协商并广告 RX 和 TX
MTU	—	1500 (范围为 46 - 9600)
Rx Ring Size	—	1000
Tx Ring Size	—	4078 (范围为 128 - 8191)
Coalesce Rx Microseconds	—	24 (范围为 0 - 255)
Coalesce Tx Microseconds	—	48
TSO	—	Enabled (已启用)

## Linux 驱动程序消息

要设置 Linux 驱动程序消息详细级别，请发出以下命令之一：

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

其中 <value> 表示 0-15 位，这些是标准的 Linux 网络值，并且 16 及更高的位特定于驱动程序。

## 统计信息

要查看详细的统计信息和配置信息，请使用 `ethtool` 公用程序。参见 `ethtool` 手册页了解更多信息。

## 安装 Windows 驱动程序软件

- [安装 Windows 驱动程序](#)
- [移除 Windows 驱动程序](#)
- [管理适配器属性](#)
- [设置电源管理选项](#)

## 安装 Windows 驱动程序

使用 Dell 更新包 (Dell Update Package, DUP) 安装 Windows 驱动程序软件：

- [在 GUI 中运行 DUP](#)
- [DUP 安装选项](#)
- [DUP 安装示例](#)

### 在 GUI 中运行 DUP

要在 GUI 中运行 DUP：

1. 双击代表 Dell 更新包文件的图标。

---

#### 注

Dell 更新包的 actual 文件名称将会不同。

---



2. 在 Dell 更新包窗口 (图 3-1) 中, 单击 **Install** (安装)。

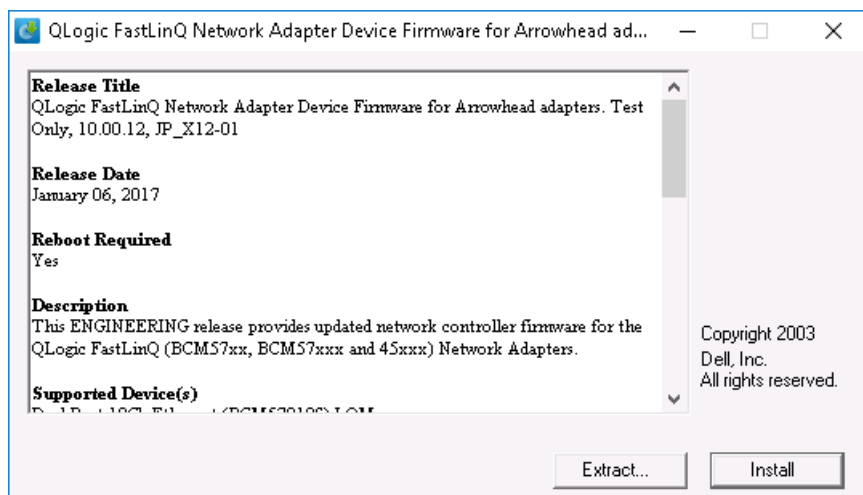


图 3-1. Dell 更新包窗口

3. 在 QLogic Super Installer—InstallShield® 向导的 Welcome (欢迎) 窗口 (图 3-2) 中, 单击 **Next** (下一步)。

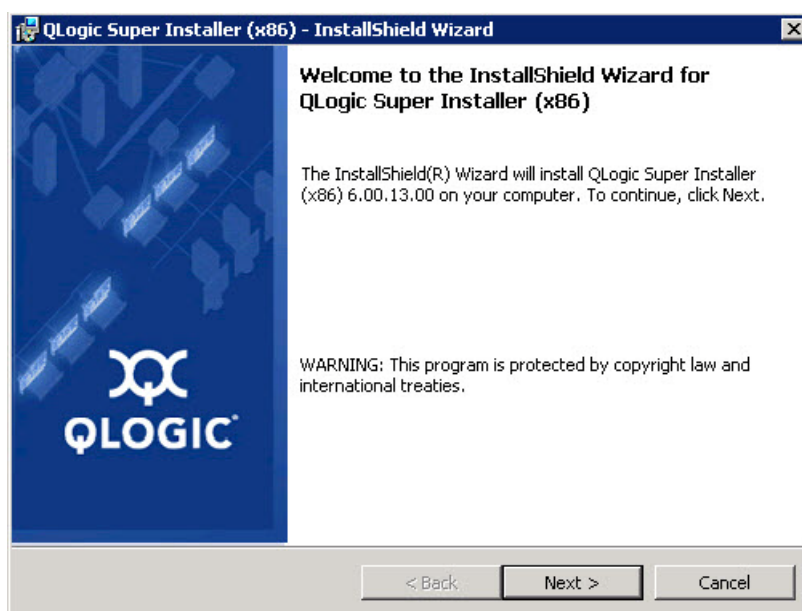


图 3-2. QLogic InstallShield 向导: Welcome (欢迎) 窗口

4. 在向导的 License Agreement（许可协议）窗口（图 3-3）中完成以下操作：
  - a. 阅读 QLogic End User Software License Agreement（QLogic 最终用户许可协议）。
  - b. 选择 **I accept the terms in the license agreement**（我接受许可协议中的条款）继续。
  - c. 单击 **Next**（下一步）。

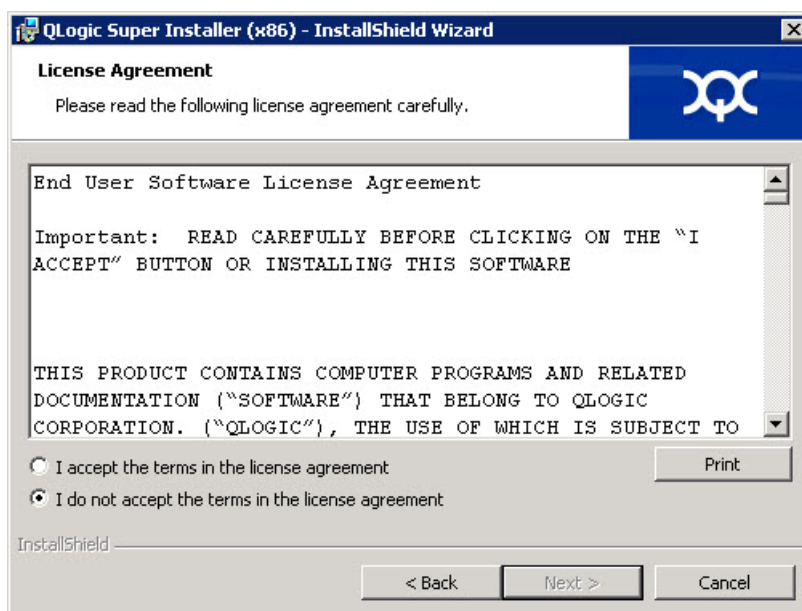
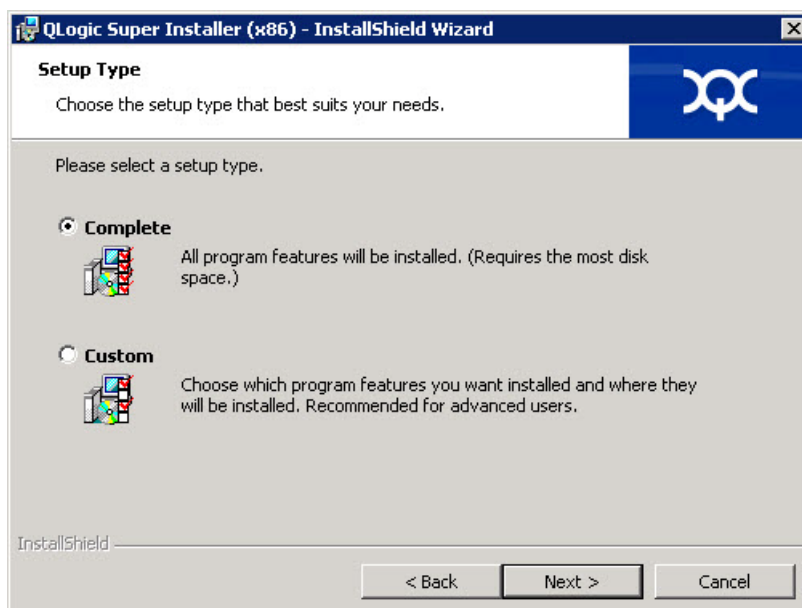


图 3-3. QLogic InstallShield 向导：License Agreement（许可协议）窗口

5. 如下完成向导的 Setup Type（设置类型）窗口（图 3-4）：
  - a. 选择以下设置类型之一：
    - 单击 **Complete**（完整）以安装所有程序功能。
    - 单击 **Custom**（自定义）以手动选择要安装的功能。
  - b. 单击 **Next**（下一步）继续。

如果单击 **Complete**（完整），则直接继续步骤 6b。



**图 3-4. InstallShield 向导：Setup Type（设置类型）窗口**

6. 如果您在步骤 5 中选择 **Custom**（自定义），则如下完成 Custom Setup（自定义设置）窗口（图 3-5）：
  - a. 选择要安装的功能。默认情况下，将选中所有功能。要更改某个功能的安装设置，请单击该功能旁边的图标，然后选择以下选项之一：
    - **This feature will be installed on the local hard drive**（此功能将安装在本地硬盘驱动器上）- 标记安装的功能，但不会影响其任何子功能。
    - **This feature, and all subfeatures, will be installed on the local hard drive**（此功能及所有子功能都将安装在本地硬盘驱动器上）- 标记安装的功能及其所有子功能。
    - **This feature will not be available**（此功能将不可用）- 阻止安装该功能。

- b. 单击 **Next**（下一步）继续。

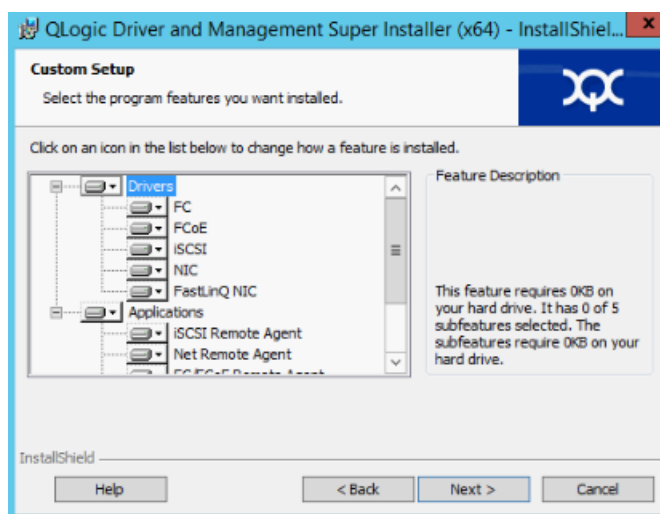


图 3-5. InstallShield 向导: Custom Setup (自定义设置) 窗口

7. 在 InstallShield 向导的 Ready To Install (准备安装) 窗口 (图 3-6) 中, 单击 **Install** (安装)。InstallShield 向导将安装 QLogic 适配器驱动程序和管理软件安装程序。

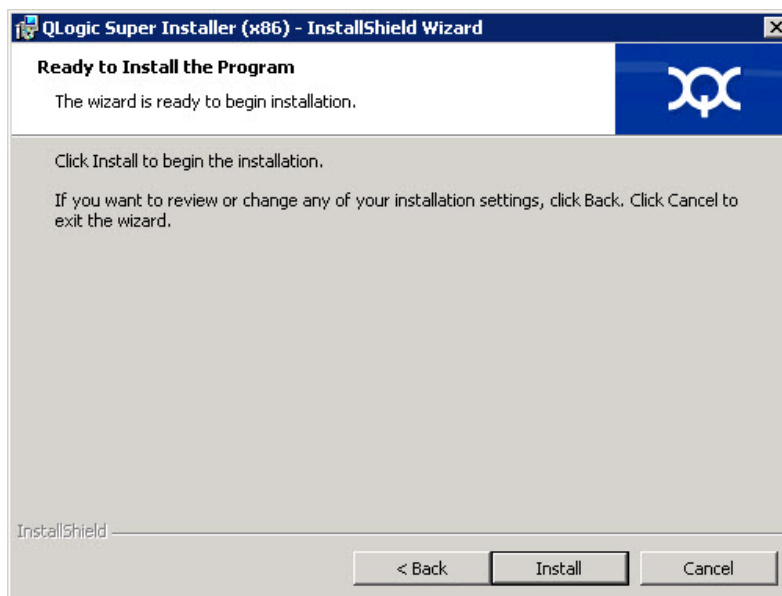
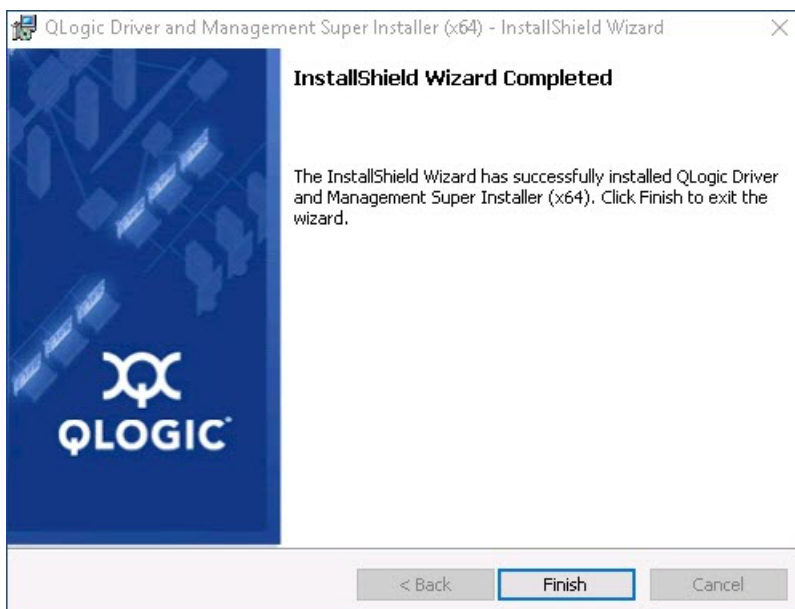


图 3-6. InstallShield 向导: Ready to Install the Program (准备安装程序) 窗口

8. 安装完成时，将显示 InstallShield Wizard Completed（InstallShield 向导完成）窗口（图 3-7）。单击 **Finish**（完成）关闭安装程序。



**图 3-7. InstallShield 向导：Completed（完成）窗口**

9. 在 Dell Update Package 窗口（图 3-8）中，“Update installer operation was successful（更新安装程序操作成功）”表示完成。
  - （可选）要打开日志文件，请单击 **View Installation Log**（查看安装日志）。日志文件显示 DUP 安装进度、任何之前安装的版本、任何错误消息，以及有关安装的其他信息。
  - 要关闭 Update Package（更新包）窗口，请单击 **CLOSE**（关闭）。

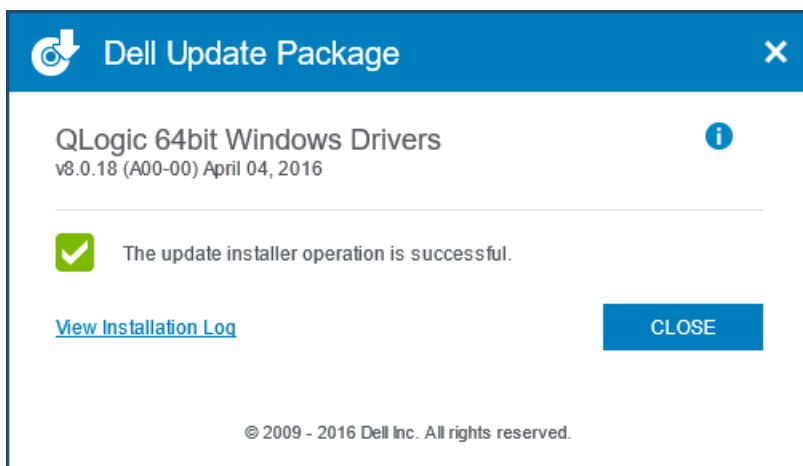


图 3-8. Dell 更新包窗口

## DUP 安装选项

要自定义 DUP 安装行为，请使用以下命令行选项。

- 要仅将驱动程序组件提取到某个目录，请使用以下选项：

`/drivers=<path>`

### 注

此命令需要 `/s` 选项。

- 要仅安装或更新驱动程序组件，请使用以下选项：

`/driveronly`

### 注

此命令需要 `/s` 选项。

- (高级) 使用 `/passthrough` 选项可将 `passthrough` 后的所有文本直接发送至 DUP 的 QLogic 安装软件。此模式禁用提供的所有 GUI，但不一定禁用 QLogic 软件的 GUI。

`/passthrough`

- (高级) 要返回此 DUP 支持的功能的代码说明，请使用以下选项：

`/capabilities`

### 注

此命令需要 `/s` 选项。

---

## DUP 安装示例

以下示例显示如何使用安装选项。

要以无提示方式更新系统，请使用以下命令：

```
<DUP_file_name>.exe /s
```

要将更新内容提取到 `C:\mydir\` 目录：

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

要将驱动程序组件提取到 `C:\mydir\` 目录：

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

要仅安装驱动程序组件，请使用以下命令：

```
<DUP_file_name>.exe /s /driveronly
```

要从默认日志位置更改为 `C:\my path with spaces\log.txt`，请使用以下命令：

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

## 移除 Windows 驱动程序

要移除 Windows 驱动程序：

1. 在控制面板中，单击 **Programs**（程序），然后单击 **Programs and Features**（程序和功能）。
2. 在程序列表中，选择 **QLogic FastLinQ Driver Installer**（QLogic FastLinQ 驱动程序安装程序），然后单击 **Uninstall**（卸载）。
3. 按照说明操作以移除驱动程序。

## 管理适配器属性

要查看或更改 **QL45212 Adapter** 属性：

1. 在控制面板中，单击 **Device Manager**（设备管理器）。
2. 在所选适配器的属性中，单击 **Advanced**（高级）选项卡。
3. 在 **Advanced**（高级）页面（[图 3-9](#)）中，选择 **Property**（属性）下的项目，然后根据需要更改该项目的 **Value**（值）。

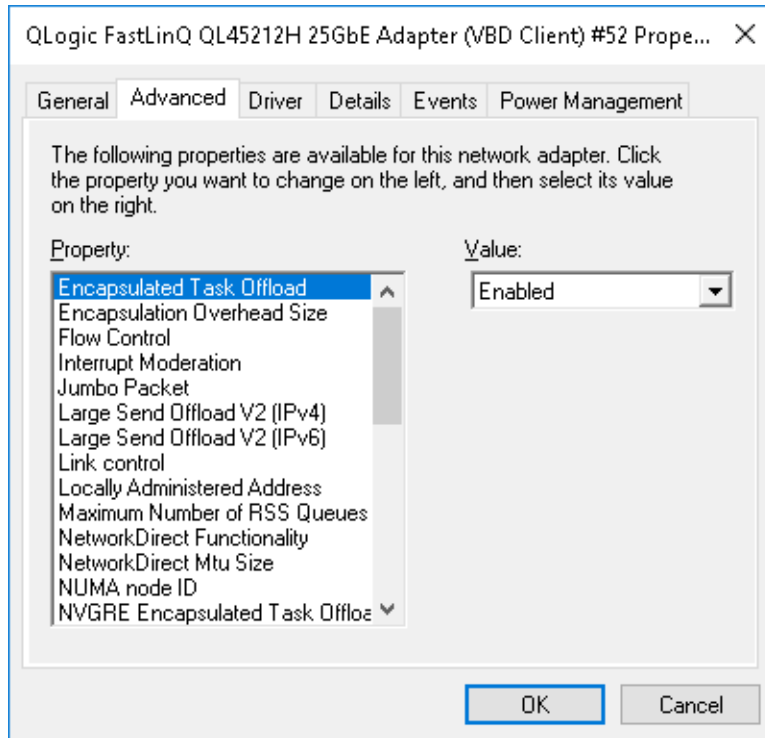


图 3-9. 设置高级适配器属性



## 设置电源管理选项

可以设置电源管理选项，以允许操作系统关闭该控制器以节约电源，或者允许该控制器唤醒计算机。如果设备正忙（例如，正在处理呼叫），操作系统将不会关闭设备。只有在计算机试图进入休眠状态时，操作系统才尝试尽可能关闭各个设备。要使控制器一直保持打开状态，不要选择 **Allow the computer to turn off the device to save power**（允许计算机关闭此设备以节约电源）复选框（图 3-10）。

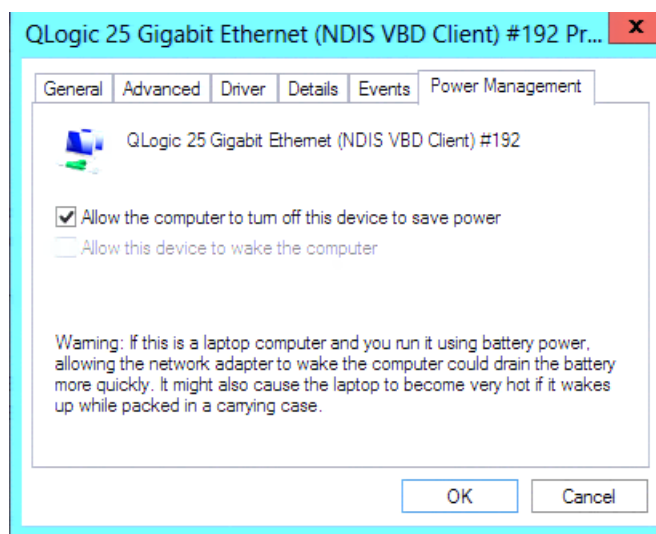


图 3-10. 电源管理选项

### 注

- 只有支持电源管理的服务器才有 Power Management（电源管理）页面。
- 对于作为组成员的任何适配器，不要选中 **Allow the computer to turn off the device to save power**（允许计算机关闭设备以节约电源）复选框。

## 安装 VMware 驱动程序软件

本节介绍用于 QL45212 Adapters 的 qedentv VMware ESXi 驱动程序：

- [VMware 驱动程序和驱动程序包](#)
- [安装 VMware 驱动程序](#)
- [VMware 驱动程序可选参数](#)
- [VMware 驱动程序参数默认值](#)
- [移除 VMware 驱动程序](#)

## VMware 驱动程序和驱动程序包

表 3-4 列出了协议的 VMware ESXi 驱动程序。

表 3-4. VMware 驱动程序

VMware 驱动程序	说明
qedentv	本机网络驱动程序
qedrntv	本机 RDMA 卸载（RoCE 和 RoCEv2）驱动程序 <sup>a</sup>
qedf	本机 FCoE 卸载驱动程序
qedil	旧版 iSCSI 卸载驱动程序

<sup>a</sup> 此版本中未包括经认证的 RoCE 驱动程序。未经认证的驱动程序可能会作为早期预览提供。

除非另有说明，否则 ESXi 驱动程序作为单独的驱动程序包被包括，而不是被捆绑在一起。表 3-5 列出了 ESXi 版本和适用的驱动程序版本。

表 3-5. 按版本的 ESXi 驱动程序包

ESXi 版本 <sup>a</sup>	协议	驱动程序名称	驱动程序版本
	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

<sup>a</sup> 发布此用户指南后，可能会提供其他 ESXi 驱动程序。有关更多信息，请参阅发行说明。

<sup>b</sup> 针对 ESXi 6.5，NIC 和 RoCE 驱动程序已被打包在一起，并且可作为单个脱机软件包使用标准 ESXi 安装命令来进行安装。程序包名为 `qedentv_3.0.7.5_qedrntv_3.0.7.5.1_signed_drivers.zip`。所建议的安装顺序为先安装 NIC 和 RoCE 驱动程序，然后再安装 FCoE 和 iSCSI 驱动程序。

使用以下任一方法安装单独的驱动程序：

- 标准 ESXi 软件包安装命令（请参阅[安装 VMware 驱动程序](#)）
- 单独驱动程序自述文件中的步骤
- 以下 VMware KB 文章中的步骤：

[https://kb.vmware.com/selfservice/microsites/search.do?language=en\\_US&cmd=displayKC&externalId=2137853](https://kb.vmware.com/selfservice/microsites/search.do?language=en_US&cmd=displayKC&externalId=2137853)

您应当首先安装 NIC 驱动程序，然后再安装存储驱动程序。

## 安装 VMware 驱动程序

可以使用驱动程序 ZIP 文件安装新驱动程序或更新现有驱动程序。请务必从同一驱动程序 ZIP 文件安装整个驱动程序集。混用来自不同 ZIP 文件的驱动程序会导致问题。

### 要安装 VMware 驱动程序：

1. 从 VMware 支持页面下载适用于 QL45212 Adapter 的 VMware 驱动程序：  
[www.vmware.com/support.html](http://www.vmware.com/support.html)
2. 启动 ESX 主机，然后登录至具有管理员权限的帐户。
3. 解压缩驱动程序 ZIP 件，然后提取 .vib 文件。
4. 使用 Linux scp 公用程序将 .vib 文件从本地系统复制到 IP 地址为 10.10.10.10 的 ESX 服务器上的 /tmp 目录中。例如，发出以下命令：

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

您可以将该文件放在 ESX 控制台 shell 可以访问的任意位置。

---

### 注

如果您没有 Linux 计算机，可使用 vSphere 数据存储文件浏览器将文件上传到服务器。

---

5. 通过发出以下命令，将主机置于维护模式：

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

6. 选择以下安装选项之一：

- **选项 1：**使用 CLI 或 VMware Update Manager (VUM) 直接在 ESX 服务器上安装 .vib。

- 要使用 CLI 安装 .vib，请发出以下命令。请务必指定完整的 .vib 文件路径。

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

- 要使用 VUM 安装 .vib，请参阅知识库文章：

[使用 VMware vCenter Update Manager 4.x 和 5.x 更新 ESXi/ESX 主机 \(1019545\)](#)

- **选项 2：**发出以下命令以一次安装所有单独的 VIB：

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

**要升级现有的驱动程序：**

按照全新安装的步骤进行操作，除了将之前选项 1 中的命令替换为以下命令：

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

## VMware 驱动程序可选参数

表 3-6 说明了可作为命令行参数提供给 esxcfg-module 命令的可选参数。

**表 3-6. VMware 驱动程序可选参数**

参数	说明
hw_vlan	全局启用 (1) 或禁用 (0) 硬件 VLAN 插入和移除。当上层需要发送或接收完整格式数据包时，禁用此参数。hw_vlan=1 为默认值。
num_queues	指定 TX/RX 队列对的数目。num_queues 可以是 1-11 或以下值之一： <ul style="list-style-type: none"><li>■ -1 允许驱动程序决定队列对的最佳数目（默认值）。</li><li>■ 0 使用默认队列。</li></ul> 对于多端口或多功能配置，您可以指定用逗号分隔的多个值。

表 3-6. VMware 驱动程序可选参数 (续)

参数	说明
multi_rx_filters	指定每个 RX 队列的 RX 过滤器数目，默认队列除外。 multi_rx_filters 可以是 1-4 或以下值之一： <ul style="list-style-type: none"> <li>■ -1 使用默认的每队列 RX 过滤器数目。</li> <li>■ 0 禁用 RX 过滤器。</li> </ul>
disable_tpa	启用 (0) 或禁用 (1) TPA (LRO) 功能。disable_tpa=0 为默认值。
max_vfs	指定每个物理功能 (PF) 的虚拟功能 (VF) 数目。max_vfs 可以是一个端口上的 0 个 (已禁用) 或 64 个 VF (已启用)。ESXi 支持最大 64 个 VF 是操作系统资源分配约束。
RSS	指定主机或 PF 的虚拟可扩展 LAN (VXLAN) 隧道流量所使用的接收端伸缩队列数目。RSS 可以是 2、3、4 或以下值之一： <ul style="list-style-type: none"> <li>■ -1 使用默认队列数目。</li> <li>■ 0 或 1 禁用 RSS 队列。</li> </ul> 对于多端口或多功能配置，您可以指定用逗号分隔的多个值。
debug	指定驱动程序在 vmkernel 日志文件中记录的数据级别。debug 可具有以下值 (按数据量升序显示)： <ul style="list-style-type: none"> <li>■ 0x80000000 表示通知级别。</li> <li>■ 0x40000000 表示信息级别 (包括通知级别)。</li> <li>■ 0x3FFFFFFF 表示针对所有驱动程序子模块的详细级别 (包括信息和通知级别)。</li> </ul>
auto_fw_reset	启用 (1) 或禁用 (0) 驱动程序自动固件恢复功能。启用此参数时，驱动程序会尝试从发送超时、固件断言和适配器奇偶校验错误等事件中恢复。默认值为 auto_fw_reset=1。
vxlan_filter_en	启用 (1) 或禁用 (0) 基于外层 MAC、内层 MAC 和 VXLAN 网络 (VNI) 的 VXLAN 过滤，将流量直接匹配至特定队列。默认值为 vxlan_filter_en=1。对于多端口或多功能配置，您可以指定用逗号分隔的多个值。
enable_vxlan_offld	启用 (1) 或禁用 (0) VXLAN 隧道流量校验和卸载和 TCP 分段卸载 (TSO) 功能。默认值为 enable_vxlan_offld=1。对于多端口或多功能配置，您可以指定用逗号分隔的多个值。

## VMware 驱动程序参数默认值

表 3-7 列出了 VMware 驱动程序参数默认值。

表 3-7. VMware 驱动程序参数默认值

参数	默认值
Speed	自动协商并广告所有速度。速度参数在所有端口上必须相同。如果在设备上启用了自动协商，则所有设备端口都将使用自动协商。
Flow Control	自动协商并广告 RX 和 TX
MTU	1,500 (范围为 46 - 9,600)
Rx Ring Size	8,192 (范围为 128 - 8,192)
Tx Ring Size	8,192 (范围为 128 - 8,192)
MSI-X	Enabled
Transmit Send Offload (TSO)	Enabled
Large Receive Offload (LRO)	Enabled
RSS	Enabled (四个 RX 队列)
HW VLAN	Enabled
Number of Queues	Enabled (八个 RX/TX 队列对)
Wake on LAN (WoL)	Disabled

## 移除 VMware 驱动程序

要移除 .vib 文件 (qedentv)，请发出以下命令：

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

要移除驱动程序，请发出以下命令：

```
# vmkload_mod -u qedentv
```

# 4 升级固件

本章提供有关使用 Dell 更新包 (Dell Update Package, DUP) 升级固件的信息。

固件 DUP 是仅用于更新闪存的公用程序；它不用于适配器配置。您可通过双击可执行文件来运行固件 DUP。或者，您可使用多个支持的命令行选项从命令行运行 DUP。

- 通过双击运行 DUP
- 第 34 页上的“从命令行运行 DUP”
- 第 35 页上的“使用 .bin 文件运行 DUP”（仅适用于 Linux）

## 通过双击运行 DUP

要通过双击可执行文件来运行固件 DUP：

1. 双击代表固件 Dell 更新包 文件的图标。

随即出现 Dell 更新包初始屏幕，如图 4-1 中所示。单击 **Install**（安装）以继续。

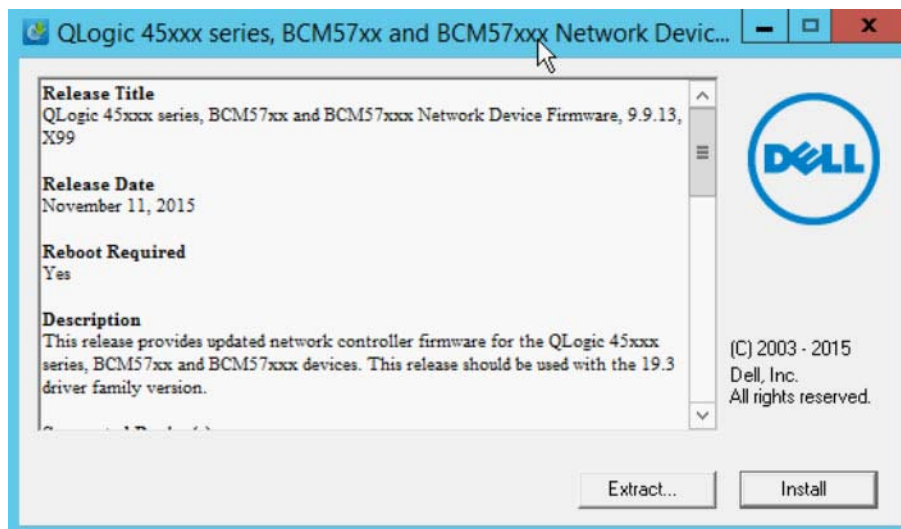


图 4-1. Dell 更新包：初始屏幕

2. 请按照屏幕说明进行操作。在 Warning（警告）对话框中，单击 **Yes**（是）继续安装，如 图 4-2 所示。

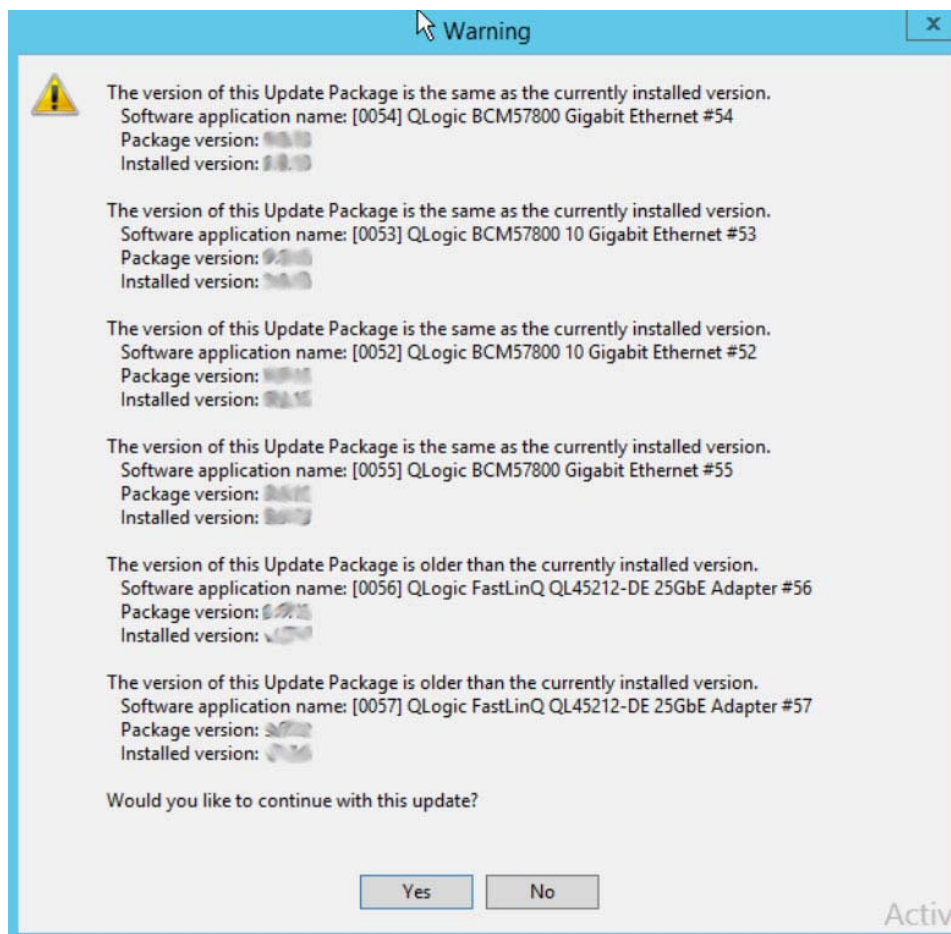
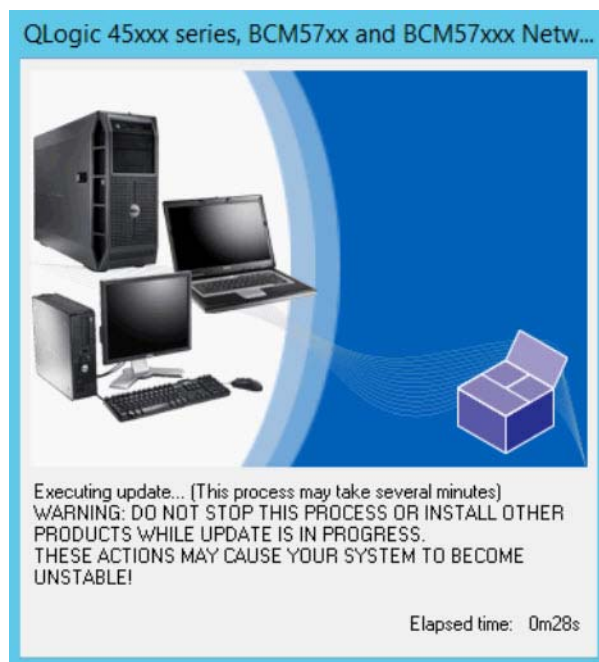


图 4-2. Dell 更新包：继续更新

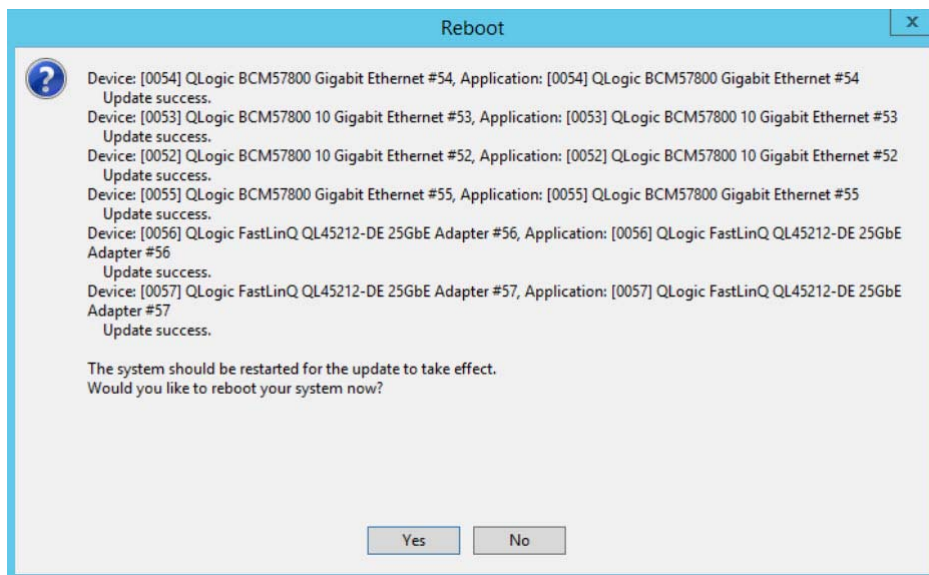


安装程序表示正在加载新固件，如 图 4-3 所示。



**图 4-3. Dell 更新包：加载新固件**

完成时，安装程序表明安装结果，如 图 4-4 所示。



**图 4-4. Dell 更新包：安装结果**

3. 单击 **Yes**（是）重新引导系统。

4. 单击 **Finish**（完成）以完成安装，如 图 4-5 所示。

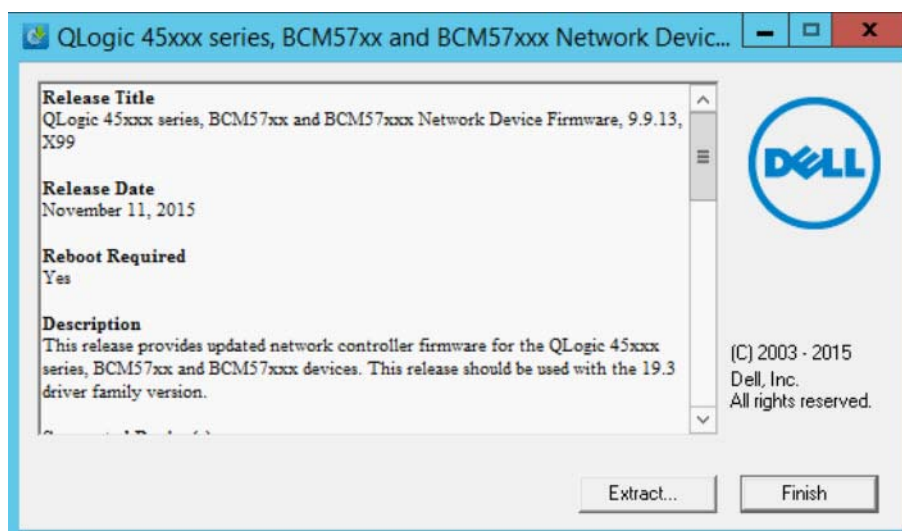


图 4-5. Dell 更新包：完成安装

## 从命令行运行 DUP

从命令行运行固件 DUP 而不指定选项时，会产生与双击 DUP 图标相同的行为。请注意，DUP 的实际文件名称各有不同。

要从命令行运行固件 DUP：

- 发出以下命令：

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

图 4-6 显示可用于自定义 Dell 更新包安装的选项。

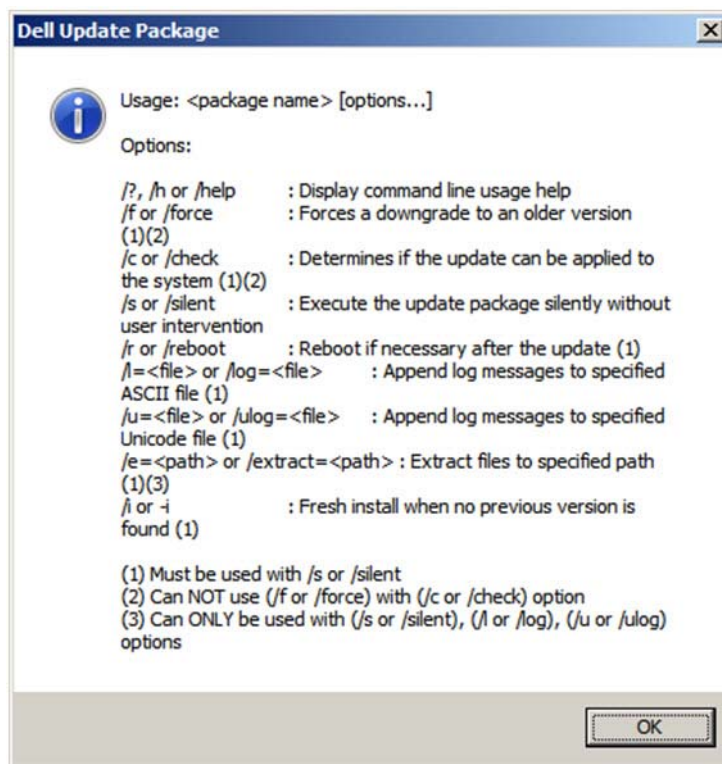


图 4-6. DUP 命令行选项

## 使用 .bin 文件运行 DUP

以下步骤仅在 Linux 操作系统上受支持。

### 要使用 .bin 文件更新 DUP:

1. 将 Network\_Firmware\_NJCX1\_LN\_X.Y.Z.BIN 文件复制到进行测试的系统 (SUT)。
2. 将文件类型更改为可执行文件，如下所示：  

```
chmod 777 Network_Firmware_NJCX1_LN_X.Y.Z.BIN
```
3. 要开始更新过程，请发出以下命令：  

```
./Network_Firmware_NJCX1_LN_X.Y.Z.BIN
```
4. 固件更新后，重新引导系统。

**DUP 更新期间 SUT 的示例输出:**

```
./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 适配器预引导配置

在主机引导过程中，您有机会暂停并使用人机界面基础设施 (HII) 应用程序执行适配器管理任务。包括以下任务：

- 第 41 页上的“显示固件映像属性”
- 第 42 页上的“配置设备级参数”
- 第 43 页上的“配置 NIC 参数”
- 第 46 页上的“配置数据中心桥接”
- 第 48 页上的“配置分区”

---

## 注

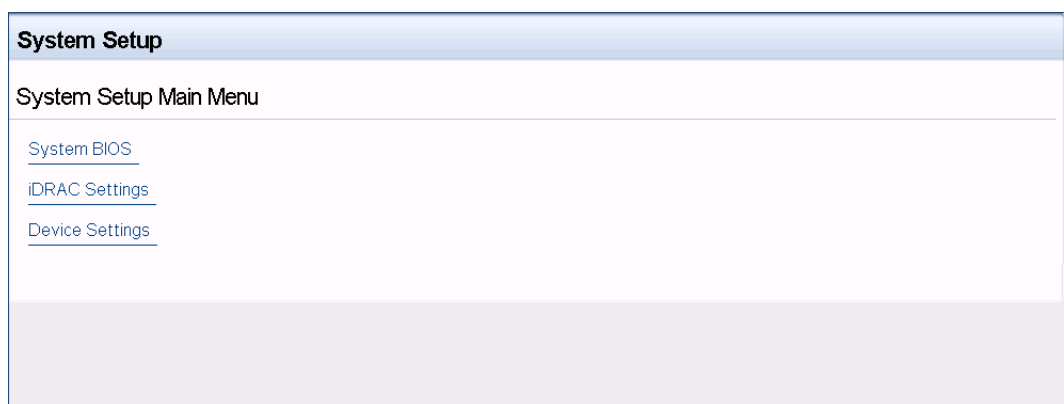
本章中的 HII 屏幕截图具有代表性，可能不符合您在系统上看到的屏幕。

---

## 使用入门

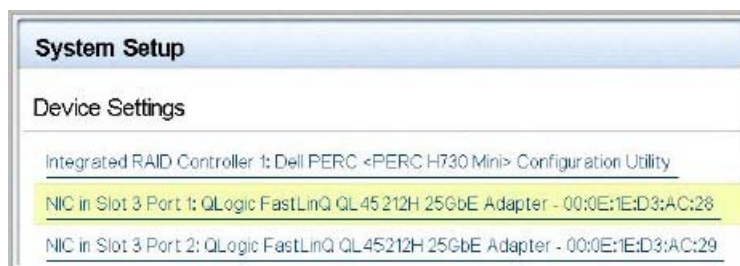
要启动 HII 应用程序：

1. 打开您的平台的 System Setup（系统设置）窗口。有关启动 System Setup（系统设置）的信息，请查询您的系统的用户指南。
2. 在 System Setup（系统设置）窗口（图 5-1）中，选择 **Device Settings**（设备设置），然后按 ENTER 键。



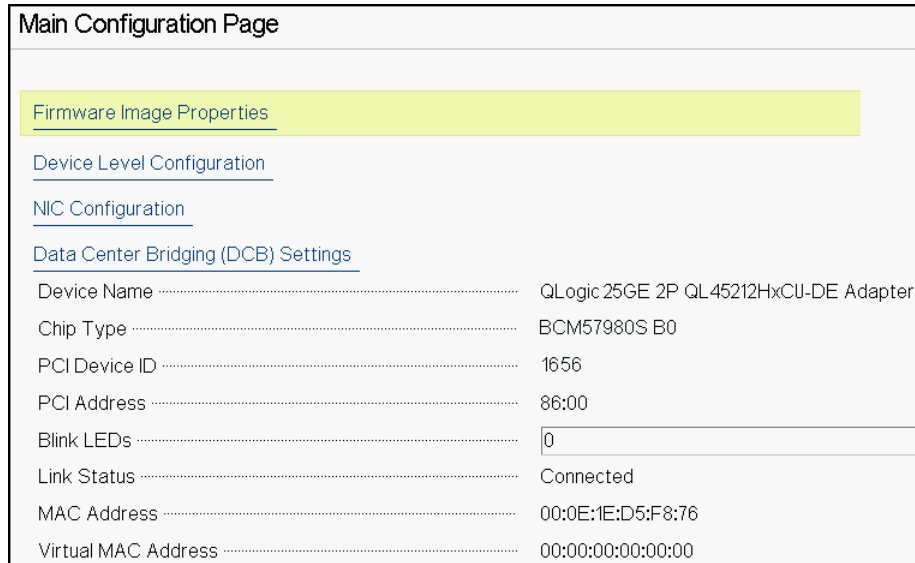
**图 5-1. 系统设置**

3. 在 Device Settings（设备设置）窗口（图 5-2）中，选择您要配置的 QL45212 Adapter 端口，然后按 ENTER 键。



**图 5-2. 系统设置：设备设置**

主配置页（图 5-3）呈现了适配器管理选项，您可通过该选项设置分区模式。



**图 5-3. 主要配置页面**

4. 在 **Device Level Configuration**（设备级配置）下，将 **Partitioning Mode**（分区模式）设置为 **NPAR**，以便将 **NIC Partitioning Configuration**（NIC 分区配置）选项添加到 **Main Configuration Page**（主要配置页面），如图 5-4 中所示。

**注**

NPAR 在最大速度为 1G 的端口上不可用。

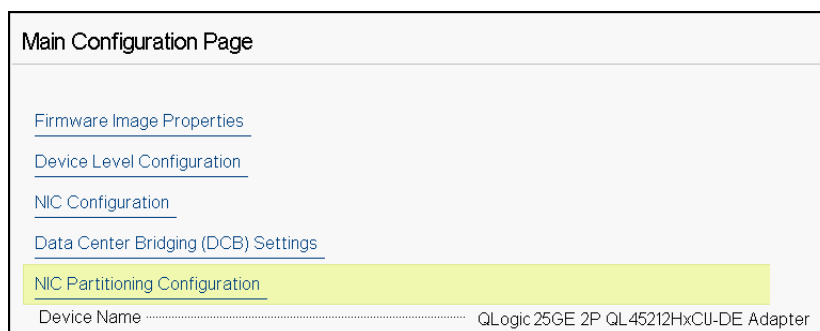


图 5-4. 主要配置页面，将分区模式设置为 NPAR

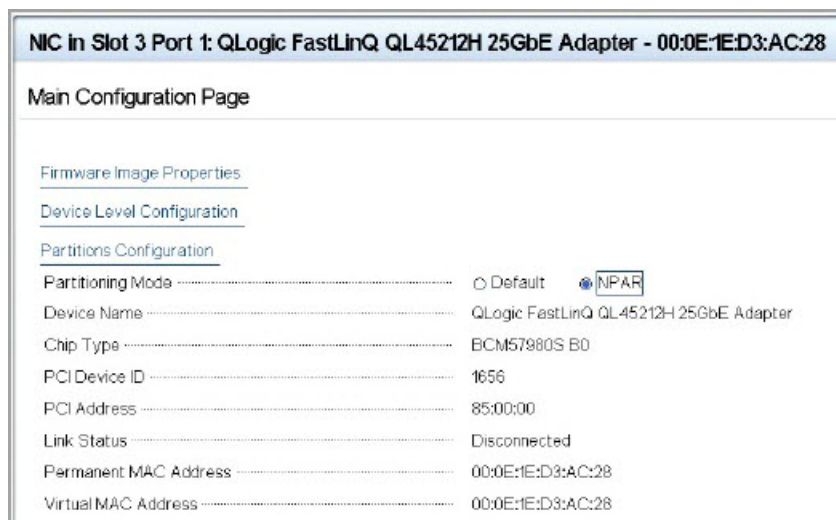


图 5-5. 主要配置页面，设置 NPAR 分区模式

在图 5-3 和图 5-4 中，Main Configuration Page（主要配置页面）显示如下：

- **Firmware Image Properties**（固件映像属性）（请参阅第 41 页上的“显示固件映像属性”）



- **Device Level Configuration**（设备级配置）（请参阅 第 42 页上的“配置设备级参数”）
- **NIC Configuration**（NIC 配置）（请参阅 第 43 页上的“配置 NIC 参数”）
- **Data Center Bridging (DCB) Settings**（数据中心桥接 (DCB) 设置）（请参阅 第 46 页上的“配置数据中心桥接”）
- **NIC Partitioning Configuration**（NIC 分区配置）（如果在 Device Level Configuration（设备级配置）页面上选择了 **NPAR**）（请参阅 第 48 页上的“配置分区”）

此外，Main Configuration Page（主要配置页面）还提供 表 5-1 中所列的适配器属性。

表 5-1. 适配器属性

适配器属性	说明
Device Name (设备名称)	工厂分配的设备名称
Chip Type (芯片类型)	ASIC 版本
PCI Device ID (PCI 设备 ID)	唯一的供应商特定 PCI 设备 ID
PCI Address (PCI 地址)	采用总线 - 设备功能格式的 PCI 设备地址
Blink LEDs (闪烁 LED)	用户定义的端口 LED 闪烁次数
Link Status (链路状态)	外部链路状态
MAC 地址	制造商分配的永久设备 MAC 地址
Virtual MAC Address (虚拟 MAC 地址)	用户定义的设备 MAC 地址

## 显示固件映像属性

要查看固件映像属性，请选择 Main Configuration Page（主要配置页面）中的 **Firmware Image Properties**（固件映像属性），然后按 ENTER 键。Firmware Image Properties（固件映像属性）页面（图 5-6）指定以下仅查看数据：

- **Family Firmware Version**（系列固件版本）是多引导映像版本，其中包含多个固件组件映像。
- **MBI Version**（MBI 版本）是设备上活动的 Cavium Qlogic 捆绑映像版本。
- **Controller BIOS Version**（控制器 BIOS 版本）是管理固件版本。

- **EFI Driver Version**（EFI 驱动程序版本）是可扩展固件接口 (EFI) 驱动程序版本。
- **L2B Firmware Version**（L2B 固件版本）是用于引导的 NIC 卸载固件版本。

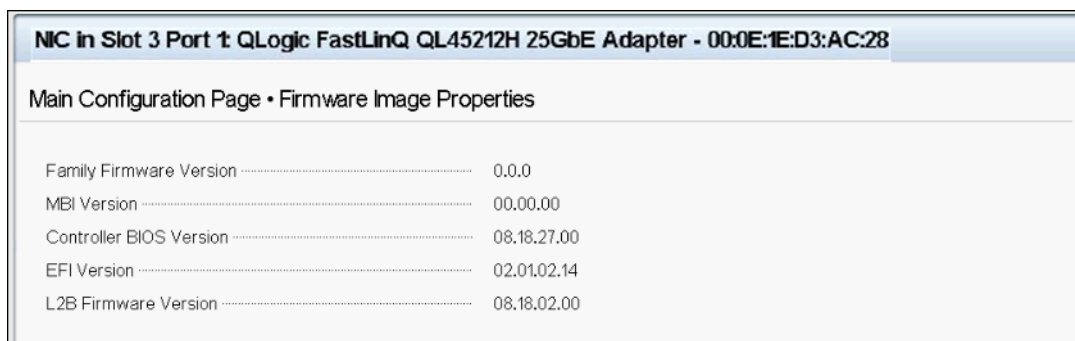


图 5-6. 固件映像属性

## 配置设备级参数

设备级别配置包括启用单域根 I/O 虚拟化 (SR-IOV) 和 NIC 分区，以及启用或禁用 NParEP。要执行设备级别配置，请在 Main Configuration Page（主要配置页面）中选择 **Device Level Configuration**（设备级别配置），并且然后单击 **Finish**（完成）。图 5-7 显示了 Device Level Configuration（设备级别配置）页面。

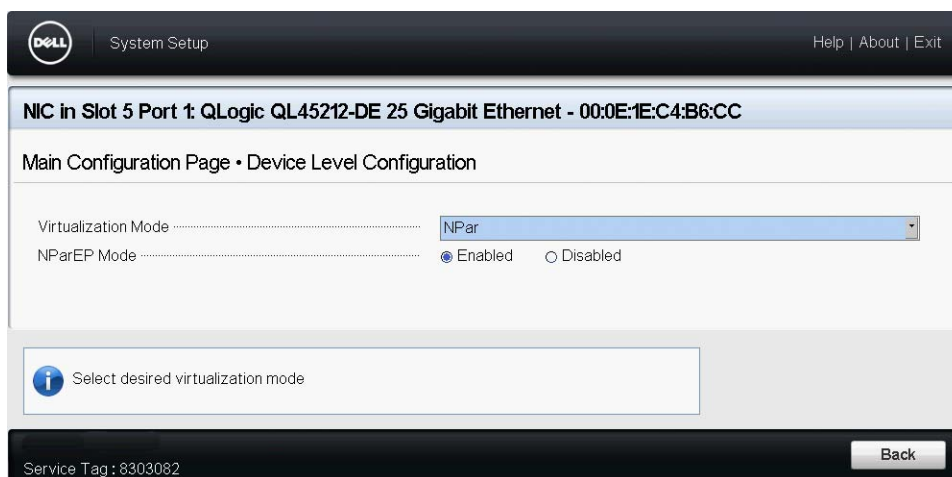


图 5-7. 系统设置：设备级配置页面

**要启用 NIC 分区 (NPAR) 和 / 或单根输入 / 输出虚拟化 (SR-IOV):**

1. 在 Main Configuration Page (主要配置页面) (第 39 页上的图 5-3) 中, 选择 **Device Level Configuration** (设备级配置), 然后单击 **Finish** (完成)。
2. 在 Device Level Configuration (设备级配置) 页面 (图 5-7) 中, 单击 **Virtualization Mode** (虚拟化模式), 然后从下列选项中选择:
  - NPAR** 启用 NPAR。
  - SRIOV** 启用 SR-IOV (非分区虚拟化)。
  - NPAR+SRIOV** 启用 NPAR 与 SR-IOV。
  - None** 禁用所有虚拟化。
3. 如果启用了 NPAR (启用或不启用 SR-IOV), 并且系统支持交替路由 ID 解析 (ARI), 则选择启用或禁用 **NPAReP mode** (NPAReP 模式):
  - 单击 **Disabled** (已禁用) 指定每个端口四个分区 (即每个适配器共 8 个 PF), 其中每个 PF 支持 16 个 SR-IOV VF。
  - 单击 **Enabled** (已启用) 指定每个端口八个分区 (即每个适配器共 16 个 PF), 其中:
    - 16 个 SR-IOV VF 位于前 8 个 PF 上 (PF 0 - 7, 每个端口的前 4 个 PF)。
    - 8 个 SR-IOV VF 位于后 8 个 PF 上 (PF 8 - 15, 每个端口的后 4 个 PF)。
    - SR-IOV Vf 当前被硬分配至各自的 PF。
4. 单击 **Back** (后退)。
5. 看到提示时, 单击 **Yes** (是) 以保存更改。更改将在系统重设后生效。

## 配置 NIC 参数

NIC 配置包括设置以下参数:

- **链路速度**
- **NIC + RDMA 模式**
- **引导模式**
- **FEC Mode (FEC 模式)**
- **PXE 虚拟 LAN 模式**
- **Virtual LAN ID (虚拟 LAN ID)**

**要配置 NIC 参数：**

1. 在 Main Configuration Page（主要配置页面）中，选择 **NIC Configuration**（NIC 配置）（第 39 页上的图 5-3），然后单击 **Finish**（完成）。

图 5-8 显示 NIC Configuration（NIC 配置）页面。

Link Speed .....	<input checked="" type="radio"/> Auto Negotiated	<input type="radio"/> 10 Gbps	<input type="radio"/> 25 Gbps
NIC + RDMA Mode .....	<input checked="" type="radio"/> Enabled	<input type="radio"/> Disabled	
Boot Mode .....	<input type="radio"/> PXE	<input checked="" type="radio"/> Disabled	
Virtual LAN Mode .....	<input type="radio"/> Enabled	<input checked="" type="radio"/> Disabled	
Virtual LAN ID .....	<input type="text" value="1"/>		

**图 5-8. NIC 配置**

2. 选择以下 **Link Speed**（链路速度）选项之一。您选择的链路速度选项适用于两个适配器端口。链路速度和前向纠错 (FEC) 必须符合所连接交换机或设备端口的相应设置。
  - Auto Negotiated**（自动协商）与所连接的交换机或设备自动协商链路参数（默认设置）。FEC 会自动启用。两端必须都处于 Auto Negotiation（自动协商）模式。
  - 10 Gbps** 指定端口固定链路速度为 10Gbps。不支持 FEC。
  - 25 Gbps** 指定端口固定链路速度为 25Gbps。
3. 对于 **NIC + RDMA Mode**（NIC + RDMA 模式），在端口上为 RDMA 选择 **Enabled**（已启用）或 **Disabled**（已禁用）。如果在 NPAR 模式下，此设置适用于端口的所有分区。
4. 在 [步骤 2](#) 中，当选择 **25 Gbps** 固定速度模式为 **Link Speed**（链路速度）时，**FEC Mode**（FEC 模式）可见。对于 **FEC Mode**（FEC 模式），选择以下选项之一。并非所有适配器都可使用 FEC 模式。
  - Disabled**（已禁用）表示未启用 FEC。
  - Fire Code**（Fire 码）表示链路被配置为使用 Fire 码 FEC；FEC 子层操作类似于第 74 条 FEC 子句。
5. 对于 **Boot Mode**（引导模式），选择以下值之一：
  - PXE** 启用 PXE 引导。
  - Disabled**（已禁用）阻止将此端口用作远程引导源。

6. 在 PXE 远程安装模式下，**Virtual LAN Mode**（虚拟 LAN 模式）参数适用于整个端口。它在 PXE 远程安装完成后不会持续。从以下 VLAN 选项中选择：
  - 对于 PXE 远程安装模式，**Enabled**（已启用）启用此端口上的 VLAN 模式。
  - Disabled**（已禁用）禁用此端口上的 VLAN 模式。
7. **Virtual LAN ID**（虚拟 LAN ID）参数指定要在此端口上用于 PXE 远程安装模式的 VLAN 标记 ID。此设置仅当在上一步中启用了 **Virtual LAN Mode**（虚拟 LAN Id）时适用。
8. 单击 **Back**（后退）。
9. 看到提示时，单击 **Yes**（是）以保存更改。更改会在系统重设后生效。

**要将端口配置为使用 RDMA：**

---

**注**

按照以下步骤在 NPAR 模式端口的所有分区上启用 RDMA。

---

1. 将 **NIC + RDMA Mode**（NIC + RDMA 模式）设置为 **Enabled**（已启用）。
2. 单击 **Back**（后退）。
3. 看到提示时，单击 **Yes**（是）以保存更改。更改会在系统重设后生效。

**要配置端口的引导模式：**

1. 对于 UEFI PXE 远程安装，请选择 **PXE** 作为 **Boot Mode**（引导模式）。
2. 单击 **Back**（后退）。
3. 看到提示时，单击 **Yes**（是）以保存更改。更改会在系统重设后生效。

**要配置端口的 PXE 远程安装以使用 VLAN：**

---

**注**

PXE 远程安装完成后，此 VLAN 不会持续。

---

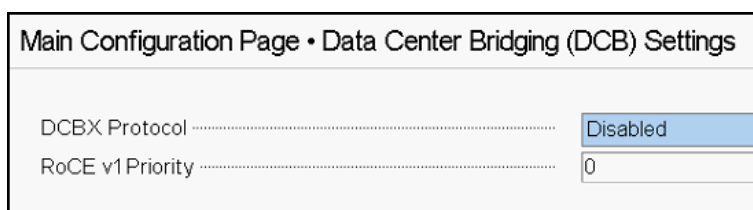
1. 将 **Virtual LAN Mode**（虚拟 LAN 模式）设置为 **Enabled**（已启用）。
2. 在 **Virtual LAN ID**（虚拟 LAN ID）框中，输入要使用的编号。
3. 单击 **Back**（后退）。
4. 看到提示时，单击 **Yes**（是）以保存更改。更改会在系统重设后生效。

## 配置数据中心桥接

数据中心桥接 (DCB) 设置包括 DCBX 协议和 RoCE 优先级。

### 要配置 DCB 设置：

1. 在 Main Configuration Page（主要配置页面）（第 39 页上的图 5-3）中，选择 **Data Center Bridging (DCB) Settings**（数据中心桥接 (DCB) 设置），然后单击 **Finish**（完成）。
2. 图 5-9 在 Data Center Bridging (DCB) Settings（数据中心桥接 (DCB) 设置）页面中，选择适当的 **DCBX Protocol**（DCBX 协议）选项：
  - Disabled**（已禁用）禁用此端口上的 DCBX。
  - CEE** 在此端口上启用旧版聚合增强型以太网 (CEE) 协议 DCBX 模式。
  - IEEE** 在此端口上启用 IEEE DCBX 协议。
  - Dynamic**（动态）启用 CEE 或 IEEE 协议的动态应用，以匹配所连接的链路伙伴。
3. 在 Data Center Bridging (DCB) Settings（数据中心桥接 (DCB) 设置）页面上，输入 **RoCE v1 Priority**（RoCE v1 优先级）的值 **0-7**。此设置指明用作 RoCE 流量的 DCB 流量类优先级编号，并且应与启用了 DCB 的交换网络用于 RoCE 流量的编号相匹配。
  - 0** 指定有损默认或通用流量类使用的正常优先级编号。
  - 3** 指定无损 FCoE 流量使用的优先级编号。
  - 4** 指定通过 DCB 的无损 iSCSI-TLV 流量使用的优先级编号。
  - 1、2、5、6 和 7** 指定可用于 RoCE 用途的 DCB 流量类优先级编号。按照相应的 OS RoCE 设置说明使用此 RoCE 控件。



Main Configuration Page • Data Center Bridging (DCB) Settings	
DCBX Protocol .....	Disabled
RoCE v1 Priority .....	0

图 5-9. 系统设置：数据中心桥接 (DCB) 设置

4. 单击 **Back**（后退）。
5. 看到提示时，单击 **Yes**（是）以保存更改。更改会在系统重设后生效。

---

### 注

在启用 DCBX 时，适配器定期发送具有作为源 MAC 地址的专用单播地址的链路层发现协议 (LLDP) 数据包。此 LLDP MAC 地址与出厂分配的适配器以太网 MAC 地址不同。如果检查连接到适配器的交换机端口的 MAC 地址表，您将看到两个 MAC 地址：一个用于 LLDP 数据包，并且一个用于适配器以太网接口。

---

## 配置分区

您可以为适配器上的每个分区配置带宽范围。

### 要配置最大和最小带宽分配：

1. 在 Main Configuration Page（主要配置页面）中，选择 **NIC Partitioning**（NIC 分区），然后按 ENTER 键。
2. 在 Partitioning Configuration（分区配置）页面（图 5-10）中，选择 **Global Bandwidth Allocation**（全局带宽分配）。

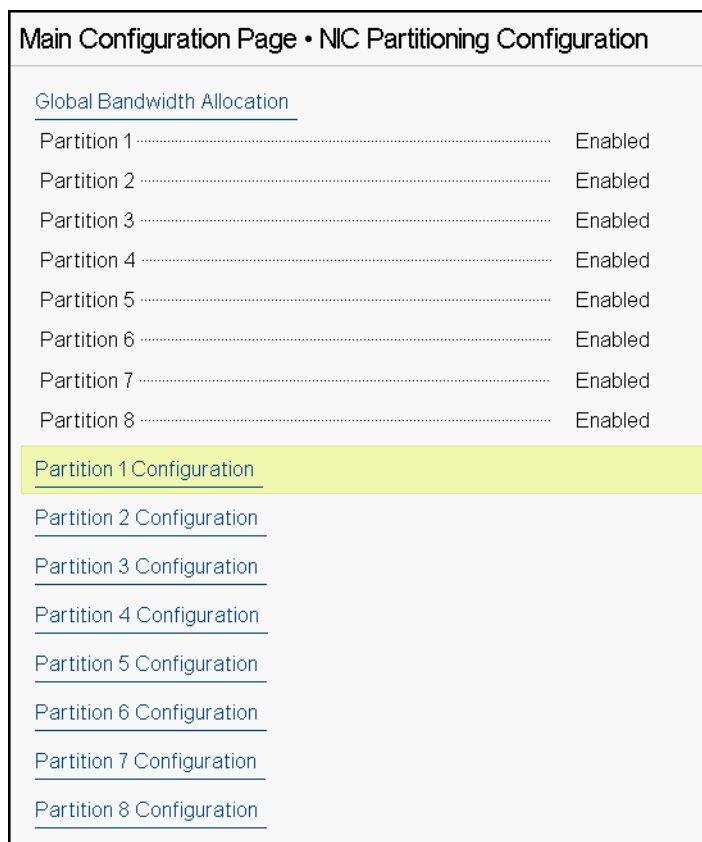


图 5-10. NIC 分区配置，全局带宽分配



图 5-11 示出了分区 1 的全球带宽分配。

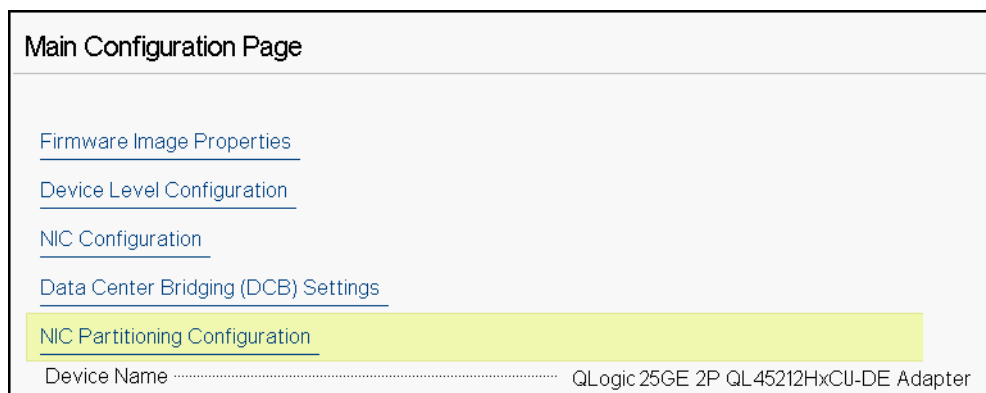


图 5-11. 全局带宽分配，分区 1 配置

3. 在 Global Bandwidth Allocation（全局带宽分配）页面（图 5-12）中，单击您要为其分配带宽的每个分区的最小和最大 TX 带宽字段。如果禁用了 NPAReP 模式，则每个端口有四个分区；如果启用了 NPAReP 模式，则每个端口有八个分区。NPAReP 模式要求在 BIOS 中为该适配器的 PCI 插槽启用 ARI。

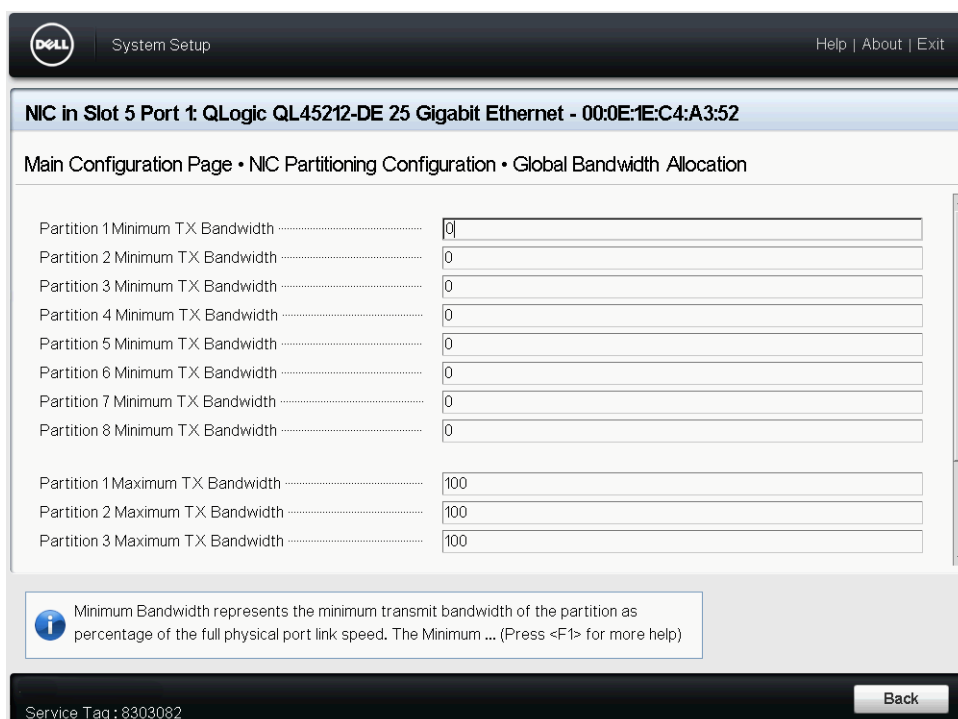


图 5-12. 全局带宽分配页面：NPAReP 模式已启用

- Partition *n* Minimum TX Bandwidth**（分区 *n* 最小 TX 带宽）是所选分区的最小发送带宽，以物理端口最大链路速度的百分比形式表示。值的有效范围是 0 - 100。如果启用了 DCBX ETS 模式，则每流量类型 DCBX ETS 最小带宽值会与每分区最小 TX 带宽值同时使用。单个端口上所有分区的最小 TX 带宽值的总和必须等于 100 或全为零。
- 将 TX 带宽全设置为零类似于在每个活动分区上等分可用带宽；然而，带宽在所有主动发送分区上动态分配。当拥塞（来自所有分区）限制 TX 带宽时，零值（将一个或多个其他值设置为非零值时）将为该分区分配百分之一的最小值。
- Partition *n* Maximum TX Bandwidth**（分区 *n* 最大 TX 带宽）是所选分区的最大发送带宽，以物理端口最大链路速度的百分比形式表示。值的有效范围是 1 - 100。无论 DCBX ETS 模式设置如何，每分区最大 TX 带宽值都适用。

在每个所选字段中键入值，然后单击 **Back**（后退）。

4. 看到提示时，单击 **Yes**（是）以保存更改。更改会在系统重设后生效。

**要配置分区：**

1. 要检查特定分区配置，在 NIC Partitioning Configuration（NIC 分区配置）页面（第 48 页上的图 5-10）中，选择 **Partition *n* Configuration**（分区 *n* 配置）。如果未启用 NParEP，则每个端口只能存在四个分区。
2. 要配置第一个分区，请选择 **Partition 1 Configuration**（分区 1 配置）以打开 Partition 1 Configuration（分区 1 配置）页面（图 5-13），其中显示以下参数：
  - NIC Mode**（NIC 模式）（始终启用）
  - PCI Device ID**（PCI 设备 ID）
  - PCI（总线） Address**（地址）
  - MAC 地址**
  - Virtual MAC Address**（虚拟 MAC 地址）

Main Configuration Page • NIC Partitioning Configuration • Partition 1 Configuration	
NIC Mode .....	Enabled
PCI Device ID .....	1656
PCI Address .....	3B:00:00
MAC Address .....	00:0E:1E:C4:CE:48
Virtual MAC Address .....	00:00:00:00:00:00

**图 5-13. 分区 1 配置**

3. 要配置第二个分区，请选择 **Partition 2 Configuration**（分区 2 配置）以打开 Partition 2 Configuration（分区 2 配置）页面。如果存在 FCoE 卸载，则 Partition 2 Configuration（分区 2 配置）（图 5-14）将显示以下参数：
  - NIC Mode**（NIC 模式）可在分区 2 及更高分区上启用或禁用 L2 以太网 NIC 个性设置。要禁用所有剩余的分区，请将 **NIC Mode**（NIC 模式）设置为 **Disabled**（已禁用）。要禁用具备卸载能力分区，请同时禁用 **NIC Mode**（NIC 模式）和相应的卸载模式。
  - PCI Device ID**（PCI 设备 ID）
  - PCI（总线） Address**（地址）
  - MAC 地址**
  - Virtual MAC Address**（虚拟 MAC 地址）

Main Configuration Page • NIC Partitioning Configuration • Partition 2 Configuration	
NIC Mode .....	<input checked="" type="radio"/> Enabled <input type="radio"/> Disabled
PCI Device ID .....	1656
PCI Address .....	3B:00:02
MAC Address .....	00:0E:1E:C4:CE:4A
Virtual MAC Address .....	00:00:00:00:00:00

**图 5-14. 分区 2 配置**

# 6 RoCE 配置

本章介绍 QL45212 Adapter、以太网交换机以及 Windows 或 Linux 主机上基于聚合以太网的 RDMA（RoCE v1 和 v2）配置，包括以下内容：

- [支持的操作系统和 OFED](#)
- [第 53 页上的“规划 RoCE”](#)
- [第 54 页上的“准备适配器”](#)
- [第 54 页上的“准备以太网交换机”](#)
- [第 56 页上的“在 Windows Server 的适配器上配置 RoCE”](#)
- [第 63 页上的“在 Linux 的适配器上配置 RoCE”](#)
- [第 73 页上的“在 VMware ESX 的适配器上配置 RoCE”](#)

## 注

某些 RoCE 功能在当前版本中可能并未完全启用。

## 支持的操作系统和 OFED

表 6-1 显示 RoCE v1、RoCE v2 和 OFED 的操作系统支持。

表 6-1. RoCE v1, RoCE v2, iWARP 和 OFED 的操作系统支持

操作系统	内建驱动程序	OFED 3.18-3 GA	OFED-4.8-1 GA
Windows Server 2012 R2	否	N/A	N/A
Windows Server 2016	否	N/A	N/A
RHEL 6.8	RoCE v1	RoCE v1	否
RHEL 6.9	RoCE v1	否	否
RHEL 7.3	RoCE v1、RoCE v2、iSER	否	RoCE v1、RoCE v2

表 6-1. RoCE v1, RoCE v2, iWARP 和 OFED 的操作系统支持 (续)

操作系统	内建驱动程序	OFED 3.18-3 GA	OFED-4.8-1 GA
RHEL 7.4	RoCE v1、RoCE v2、iSER	否	否
SLES 12 SP3	RoCE v1、RoCE v2、iSER	否	否
CentOS 7.3	RoCE v1、RoCE v2、iSER	否	RoCE v1、RoCE v2
CentOS 7.4	RoCE v1、RoCE v2、iSER	否	否
VMware ESXi 6.0 u3	否	N/A	N/A
VMware ESXi 6.5、6.5U1	RoCE v1、RoCE v2	N/A	N/A
VMware ESXi 6.7	RoCE v1、RoCE v2	N/A	N/A

## 规划 RoCE

在准备实施 RoCE 时，请考虑以下限制：

- 如果使用的是内建 OFED，则服务器和客户端系统上的操作系统应该相同。一些应用程序也许可以在不同的操作系统之间正常工作，但无法保证。这是 OFED 限制。
- 对于 OFED 应用程序（最常见的是 perftest 应用程序），服务器和客户端应用程序应使用相同的选项和值。如果操作系统和 perftest 应用程序具有不同的版本，则可能会出现问題。要确认 perftest 版本，请发出以下命令：  

```
# ib_send_bw --version
```
- 在内建 OFED 中生成 libqedr 需要安装 libibverbs-devel。
- 在内建 OFED 中运行用户空间应用程序需要安装 InfiniBand® 支持组，通过 yum groupinstall，“InfiniBand 支持”包含 libibcm、libibverbs 等等。
- 依赖于 libibverbs 的 OFED 和 RDMA 应用程序也需要 QLogic RDMA 用户空间库 libqedr。使用 libqedr RPM 或源文件包安装 libqedr。
- RoCE 仅支持小字节序。
- RoCE 无法在 SR-IOV 环境中的 VF 上工作。

## 准备适配器

按照以下步骤操作以启用 DCBX，并使用 HII 管理应用程序指定 RoCE 优先级。有关 HII 应用程序的信息，请参阅[第 5 章 适配器预引导配置](#)。

### 要准备适配器：

1. 在 Main Configuration Page（主要配置页面）中，选择 **Data Center Bridging (DCB) Settings**（数据中心桥接 (DCB) 设置），然后单击 **Finish**（完成）。
2. 在 Data Center Bridging(DCB)Settings（数据中心桥接 (DCB) 设置）窗口中，单击 **DCBX Protocol**（DCBX 协议）选项。QL45212 Adapter 支持 CEE 协议和 IEEE 协议两者。此值应匹配 DCB 交换机上的对应值。在本示例中，选择 **CEE** 或 **Dynamic**（动态）。
3. 在 **RoCE Priority**（RoCE 优先级）框中，键入优先级值。此值应该符合 DCB 交换机上的相应值。在本例中，键入 5。通常，0 用于默认有损流量类，3 用于 FCoE 流量类，4 用于通过 DCB 的无损 iSCSI-TLV 流量类。
4. 单击 **Back**（后退）。
5. 看到提示时，单击 **Yes**（是）以保存更改。更改将在系统重设后生效。

对于 Windows，您可以使用 HII 或 QoS 方法配置 DCBX。本节中所示的配置为通过 HII。对于 QoS，请参阅[第 115 页上的“为 RoCE 配置 QoS”](#)。

## 准备以太网交换机

本节介绍如何为 RoCE 配置 Cisco® Nexus® 6000 以太网交换机和 Dell® Z9100 以太网交换机。

- [配置 Cisco Nexus 6000 以太网交换机](#)
- [配置 Dell Z9100 以太网交换机](#)

### 配置 Cisco Nexus 6000 以太网交换机

用于针对 RoCE 来配置 Cisco Nexus 6000 以太网交换机的步骤包括配置类别映射、配置策略映射、应用策略，以及向交换机端口分配 VLANID。

#### 要配置 Cisco 交换机：

1. 按照以下方式来打开配置终端会话：

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

2. 按照以下方式来配置服务质量 (QoS) 类别映射并设置 RoCE 优先级以匹配适配器 (5):

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

3. 按照以下方式来配置排队类别映射:

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

4. 按照以下方式来配置网络 QoS 类别映射:

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

5. 按照以下方式来配置 QoS 策略映射:

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

6. 配置排队策略映射, 以分配网络带宽。在本示例中, 使用 50% 的值:

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

7. 按照以下方式来配置网络 QoS 策略映射, 以设置无丢弃流量类别的优先级流控制:

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

8. 按照以下方式来在系统级别应用新策略

```
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. 为交换机端口分配 VLAN ID, 以匹配分配给适配器 (5) 的 VLAN ID。

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

## 配置 Dell Z9100 以太网交换机

要为 RoCE 配置 Dell Z9100 以太网交换机，请参阅 [附录 C Dell Z9100 交换机配置](#) 中的步骤。

## 在 Windows Server 的适配器上配置 RoCE

在 Windows Server 的适配器上配置 RoCE 涉及在适配器上启用 RoCE 以及验证 Network Direct MTU 大小。

### 要在 Windows Server 主机上配置 RoCE：

1. 在适配器上启用 RoCE。
  - a. 打开 Windows 设备管理器，然后打开 QL45212 Adapters NDIS 微型端口属性。
  - b. 在 QLogic FastLinQ Adapter Properties（QLogic FastLinQ 适配器属性）中，单击 **Advanced**（高级）选项卡。
  - c. 在 Advanced（高级）页面中，通过选择 **Property**（属性）下的每个项目，然后选择该项目相应的 **Value**（值），配置表 6-2 中列出的属性。然后单击 **OK**（确定）。

**表 6-2. RoCE 的高级属性**

属性	值或说明
Network Direct Functionality (Network Direct 功能)	Enabled（已启用）
Network Direct Mtu Size (Network Direct Mtu 大小)	Network Direct MTU 大小必须小于巨型数据包大小。
RDMA Mode（RDMA 模式）	<b>RoCE v1</b> 或 <b>RoCE v2</b> 。
VLAN ID	将任何 VLAN ID 分配给接口。该值必须与交换机上分配的值相同。



图 6-1 显示配置属性值的示例。

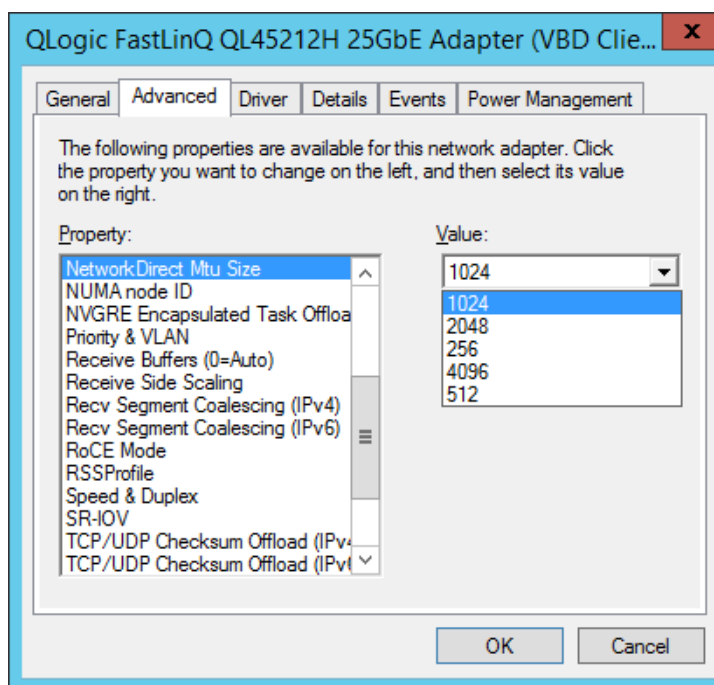


图 6-1. 配置 RoCE 属性

2. 使用 Windows PowerShell 时，请验证适配器上是否已启用 RDMA。  
Get-NetAdapterRdma 命令列出支持 RDMA 的适配器 - 两个端口同时启用。

### 注

如果您通过 Hyper-V 配置 RoCE，不要将 VLAN ID 分配给物理接口。

```
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. 使用 Windows PowerShell 时，请验证主机操作系统是否已启用 NetworkDirect。Get-NetOffloadGlobalSetting 命令显示 NetworkDirect 已启用。

```
PS C:\Users\Administrators> Get-NetOffloadGlobalSetting
ReceiveSideScaling                  : Enabled
ReceiveSegmentCoalescing           : Enabled
```

```
Chimney : Disabled
TaskOffload : Enabled
NetworkDirect : Enabled
NetworkDirectAcrossIPSubnets : Blocked
PacketCoalescingFilter : Disabled
```

4. 连接服务器消息块 (SMB) 驱动器，运行 RoCE 流量并验证结果。

要设置并连接到 SMB 驱动器，请查看 Microsoft 在线提供的信息：

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

5. 默认情况下，Microsoft 的 SMB Direct 为每个端口建立两个 RDMA 连接，以提供优质性能，包括较高块大小（例如 64KB）下的线速率。为优化性能，您可以将每个 RDMA 接口的 RDMA 连接数更改为四个（或更多）。

要将 RDMA 连接的数量增加到四个（或更多），请在 Windows PowerShell 中发出以下命令：

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

## 查看 RDMA 计数器

为了查看 RoCE 的 RDMA 计数器：

1. 启动性能监测器。
2. 打开 Add Counters（添加计数器）对话框。图 6-2 显示一个实例。

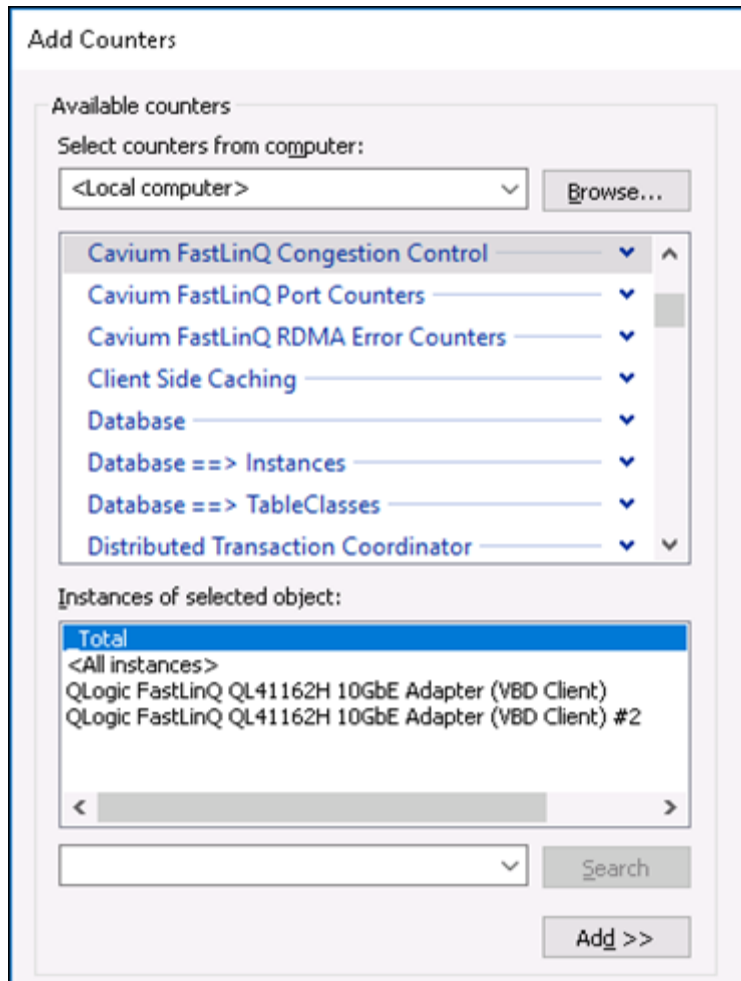


图 6-2. Add Counters（添加计数器）对话框

3. 选择以下计数器类型之一：
  - Cavium FastLinQ 拥塞控制：**
    - 当网络中出现拥塞并且交换机上启用 ECN 时的增量。
    - 描述成功发送和接收的 RoCE v2 ECN 标记数据包和拥塞通知数据包 (CNP)。
    - 仅适用于 RoCE v2。
  - Cavium FastLinQ 端口计数器：**
    - 当网络出现拥塞时的增量。
    - 当配置流量控制或全局暂停且网络出现拥塞时，暂停计数器。
    - 当配置优先级流量控制且网络出现拥塞时的 PFC 计数器增量。
  - Cavium FastLinQ RDMA 错误计数器：**
    - 传输操作出现任何错误时的增量。
    - 有关详细信息，请参阅 [表 6-3](#)。
4. 在 **Instances of selected object** 栏下，选择 **Total**，然后单击 **Add**。

图 6-3 示出了 计数器监测输出的实例。

The figure consists of three screenshots of the Windows Performance Monitor tool, each displaying a different set of performance counters for the Cavium FastLinQ RDMA adapter. All counters show a value of 0.000.

**Top Screenshot: Cavium FastLinQ Congestion Control**

Counter	_Total	QLogic FastLinQ QL41162H 10GbE Adapter (VBD Client)	QLogic FastLinQ QL41162H 10GbE Adapter (VBD Client) #2
Notification Point - CNPs Sent Successfully	0.000	0.000	0.000
Notification Point - RoCEv2 ECN Marked Packets	0.000	0.000	0.000
Reaction Point - CNPs Received Successfully	0.000	0.000	0.000

**Middle Screenshot: Cavium FastLinQ Port Counters**

Counter	_Total	QLogic FastLinQ QL41162H 10GbE Adapter (VBD Client)	QLogic FastLinQ QL41162H 10GbE Adapter (VBD Client) #2
Pause Frames Received	0.000	0.000	0.000
Pause Frames Transmitted	0.000	0.000	0.000
PFC Frames Received	0.000	0.000	0.000
PFC Frames Transmitted	0.000	0.000	0.000

**Bottom Screenshot: Cavium FastLinQ RDMA Error Counters**

Counter	_Total	QLogic FastLinQ QL41162H 10GbE Adapter (VBD Client)	QLogic FastLinQ QL41162H 10GbE Adapter (VBD Client) #2
CQ Overflow	0.000	0.000	0.000
Requestor Bad Response	0.000	0.000	0.000
Requestor CQE Flushed	0.000	0.000	0.000
Requestor Local Length	0.000	0.000	0.000
Requestor Local Protection	0.000	0.000	0.000
Requestor Local QP Operation	0.000	0.000	0.000
Requestor Remote Access	0.000	0.000	0.000
Requestor Remote Invalid Request	0.000	0.000	0.000
Requestor Remote Operation	0.000	0.000	0.000
Requestor Retry Exceeded	0.000	0.000	0.000
Requestor RNR NAK Retry Exceeded	0.000	0.000	0.000
Responder CQE Flushed	0.000	0.000	0.000
Responder Local Length	0.000	0.000	0.000
Responder Local Protection	0.000	0.000	0.000
Responder Local QP Operation	0.000	0.000	0.000
Responder Remote Invalid Request	0.000	0.000	0.000

图 6-3. 性能监测器：Cavium FastLinQ 计数器

表 6-3 提供错误计数器的有关详细信息。

表 6-3. Cavium FastLinQ RDMA 错误计数器

RDMA 错误计数器	说明	适用于 RoCE?	故障排除
CQ 溢出	发布 RDMA 工作请求的完成队列。此计数器指定在发送或接收队列上的工作请求已完成但在关联的完成队列中没有空间的示例数量。	是	表明导致完成队列空间不足的软件设计问题。
请求者响应差	响应者返回故障响应。	是	—
请求者 CQE 发生错误	当 QP 因任何原因进入错误状态并且存在待处理的工作请求时，可通过向 CQ 发送处于刷新状态的完成（而非完成实际的工作请求）来刷新发布的工作请求。如果工作请求在错误状态下完成，那么该 QP 的所有其它待处理的工作请求将刷新。	是	关闭 RDMA 连接时出现。

表 6-3. Cavium FastLinQ RDMA 错误计数器 (续)

RDMA 错误计数器	说明	适用于 RoCE?	故障排除
请求者的本地长度	包含太多或太少有效载荷数据的 RDMA READ 响应消息。	是	通常表明主机软件组件相关的问题。
请求者的本地保护	本地发布的工作请求数据段不引用对所请求操作有效的内存区域。	是	通常表明主机软件组件相关的问题。
请求者本地 QP 操作	处理此工作请求时检测到内部 QP 一致性错误。	是	—
请求者的远程访问	远程数据缓冲区出现保护错误，该错误可通过 RDMA Read 读取、通过 RDMA Write 写入或通过原子操作访问。	是	—
请求者的远程无效请求	远程端收到信道上的无效消息。无效请求可为 Send 消息或 RDMA 请求。	是	可能的原因包括：该接收队列不支持此操作；没有足够的缓冲来接收新的 RDMA 或原子操作请求；或 RDMA 请求所指定的长度大于 231 字节。
请求者的远程操作	由于其本地问题，远程端无法完成所请求的操作	是	出现在远程端的妨碍操作完成的软件问题（例如，在 RQ 上引发 QP 错误或 WQE 故障的软件问题）。
请求者重试次数超限	传输重试次数已超过最大限制	是	远程对等方可能已停止响应，或网络问题正阻止消息确认。
请求者 RNR 重试次数超限	接收到的 RNR NAK 重试次数已达上限，且无一次成功。	是	远程对等方可能已停止响应，或网络问题正阻止消息确认。
响应器 CQE 已刷新	当 QP 因任何原因进入错误状态并且存在待处理的工作请求时，可通过向 CQ 发送处于刷新状态的完成来刷新发布的工作请求（RQ 上的接收缓冲区）。如果工作请求在错误状态下完成，那么该 QP 的所有其它待处理的工作请求将刷新。	是	—

表 6-3. Cavium FastLinQ RDMA 错误计数器 (续)

RDMA 错误计数器	说明	适用于 RoCE?	故障排除
响应器的本地长度	入站消息中的无效长度。	是	恶意的远程对等机。例如，入站发送消息的长度大于接收缓冲区的大小。
响应器的本地保护	本地发布的工作请求数据段不引用对所请求操作有效的内存区域。	是	表明内存管理相关的软件问题。
响应器本地 QP 操作错误	处理此工作请求时检测到内部 QP 一致性错误。	是	表明软件问题。
响应器远程无效请求	响应器在信道上检测到无效入站消息。	是	表明来自远程对等机的可能的恶意行为。可能的原因包括：该接收队列不支持此操作；没有足够的缓冲来接收新的 RDMA 请求；或 RDMA 请求所指定的长度大于 $2^{31}$ 字节。

## 在 Linux 的适配器上配置 RoCE

本节介绍 RHEL 和 SLES 的 RoCE 配置步骤。本节还介绍如何验证 RoCE 配置，并提供有关对 VLAN 接口使用组 ID (GID) 的一些指导。

- [RHEL 的 RoCE 配置](#)
- [SLES 的 RoCE 配置](#)
- [验证 Linux 上的 RoCE 配置](#)
- [VLAN 接口和 GID 索引值](#)
- [Linux 的 RoCE v2 配置](#)

## RHEL 的 RoCE 配置

要在适配器上配置 RoCE，必须在 RHEL 主机上安装并配置开放结构企业分布 (OFED)。

### 要为 RHEL 准备内建 OFED：

1. 在安装或升级操作系统时，选择 InfiniBand® 和 OFED 支持软件包。
2. 从 RHEL ISO 映像安装以下 RPM：

```
libibverbs-devel-x.x.x.x86_64.rpm  
(libqedr 库所需)  
perftest-x.x.x.x86_64.rpm  
(InfiniBand 带宽和延迟应用程序所需)
```

或使用 Yum，安装内建 OFED：

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

---

### 注

在安装过程中，如果您已经选择了前面提到的软件包，则不需要重新安装它们。内建 OFED 和支持软件包可能因操作系统版本而异。

---

3. 如第 14 页上的“[安装包含 RDMA 的 Linux 驱动程序](#)”中所述，安装新 Linux 驱动程序。

## SLES 的 RoCE 配置

要在 SLES 主机的适配器上配置 RoCE，必须在 SLES 主机上安装并配置 OFED。

### 要为 SLES Linux 安装内建 OFED：

1. 在安装或升级操作系统时，选择 InfiniBand 支持软件包。
2. 从相应的 SLES SDK 套件映像安装以下 RPM：

```
libibverbs-devel-x.x.x.x86_64.rpm  
(libqedr 安装所需)  
perftest-x.x.x.x86_64.rpm  
(带宽和延迟应用程序所需)
```

3. 如第 14 页上的“[安装包含 RDMA 的 Linux 驱动程序](#)”中所述，安装 Linux 驱动程序。



## 验证 Linux 上的 RoCE 配置

安装 OFED 后，安装 Linux 驱动程序，然后加载 RoCE 驱动程序，验证在所有 Linux 操作系统上是否检测到 RoCE 设备。

要在 Linux 上验证 RoCE 配置：

1. 使用 `service/systemctl` 命令停止防火墙表。
2. 仅用于 RHEL：如果安装了 RDMA 服务 (`yum install rdma`)，请验证 RDMA 服务是否已启动。

---

### 注

对于 RHEL 6.x 和 SLES 11 SP4，重新引导后必须启动 RDMA 服务。  
对于 RHEL 7.x 和 SLES 12 SPX 及更高版本，RDMA 服务会在重新引导后自行启动。

---

在 RHEL 或 CentOS 上：使用 `service rdma` 状态命令以启动服务：

- ❑ 如果 RDMA 尚未启动，请发出以下命令：

```
# service rdma start
```

- ❑ 如果 RDMA 未启动，则发出以下备选命令之一：

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

或者

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

3. 通过检查 `dmesg` 日志验证是否检测到 RoCE 设备：

```
# dmesg|grep qedr
```

```
[87910.988411] qedr: discovered and registered 2 RoCE funcs
```

4. 验证是否已加载所有模块。例如：

```
# 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. 使用配置方法（如 `ifconfig`）配置 IP 地址并启用端口：

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

6. 发出 `ibv_devinfo` 命令。对于每个 PCI 功能，您应该会看到一个单独的 `hca_id`，如下例中所示：

```
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. 验证所有服务器之间的 L2 和 RoCE 连接：一个服务器充当服务器，另一个服务器充当客户端。

- 使用简单的 `ping` 命令验证 L2 连接。
- 通过在服务器或客户端上执行 RDMA ping 验证 RoCE 连接：

在服务器上发出以下命令：

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

在客户端上发出以下命令：

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

下面是服务器和客户端上的成功 ping pong 测试的示例。

### 服务器 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
```

### 客户端 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
```

■ 要显示 RoCE 统计信息，请发出以下命令，其中 *x* 是设备号：

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

## VLAN 接口和 GID 索引值

如果在服务器和客户端上均使用 VLAN 接口，您还必须在交换机上配置相同的 VLAN ID。如果是通过交换机运行流量，则 InfiniBand 应用程序必须使用正确的 GID 值，该值基于 VLAN ID 和 VLAN IP 地址。

根据以下结果，应该为任意 perftest 应用程序使用 GID 值 (-x 4/-x 5)。

```
# 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 VLAN 接口的 IP 地址
GID[ 5]: fe80:0000:0000:0000:020e:1e00:0350:c5b0 VLAN ID 3
```

### 注

背靠背或暂停设置的默认 GID 值为零 (0)。对于服务器 / 交换机配置，您必须确定合适的 GID 值。如果使用的是交换机，请参阅相应的交换机配置说明文件以了解正确设置。

---

## Linux 的 RoCE v2 配置

要验证 RoCE v2 功能，您必须使用 RoCE v2 支持的内核。

### 要为 Linux 配置 RoCE v2：

1. 确保您使用以下支持的内核之一：
  - SLES 12 SP2 GA
  - RHEL 7.3 GA
2. 按如下方式配置 RoCE v2：
  - a. 识别 RoCE v2 的 GID 索引。
  - b. 配置服务器和客户端的路由地址。
  - c. 在交换机上启用 L3 路由。

### 注

您可以通过使用 RoCE v2 支持的内核配置 RoCE v1 和 RoCE v2。这些内核允许您在同一子网以及不同子网（例如 RoCE v2）和任何可路由环境中运行 RoCE 流量。RoCE v2 只需少量设置，所有其他交换机和适配器设置均通用于 RoCE v1 和 RoCE v2。

---

## 识别 RoCE v2 GID 索引或地址

要查找 RoCE v1 和 RoCE v2 特定 GID，请使用 `sys` 或 `class` 参数，或从 QL45212 FastLinQ 源文件包运行 RoCE 脚本。要检查默认 **RoCE GID** 索引和地址，请发出 `ibv_devinfo` 命令并将其与 `sys` 或 `class` 参数进行比较。例如：

```
#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
```

## 通过 `sys` 和 `class` 参数 验证 RoCE v1 或 RoCE v2 GID 索引和地址

使用以下选项之一，通过 `sys` 和 `class` 参数验证 RoCE v1 或 RoCE v2 GID 索引和地址：

### ■ 选项 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
```

### ■ 选项 2

使用 FastLinQ 源文件包的脚本。

```
#/.../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
```

## 6-RoCE 配置

### 在 Linux 的适配器上配置 RoCE

---

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

---

#### 注

您必须基于服务器或交换机配置 (Pause/PFC) 指定 RoCE v1 或 RoCE v2 的 GID 索引值。使用链路本地 Ipv6 地址、IPv4 地址或 Ipv6 地址的 GID 索引。要将带 VLAN 标记的帧用于 RoCE 流量，您必须指定从 VLAN IPv4 或 IPv6 地址得出的 GID 索引值。

---

### 通过 perfest 应用程序验证 RoCE v1 或 RoCE v2 功能

本节展示如何通过 perfest 应用程序验证 RoCE v1 或 RoCE v2 功能。在本例中，使用以下服务器 IP 和客户端 IP：

- 服务器 IP: 192.168.100.3
- 客户端 IP: 192.168.100.4

#### 验证 RoCE v1

在同一子网上运行并使用 RoCE v1 GID 索引。

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

#### 验证 RoCE v2

在同一子网上运行并使用 RoCE v2 GID 索引。

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

#### 注

如果您通过交换机 PFC 配置运行，请对通过同一子网的 RoCE v1 或 v2 使用 VLAN GID。

---

### 通过不同子网验证 RoCE v2

#### 注

您必须首先配置交换机和服务器的路由设置。在适配器上，使用 HII 或 UEFI 用户界面设置 RoCE 优先级和 DCBX 模式。

---

**要通过不同子网验证 RoCE v2:**

1. 使用 DCBX-PFC 配置设置服务器和客户端的路由配置。

 **系统设置:**

**服务器 VLAN IP:** 192.168.100.3 和**网关:** 192.168.100.1

**客户端 VLAN IP:** 192.168.101.3 和**网关:** 192.168.101.1

 **服务器配置:**

```
#!/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
```

 **客户端配置:**

```
#!/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. 使用以下程序设置交换机设置。

- 使用任何流控制方法（Pause、DCBX-CEE 或 DCBX-IEEE），并为 RoCE v2 启用 IP 路由。请参阅 [第 54 页上的“准备以太网交换机”](#) 以了解 RoCE v2 配置，或参阅供应商交换机文档。
- 如果您使用 PFC 配置和 L3 路由，则在使用不同子网的 VLAN 上运行 RoCE v2 流量，并使用 RoCE v2 VLAN GID 索引。

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

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

**服务器交换机设置：**

```
[root@RoCE-Auto-2 /]# ib_send_bw -d qedr0 -F -x 5 -q 2 --report_gbits
*****
* Waiting for client to connect... *
*****
-----
                Send BW Test
Dual-port      : OFF           Device       : qedr0
Number of qps  : 2             Transport type : IB
Connection type : RC           Using SRQ    : OFF
RX depth       : 512
CQ Moderation  : 100
Mt看           : 1024[B]
Link type      : Ethernet
Gid index      : 5
Max inline data : 0[B]
rdma_cm QPs   : OFF
Data_ex. method : Ethernet
-----
local address: LID 0000 QPN 0xff0000 PSN 0xf0b2c3
GID: 00:00:00:00:00:00:00:00:00:00:00:00:255:255:192:168:100:03
local address: LID 0000 QPN 0xff0002 PSN 0xa2b8f1
GID: 00:00:00:00:00:00:00:00:00:00:00:00:255:255:192:168:100:03
remote address: LID 0000 QPN 0xff0000 PSN 0x40473a
GID: 00:00:00:00:00:00:00:00:00:00:00:00:255:255:192:168:101:03
remote address: LID 0000 QPN 0xff0002 PSN 0x124cd3
GID: 00:00:00:00:00:00:00:00:00:00:00:00:255:255:192:168:101:03
-----
#bytes    #iterations    BW peak[Gb/sec]    BW average[Gb/sec]    MsgRate[Mpps]
65536     1000             0.00                23.07                  0.043995
-----
```

**图 6-4. 交换机设置, 服务器**

**客户端交换机设置：**

```
[root@roce-auto-1 ~]# ib_send_bw -d qedr0 -F -x 5 192.168.100.3 -q 2 --report_gbits
-----
                Send BW Test
Dual-port      : OFF           Device       : qedr0
Number of qps  : 2             Transport type : IB
Connection type : RC           Using SRQ    : OFF
TX depth       : 128
CQ Moderation  : 100
Mt看           : 1024[B]
Link type      : Ethernet
Gid index      : 5
Max inline data : 0[B]
rdma_cm QPs   : OFF
Data_ex. method : Ethernet
-----
local address: LID 0000 QPN 0xff0000 PSN 0x40473a
GID: 00:00:00:00:00:00:00:00:00:00:00:00:255:255:192:168:101:03
local address: LID 0000 QPN 0xff0002 PSN 0x124cd3
GID: 00:00:00:00:00:00:00:00:00:00:00:00:255:255:192:168:101:03
remote address: LID 0000 QPN 0xff0000 PSN 0xf0b2c3
GID: 00:00:00:00:00:00:00:00:00:00:00:00:255:255:192:168:100:03
remote address: LID 0000 QPN 0xff0002 PSN 0xa2b8f1
GID: 00:00:00:00:00:00:00:00:00:00:00:00:255:255:192:168:100:03
-----
#bytes    #iterations    BW peak[Gb/sec]    BW average[Gb/sec]    MsgRate[Mpps]
65536     1000             23.04                23.04                  0.043936
-----
```

**图 6-5. 交换机设置, 客户端**



## 为 RDMA\_CM 应用程序配置 RoCE v1 或 RoCE v2 设置

要配置 RoCE，请使用 FastLinQ 源文件包的以下脚本：

```
# ./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
```

服务器设置：

```
[root@RoCE-Auto-2 /]# rping -s -v -C 10
server ping data: rdma-ping-0: ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-1: BCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-2: CDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-3: DEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-4: EFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-5: FGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-6: GHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-7: HIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-8: IJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server ping data: rdma-ping-9: JKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
server DISCONNECT EVENT...
wait for RDMA_READ_ADV state 10
[root@RoCE-Auto-2 /]#
```

图 6-6. 配置 RDMA\_CM 应用程序：服务器

客户端设置：

```
[root@roce-auto-1 ~]# rping -c -v -C 10 -a 192.168.100.3
ping data: rdma-ping-0: ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-1: BCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-2: CDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-3: DEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-4: EFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-5: FGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-6: GHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-7: HIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-8: IJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
ping data: rdma-ping-9: JKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz
client DISCONNECT EVENT...
[root@roce-auto-1 ~]#
```

图 6-7. 配置 RDMA\_CM 应用程序：客户端

## 在 VMware ESX 的适配器上配置 RoCE

本节提供 RoCE 配置的以下步骤和信息：

- 配置 RDMA 接口
- 配置 MTU

- [RoCE 模式和统计信息](#)
- [配置半虚拟 RDMA 设备 \(PVRDMA\)](#)

## 配置 RDMA 接口

要配置 RDMA 接口：

1. 安装 QLogic NIC 和 RoCE 驱动程序。
2. 使用模块参数，通过发出以下命令，从 NIC 驱动程序启用 RoCE 功能：

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

要应用更改，请重新加载 NIC 驱动程序或重新引导系统。

3. 要查看 NIC 接口列表，请发出 `esxcfg-nics -l` 命令。例如：

```
esxcfg-nics -l
Name      PCI          Driver      Link Speed    Duplex MAC Address      MTU      Description
Vmnict0   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
Vmnict1   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. 要查看 RDMA 设备列表，请发出 `esxcli rdma device list` 命令。例如：

```
esxcli rdma device list
Name      Driver      State      MTU      Speed      Paired Uplink      Description
-----
vmrdma0   qedrntv    Active    1024     25 Gbps    vmnict0      QLogic FastLinQ QL45xxx RDMA Interface
vmrdma1   qedrntv    Active    1024     25 Gbps    vmnict1      QLogic FastLinQ QL45xxx RDMA Interface
```

5. 要创建新的虚拟交换机，请发出以下命令：

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

例如：

```
# esxcli network vswitch standard add -v roce_vs
```

这样将创建一个名为 `roce_vs` 的新虚拟交换机。

6. 要将 QLogic NIC 端口与 vSwitch 关联，请发出以下命令：

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

例如：

```
# esxcli network vswitch standard uplink add -u vmnict0 -v roce_vs
```

7. 要在此 vSwitch 上创建新的端口组，请发出以下命令：

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

例如:

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

8. 要在此端口组上创建 vmknic 接口并配置 IP, 请发出以下命令:

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

例如:

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

9. 要配置 VLAN ID, 请发出以下命令:

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

要使用 VLAN ID 运行 RoCE 流量, 请在相应的 Vmkernel 端口组上配置 VLAN ID。

## 配置 MTU

要修改 RoCE 接口的 MTU, 请更改相应 vSwitch 的 MTU。通过发出以下命令, 基于 vSwitch 的 MTU 设置 RDMA 接口的 MTU 大小:

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

例如:

```
# 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 模式和统计信息

对于 RoCE 模式, ESXi 需要同时支持 RoCE v1 和 v2。在队列对创建过程中, 确定使用哪种 RoCE 模式。在注册和初始化期间, ESXi 驱动程序将广告两种模式。

要查看 RoCE 统计信息, 请发出以下命令:

```
# 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
```

## 配置半虚拟 RDMA 设备 (PVRDMA)

### 利用 vCenter 界面配置 PVRDMA:

1. 请按如下方式创建并配置新的分布式虚拟交换机:
  - a. 在 VMware vSphere Web Client 中, 右键单击 Navigator (导航) 窗口左侧窗格中的 **RoCE** 节点。
  - b. 在 Actions (操作) 菜单上, 指向 **Distributed Switch** (分布式交换机), 然后单击 **New Distributed Switch** (新的分布式交换机)。
  - c. 选择 6.5.0 版本。

- d. 在 **New Distributed Switch**（新的分布式交换机）下，单击 **Edit settings**（编辑设置），然后配置以下内容：
  - **Number of uplinks**（上行链路数）。选择适当的值。
  - **Network I/O Control**（网络 I/O 控制）。选择 **Disabled**（已禁用）。
  - **Default port group**（默认端口组）。选中此复选框。
  - **Port group name**（端口组名称）。输入端口组的名称。

图 6-8 显示一个实例。

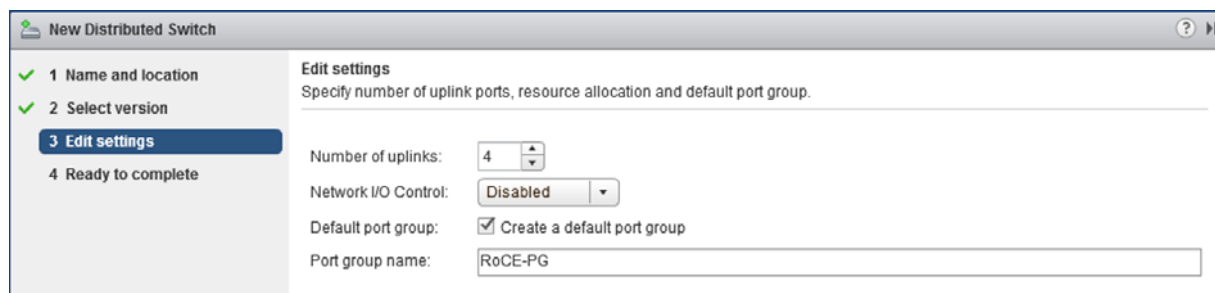


图 6-8. 配置新的分布式交换机

2. 请按如下方式配置分布式虚拟交换机：
  - a. 在 VMware vSphere Web Client 中，展开 Navigator（导航）窗口左侧窗格中的 **RoCE** 节点。
  - b. 右键单击 **RoCE-VDS**，然后单击 **Add and Manage Hosts**（添加和管理主机）。
  - c. 在 **Add and Manage Hosts**（添加和管理主机）下，配置以下内容：
    - **Assign uplinks**（分配上行链路）。从可用上行链路列表中选择。
    - **Manage VMkernel network adapters**（管理 Vmkernel 网络适配器）。接受默认设置，然后单击 **Next**（下一步）。
    - **Migrate VM networking**（迁移 VM 网络）。分配在 [步骤 1](#) 中创建的端口组。
3. 为 PVRDMA 分配一个在 ESX 主机上使用的 vmknics：
  - a. 右键单击主机，然后选择 **Settings**（设置）。
  - b. 在 Settings（设置）页面中，展开 **System**（系统）节点，然后单击 **Advanced System Settings**（高级系统设置）。
  - c. Advanced System Settings（高级系统设置）页面显示密钥对的值及其摘要。单击 **Edit**（编辑）。

- d. 在 Edit Advanced System Settings（编辑高级系统设置）页面中，在 **PVRDMA** 上筛选以将所有设置缩小为只有 Net.PVRDMAvmknics。
- e. 将 **Net.PVRDMAvmknics** 值设置为 **vmknics**；例如，**vmk1**。

图 6-9 显示一个实例。

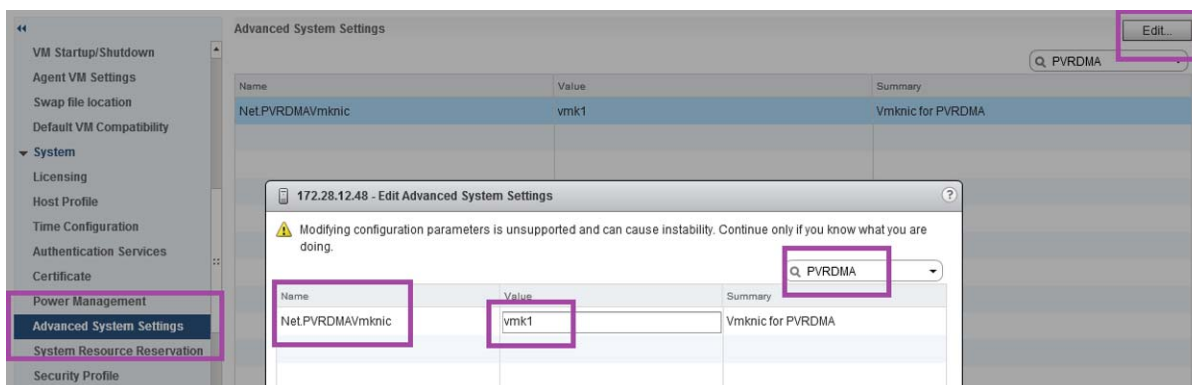


图 6-9. 为 PVRDMA 分配 vmknics

4. 设置 PVRDMA 的防火墙规则：
  - a. 右键单击主机，然后选择 **Settings**（设置）。
  - b. 在 Settings（设置）页面中，展开 **System**（系统）节点，然后单击 **Security Profile**（安全配置文件）。
  - c. 在 Firewall Summary（防火墙摘要）页面中，单击 **Edit**（编辑）。
  - d. 在 **Name**（名称）下的 Edit Security Profile（编辑安全配置文件）对话框中，向下滚动，选中 **pvrdma** 复选框，然后选中 **Set Firewall**（设置防火墙）复选框。

图 6-10 显示一个实例。

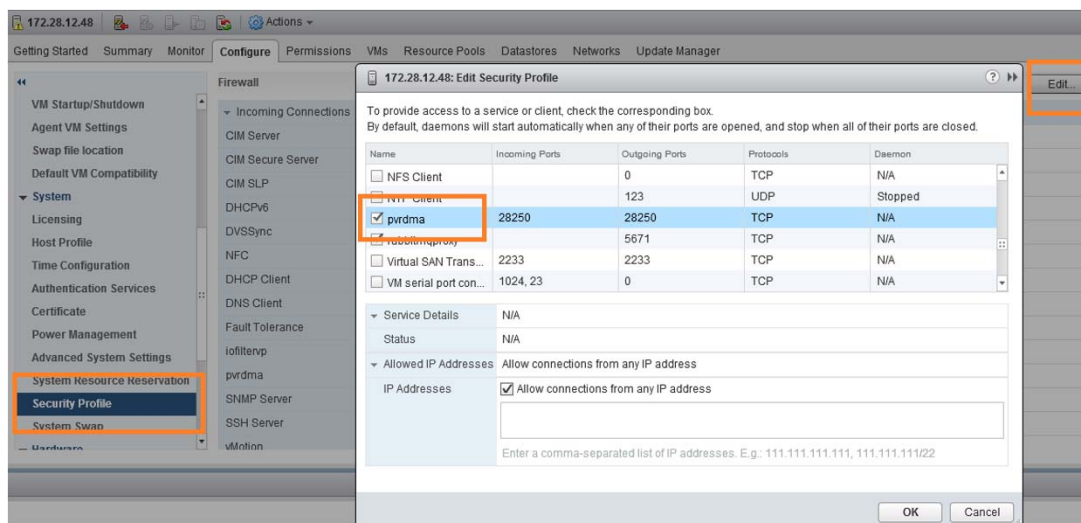


图 6-10. 设置防火墙规则

5. 请按如下方式设置 PVRDMA 的 VM:
  - a. 安装以下支持的客户操作系统之一：
    - RHEL 7.2
    - Ubuntu 14.04（内核版本 4.0）
  - b. 安装 OFED-3.18。
  - c. 编译并安装 PVRDMA 客户驱动程序和库。
  - d. 请按如下方式将新 PVRDMA 网络适配器添加至 VM：
    - 编辑 VM 设置。
    - 添加新的网络适配器。
    - 选择新添加的 DVS 端口组作为 **Network**（网络）。
    - 选择 **PVRDMA** 作为适配器类型。
  - e. 引导 VM 后，确保加载了 PVRDMA 客户驱动程序。

# 7 iSER 配置

本章提供了一种用于针对 Linux（RHEL 和 SLES）来配置 RDMA 的 iSCSI 扩展 (iSER) 的过程，该过程包括：

- [在您开始之前](#)
- [针对 RHEL 来配置 iSER](#)
- [针对 SLES 12 来配置 iSER](#)
- [优化 Linux 性能](#)
- [在 ESXi 6.7 上配置 iSER](#)

## 在您开始之前

在准备配置 iSER 时，请考虑以下事项：

- 仅在以下操作系统的内建 OFED 中支持 iSER：
  - RHEL 7.1 和 7.2
  - SLES 12 和 12 SP1
  - CentOS 7.2
- 非收件箱 OFED 不支持 iSER 目标配置，并且 `ib_isert` 模块不可用于最新的非收件箱 OFED 版本。
- 在登录目标之后或运行 I/O 流量时，卸载 Linux RoCE `qedr` 驱动程序可能导致系统崩溃。
- 在运行 I/O 时，执行接口关闭/打开测试或执行电缆拉力测试可使得驱动程序或可能导致系统崩溃的 iSER 模块错误。如果发生这种情况，请重新引导系统。



## 针对 RHEL 来配置 iSER

### 要针对 RHEL 来配置 iSER：

1. 如第 64 页上的“RHEL 的 RoCE 配置”中所述，安装收件箱 OFED。iSER 不支持非收件箱 OFED，因为 `ib_isert` 模块在非收件箱 OFED 3.18-2 GA/3.18-3 GA 版本中不可用。收件箱 `ib_isert` 模块不能与任何非收件箱 OFED 版本一起使用。
2. 如第 10 页上的“移除 Linux 驱动程序”中所述，卸载任何现有 FastLinQ 驱动程序。
3. 如第 14 页上的“安装包含 RDMA 的 Linux 驱动程序”中所述，安装最新的 FastLinQ 驱动程序和 `libqedr` 软件包。
4. 如下所示加载 RDMA 服务：

```
systemctl start rdma
modprobe qedr
modprobe ib_iser
modprobe ib_isert
```
5. 通过发出 `lsmod | grep qed` 和 `lsmod | grep iser` 命令来验证在启动器和目标设备上加载的所有 RDMA 模块和 iSER 模块。
6. 如页 66 上的步骤 6 中所示，通过发出 `ibv_devinfo` 命令来验证存在独立的 `hca_id` 实例。
7. 检查启动器设备和目标设备上的 RDMA 连接。
  - a. 在启动器设备上，发出以下命令：

```
rping -s -C 10 -v
```
  - b. 在目标设备上，发出以下命令：

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

图 7-1 显示成功 RDMA ping 操作的示例。

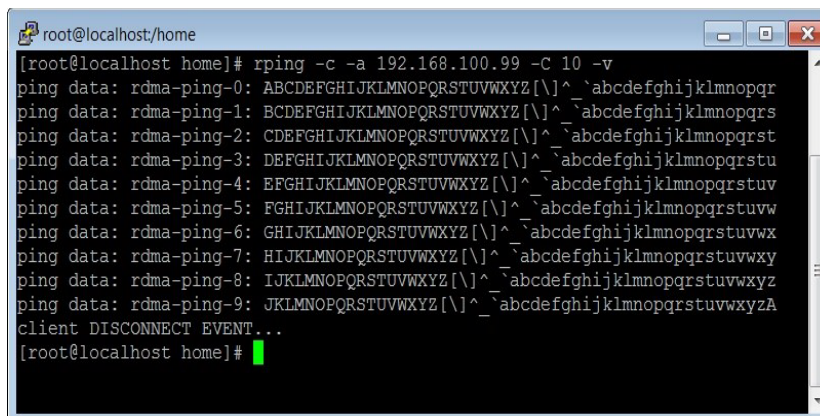


图 7-1. RDMA Ping 成功

8. 可使用 Linux TCM-LIO 目标来测试 iSER。其设置与任何 iSCSI 目标相同，但除了要在适用门户上发出命令 `enable_iser Boolean=true` 之外。门户实例在图 7-2 中被识别为 `iser`。

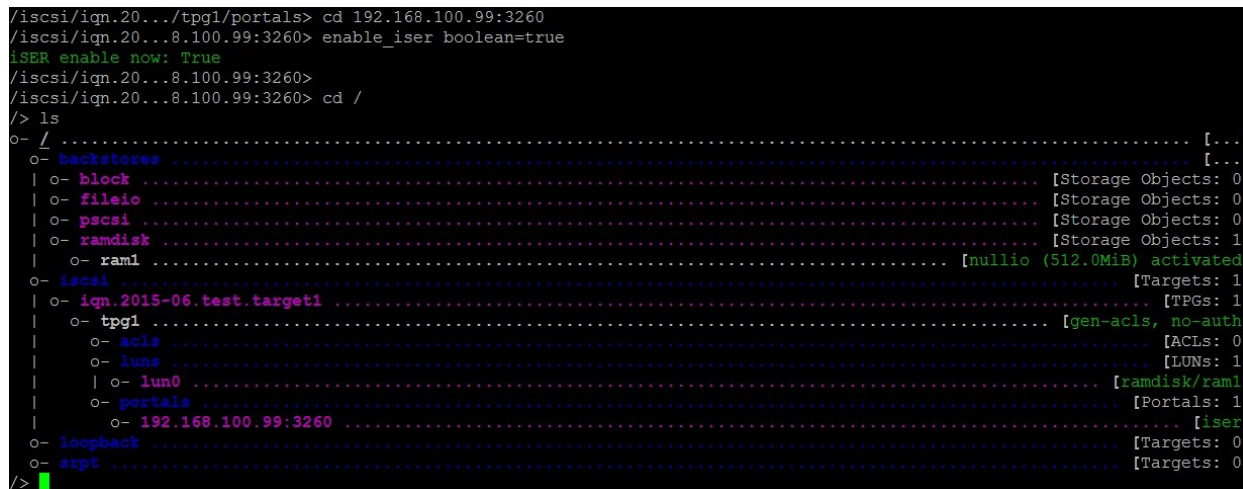


图 7-2. 利用 iSER 门户实例来测试 iSER 门户实例

9. 使用 `yum install iscsi-initiator-utils` 命令来安装 Linux iSCSI 启动器实用程序。
  - a. 要查找 iSER 目标，请发出 `iscsiadm` 命令。例如：

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

- b. 要将传输模式更改为 iSER，请发出 `iscsiadm` 命令。例如：  

```
iscsiadm -m node -T iqn.2015-06.test.target1 -o update -n iface.transport_name -v iser
```
- c. 要连接到或登录到 iSER 目标，请发出 `iscsiadm` 命令。例如：  

```
iscsiadm -m node -l -p 192.168.100.99:3260 -T iqn.2015-06.test.target1
```
- d. 确认 `Iface Transport` 为目标连接中的 `iser`，如图 7-3 中所示的。发出 `iscsiadm` 命令；例如：  

```
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 ~]# 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
```

图 7-3. Iface 传输确认

- e. 要检查新的 iSCSI 设备，如图 7-4 中所示，发出 `lsscsi` 命令。

```
[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 ~]#
```

图 7-4. 检查新的 iSCSI 设备

## 针对 SLES 12 来配置 iSER

由于 targetcli 不是 SLES 12.x 上的收件箱，因此您必须完成以下步骤。

### 要针对 SLES 12 来配置 iSER：

1. 要安装 targetcli，从 ISO 映像（x86\_64 和 noarch 位置）复制并安装以下 RPM：

```
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. 在启动 targetcli 之前，请如下来加载所有 RoCE 设备驱动程序和 iSER 模块：

```
# modprobe qed
# modprobe qede
# modprobe qedr
# modprobe ib_iser （启动器）
# modprobe ib_isert （目标）
```

3. 在配置 iSER 目标之前，配置 NIC 接口并运行 L2 和 RoCE 流量，如[页 66](#)上的[步骤 7](#)中所述的。
4. 启动 targetcli 实用程序，并且在 iSER 目标系统上配置您的目标。

---

### 注

targetcli 版本在 RHEL 和 SLES 中为不同的。请务必使用合适的后备存储来配置您的目标：

- RHEL 使用 *ramdisk*
  - SLES 使用 *rd\_mcp*
-

## 优化 Linux 性能

考虑在本节中描述的以下 Linux 性能配置增强。

- 将 CPU 配置为最高性能模式
- 配置内核 sysctl 设置
- 配置 IRQ 关联设置
- 配置块设备暂存

### 将 CPU 配置为最高性能模式

通过使用以下脚本来将 CPU 缩放调节器配置成性能，以将所有 CPU 设置为最高性能模式：

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

通过发出以下命令来验证所有 CPU 核心是否被设置为最高性能模式：

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

### 配置内核 sysctl 设置

按照以下方式来设置内核 sysctl：

```
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
```

## 配置 IRQ 关联设置

以下示例将 CPU 核心 0、1、2 和 3 分别设置为中断请求 (IRQ) XX、YY、ZZ 和 XYZ。对分配给端口的各 IRQ 执行以下步骤（默认为每端口八个队列）。

```
systemctl disable irqbalance
systemctl stop irqbalance
cat /proc/interrupts | grep qedr 显示被分配至每个端口队列的 IRQ
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
```

## 配置块设备暂存

按照以下方式来设置每个 iSCSI 设备或目标的块设备暂存设置：

```
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
```

## 在 ESXi 6.7 上配置 iSER

本节介绍为 VMware ESXi 6.7 配置 iSER 的有关信息。

### 准备工作

为 ESXi 6.7 配置 iSER 前，请确保完成以下操作：

- 在 ESXi 6.7 系统上安装带 NIC 和 RoCE 驱动器的 CNA 程序包并列设备。为了查看 RDMA 设备，请发出以下命令：

```
esxcli rdma device list
Name      Driver  State  MTU  Speed  Paired Uplink  Description
-----
vmrdma0  qedrntv Active 1024  40 Gbps  vmnic4          QLogic FastLinQ QL45xxx RDMA Interface
vmrdma1  qedrntv Active 1024  40 Gbps  vmnic5          QLogic FastLinQ QL45xxx RDMA Interface
[root@localhost:~] esxcfg-vmknics -l
Interface  Port Group/DVPort/Opaque Network  IP Family IP Address
Netmask    Broadcast  MAC Address  MTU  TSO MSS  Enabled Type
NetStack
vmk0       Management Network  IPv4        172.28.12.94
255.255.240.0  172.28.15.255  e0:db:55:0c:5f:94 1500  65535  true  DHCP
defaultTcpipStack
```

## 7-iSER 配置

### 在 ESXi 6.7 上配置 iSER

```
vmk0      Management Network      IPv6      fe80::e2db:55ff:fe0c:5f94
64                e0:db:55:0c:5f:94 1500      65535      true      STATIC, PREFERRED
defaultTcpipStack
```

- 配置 iSER 目标，从而实现与 iSER 启动器的通信。

## 为 ESXi 6.7 配置 iSER

### 为 ESXi 6.7 配置 iSER 的步骤：

1. 通过发出以下命令添加 iSER 设备：

```
esxcli rdma iser add
esxcli iscsi adapter list
Adapter Driver State UID Description
-----
vmhba64 iser unbound iscsi.vmhba64 VMware iSCSI over RDMA (iSER) Adapter
vmhba65 iser unbound iscsi.vmhba65 VMware iSCSI over RDMA (iSER) Adapter
```

2. 如下所示禁用防火墙。

```
esxcli network firewall set --enabled=false
esxcli network firewall unload
vsish -e set /system/modules/iscsi_trans/loglevels/iscsitrans 0
vsish -e set /system/modules/iser/loglevels/debug 4
```

3. 创建标准 vSwitch Vmkernel 端口组并分配 IP：

```
esxcli network vswitch standard add -v vSwitch_iser1
esxcfg-nics -l
Name PCI Driver Link Speed Duplex MAC Address MTU Description
vmnic0 0000:01:00.0 ntg3 Up 1000Mbps Full e0:db:55:0c:5f:94 1500 Broadcom Corporation NetXtreme BCM5720 Gigabit Ethernet
vmnic1 0000:01:00.1 ntg3 Down 0Mbps Half e0:db:55:0c:5f:95 1500 Broadcom Corporation NetXtreme BCM5720 Gigabit Ethernet
vmnic2 0000:02:00.0 ntg3 Down 0Mbps Half e0:db:55:0c:5f:96 1500 Broadcom Corporation NetXtreme BCM5720 Gigabit Ethernet
vmnic3 0000:02:00.1 ntg3 Down 0Mbps Half e0:db:55:0c:5f:97 1500 Broadcom Corporation NetXtreme BCM5720 Gigabit Ethernet
vmnic4 0000:42:00.0 qedentv Up 40000Mbps Full 00:0e:1e:d5:f6:a2 1500 QLogic Corp. QLogic FastLinQ QL45xxx 10/25/40/50/100 GbE Ethernet Adapter
vmnic5 0000:42:00.1 qedentv Up 40000Mbps Full 00:0e:1e:d5:f6:a3 1500 QLogic Corp. QLogic FastLinQ QL45xxx 10/25/40/50/100 GbE Ethernet Adapter

esxcli network vswitch standard uplink add -u vmnic5 -v vSwitch_iser1
esxcli network vswitch standard portgroup add -p "rdma_group1" -v vSwitch_iser1
esxcli network ip interface add -i vmk1 -p "rdma_group1"
esxcli network ip interface ipv4 set -i vmk1 -I 192.168.10.100 -N 255.255.255.0 -t static
esxcfg-vswitch -p "rdma_group1" -v 4095 vSwitch_iser1
```

## 7-iSER 配置

### 在 ESXi 6.7 上配置 iSER

---

```
esxcli iscsi networkportal add -n vmk1 -A vmhba65
esxcli iscsi networkportal list
esxcli iscsi adapter get -A vmhba65
vmhba65
  Name: iqn.1998-01.com.vmware:localhost.punelab.qlogic.com qlogic.org qlogic.com
mv.qlogic.com:1846573170:65
  Alias: iser-vmnic5
  Vendor: VMware
  Model: VMware iSCSI over RDMA (iSER) Adapter
  Description: VMware iSCSI over RDMA (iSER) Adapter
  Serial Number: vmnic5
  Hardware Version:
  Asic Version:
  Firmware Version:
  Option Rom Version:
  Driver Name: iser-vmnic5
  Driver Version:
  TCP Protocol Supported: false
  Bidirectional Transfers Supported: false
  Maximum Cdb Length: 64
  Can Be NIC: true
  Is NIC: true
  Is Initiator: true
  Is Target: false
  Using TCP Offload Engine: true
  Using iSCSI Offload Engine: true
```

#### 4. 如下所示将目标添加至 iSER 启动器:

```
esxcli iscsi adapter target list
esxcli iscsi adapter discovery sendtarget add -A vmhba65 -a 192.168.10.11
esxcli iscsi adapter target list
Adapter Target Alias Discovery Method Last Error
-----
vmhba65 iqn.2015-06.test.target1 SENDTARGETS No Error
esxcli storage core adapter rescan --adapter vmhba65
```

#### 5. 如下所示列出所附接的目标:

```
esxcfg-scsidevs -l
mpx.vmhba0:C0:T4:L0
  Device Type: CD-ROM
  Size: 0 MB
  Display Name: Local TSSTcorp CD-ROM (mpx.vmhba0:C0:T4:L0)
  Multipath Plugin: NMP
```



## 7-iSER 配置

### 在 ESXi 6.7 上配置 iSER

---

```
Console Device: /vmfs/devices/cdrom/mpx.vmhba0:C0:T4:L0
Devfs Path: /vmfs/devices/cdrom/mpx.vmhba0:C0:T4:L0
Vendor: TSSTcorp Model: DVD-ROM SN-108BB Revis: D150
SCSI Level: 5 Is Pseudo: false Status: on
Is RDM Capable: false Is Removable: true
Is Local: true Is SSD: false
Other Names:
    vml.0005000000766d686261303a343a30
VAAI Status: unsupported
naa.6001405e81ae36b771c418b89c85dae0
Device Type: Direct-Access
Size: 512 MB
Display Name: LIO-ORG iSCSI Disk (naa.6001405e81ae36b771c418b89c85dae0)
Multipath Plugin: NMP
Console Device: /vmfs/devices/disks/naa.6001405e81ae36b771c418b89c85dae0
Devfs Path: /vmfs/devices/disks/naa.6001405e81ae36b771c418b89c85dae0
Vendor: LIO-ORG Model: raml Revis: 4.0
SCSI Level: 5 Is Pseudo: false Status: degraded
Is RDM Capable: true Is Removable: false
Is Local: false Is SSD: false
Other Names:
    vml.02000000006001405e81ae36b771c418b89c85dae072616d312020
VAAI Status: supported
naa.690b11c0159d050018255e2d1d59b612
```

# 8 SR-IOV 配置

单根输入 / 输出虚拟化 (SR-IOV) 为一种 PCISIG 的规格，使得单个 PCI Express (PCIe) 设备能够被显示为多个单独的物理 PCIe 设备。SR-IOV 允许隔离 PCIe 资源，以实现性能、互操作性和可管理性。

---

## 注

某些 SR-IOV 功能在当前版本中可能并未完全启用。

---

本章提供以下内容的说明：

- [在 Windows 上配置 SR-IOV](#)
- [第 97 页上的“在 Linux 上配置 SR-IOV”](#)
- [第 103 页上的“在 VMware 上配置 SR-IOV”](#)

## 在 Windows 上配置 SR-IOV

要在 Windows 上配置 SR-IOV：

1. 访问服务器的 BIOS System Setup（BIOS 系统设置），并且然后单击 **System BIOS Settings**（系统 BIOS 设置）。
2. 在 System BIOS Settings（系统 BIOS 设置）页面中，单击 **Integrated Devices**（集成设备）。

3. 在 Integrated Devices（集成设备）页面（图 8-1）中：
  - a. 将 **SR-IOV Global Enable**（SR-IOV 全局启用）选项设置为 **Enabled**（已启用）。
  - b. 单击 **Back**（后退）。

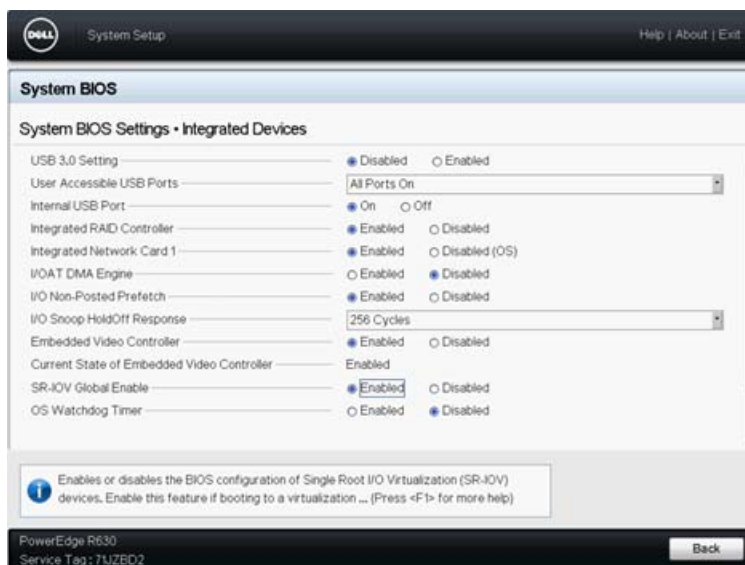


图 8-1. SR-IOV 的系统设置：集成设备

4. 在所选适配器的 Main Configuration Page（主要配置页面）上，单击 **Device Level Configuration**（设备级配置）。
5. 在 Main Configuration Page - Device Level Configuration（主要配置页面 - 设备级配置）（图 8-2）中：
  - a. 将 Virtualization Mode（虚拟化模式）设置为 **SR-IOV**。
  - b. 单击 Back（后退）。

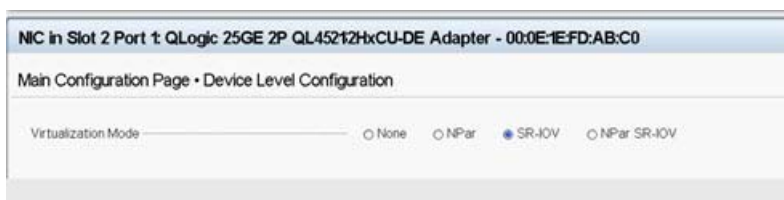


图 8-2. SR-IOV 的系统设置：设备级配置

6. 在 Main Configuration Page（主要配置页面）中，单击 **Finish**（完成）。
7. 在 Warning - Saving Changes（警告 - 保存更改）消息框中，单击 **Yes**（是）保存配置。

8. 在 Success - Saving Changes（成功 - 保存更改）消息框中，单击 **OK**（确定）。
9. 要在微型端口适配器上启用 SR-IOV：
  - a. 访问设备管理器。
  - b. 打开微型端口适配器属性，然后单击 **Advanced**（高级）选项卡。
  - c. 在 Advanced（高级）属性页面（图 8-3）的 **Property**（属性）下，选择 **SR-IOV**，然后将值设置为 **Enabled**（已启用）。
  - d. 单击 **OK**（确定）。

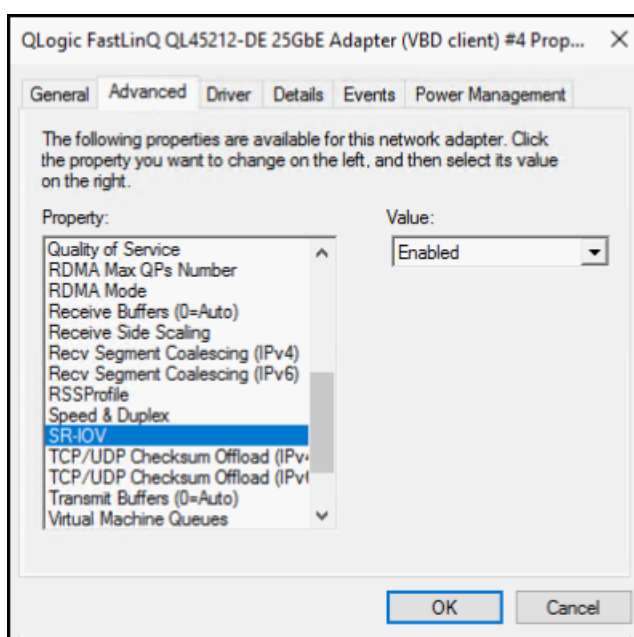
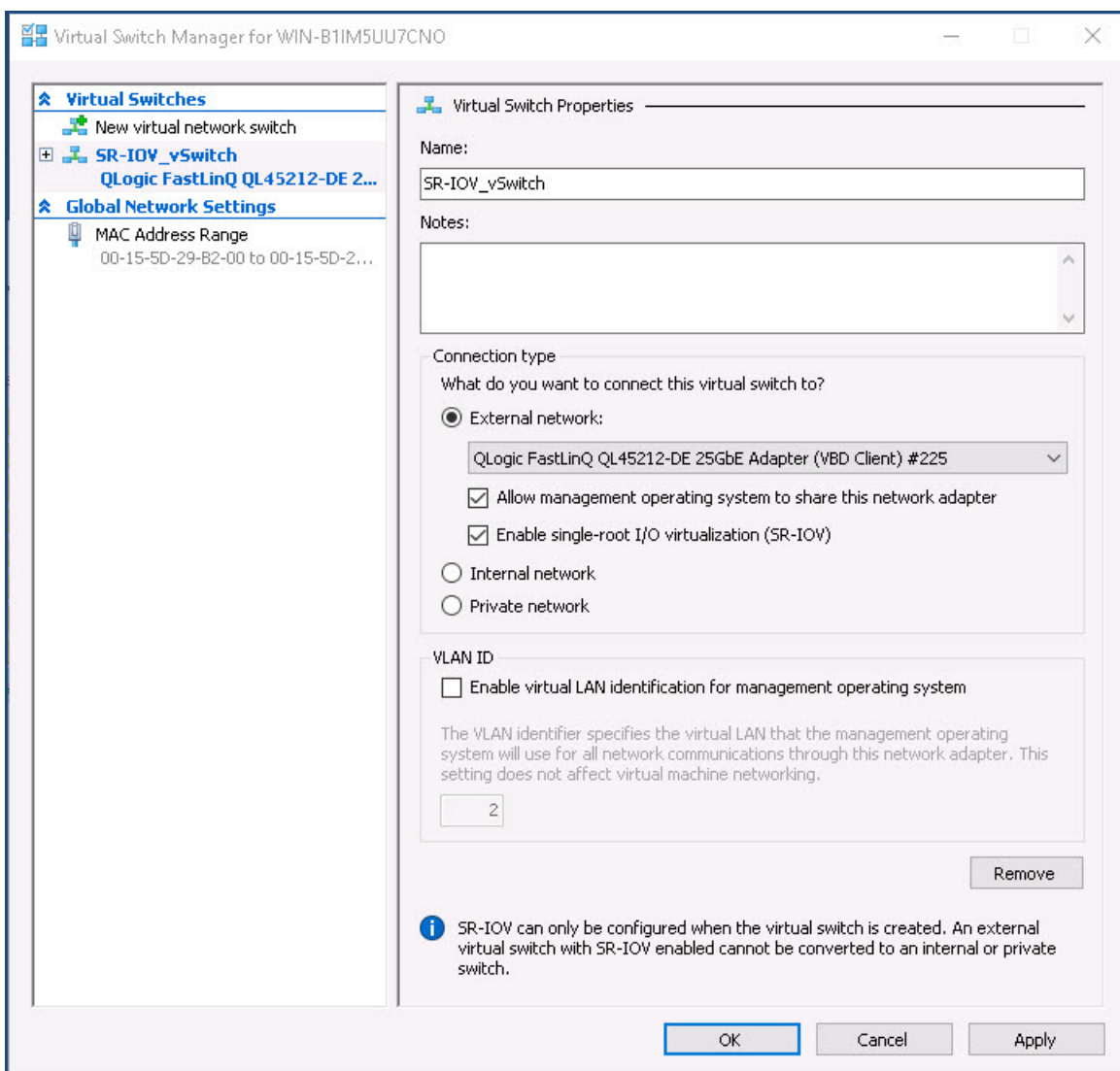


图 8-3. 适配器属性，高级：启用 SR-IOV

10. 要使用 SR-IOV 创建虚拟机交换机（第 93 页上的图 8-4）：
  - a. 启动 Hyper-V 管理器。
  - b. 选择 **Virtual Switch Manager**（虚拟交换机管理器）。
  - c. 在 **Name**（名称）框中，键入虚拟交换机的名称。
  - d. 在 **Connection type**（连接类型）下，选择 **External network**（外部网络）。
  - e. 选择 **Enable single-root I/O virtualization (SR-IOV)**（启用单根 I/O 虚拟化 (SR-IOV)）复选框，然后单击 **Apply**（应用）。

**注**

确保在创建 vSwitch 时启用 SR-IOV。创建 vSwitch 后，此选项将不可用。



**图 8-4. 虚拟交换机管理器：启用 SR-IOV**

- f. Apply Networking Changes（应用网络更改）消息框告知您待处理更改可能会中断网络连接。要保存您的更改并继续，请单击 **Yes**（是）。

11. 要获取虚拟机交换机功能，请发出以下 Windows PowerShell 命令：

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

Get-VMSwitch 命令的输出包括以下 SR-IOV 功能：

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

12. 要创建虚拟机 (VM) 并导出 VM 中的虚拟功能 (VF)：
- 创建一个虚拟机。
  - 将 VMNetworkadapter 添加到虚拟机。
  - 将虚拟交换机分配给 VMNetworkadapter。
  - 在 Settings for VM <VM\_Name> (VM <虚拟机名称> 的设置) 对话框 (图 8-5) 中，Hardware Acceleration (硬件加速) 页面的 **Single-root I/O virtualization** (单根 I/O 虚拟化) 下，选择 **Enable SR-IOV** (启用 SR-IOV) 复选框，然后单击 **OK** (确定)。

---

#### 注

创建虚拟适配器连接后，可以随时启用或禁用 SR-IOV 设置 (即使流量正在运行)。

---

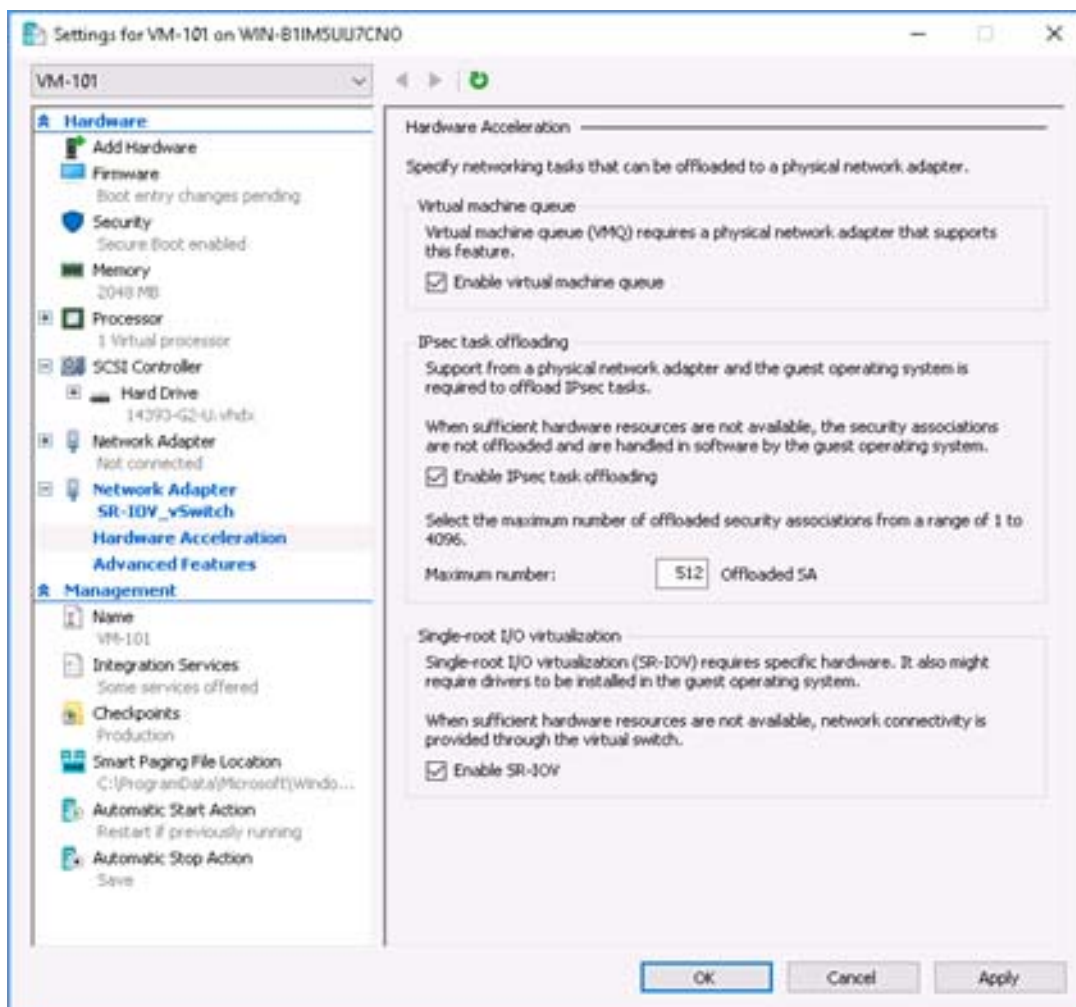


图 8-5. VM 的设置：启用 SR-IOV

13. 针对在虚拟机中检测到的适配器来安装 QLogic 驱动程序。针对主机 OS 使用可从供应商获取的最新驱动程序（请勿使用收件箱驱动程序）。

### 注

确保在 VM 和主机系统上使用相同的驱动程序包。例如，在 Windows VM 和 Windows Hyper-V 主机上使用相同的 qeVBD 和 qeND 驱动程序版本。

安装驱动程序后，QLogic 适配器在 VM 中列出。图 8-6 显示一个实例。

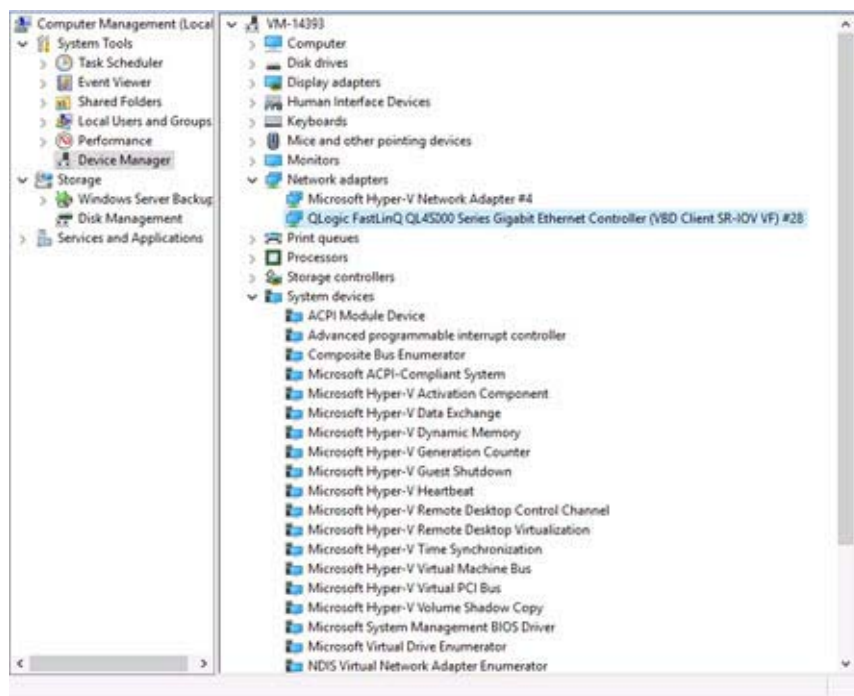


图 8-6. 设备管理器：带有 QLogic 适配器的 VM

14. 要查看 SR-IOV VF 详细信息，请发出以下 Windows PowerShell 命令：

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

图 8-7 显示示例输出。

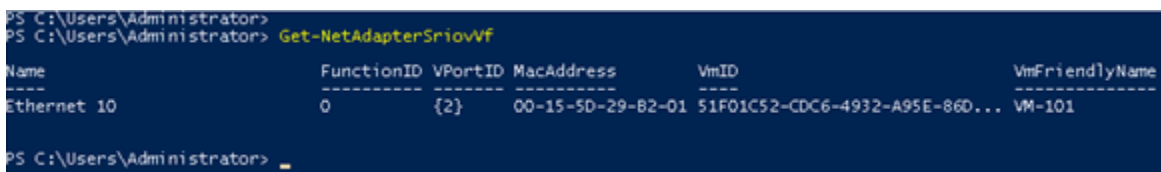


图 8-7. Windows PowerShell 命令：Get-NetadapterSriovVf

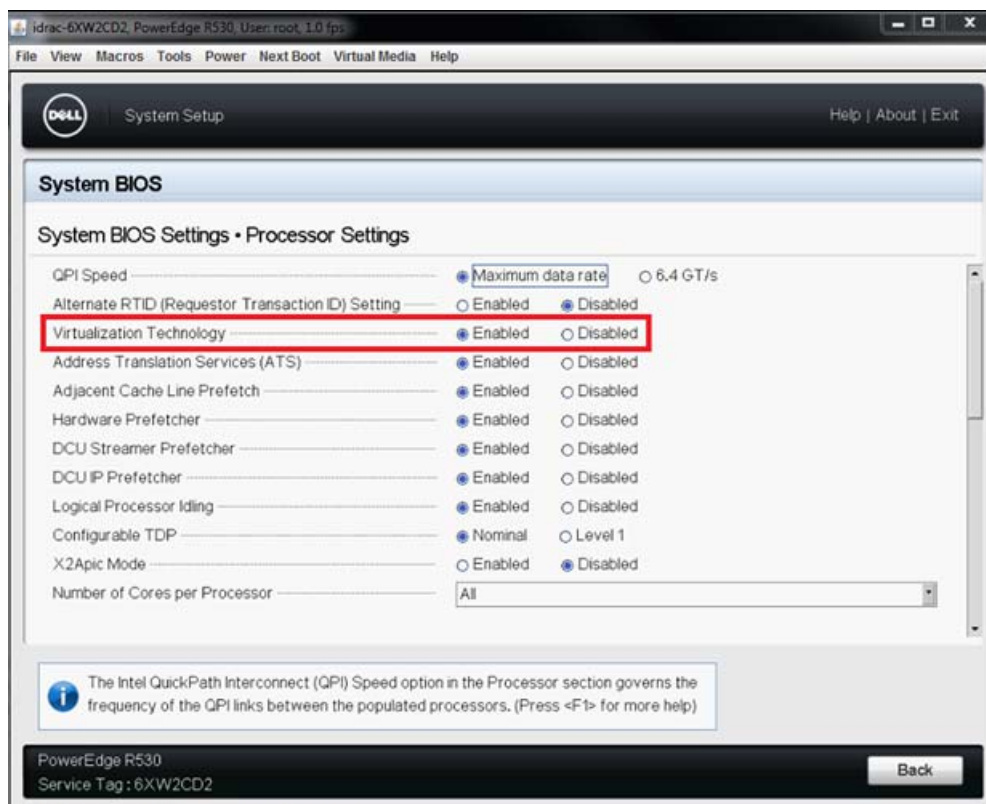


## 在 Linux 上配置 SR-IOV

要在 Linux 上配置 SR-IOV：

1. 访问服务器 BIOS System Setup（BIOS 系统设置），然后单击 **System BIOS Settings**（系统 BIOS 设置）。
2. 在 System BIOS Settings（系统 BIOS 设置）页面中，单击 **Integrated Devices**（集成的设备）。
3. 在 System Integrated Devices（系统集成设备）页面（请参阅[第 91 页上的图 8-1](#)）中：
  - a. 将 **SR-IOV Global Enable**（SR-IOV 全局启用）选项设置为 **Enabled**（已启用）。
  - b. 单击 **Back**（后退）。
4. 在 System BIOS Settings（系统 BIOS 设置）页面中，单击 **Processor Settings**（处理器设置）。

5. 在 Processor Settings（处理器设置）（图 8-8）页面中：
  - a. 将 **Virtualization Technology**（虚拟化技术）选项设置为 **Enabled**（已启用）。
  - b. 单击 **Back**（后退）。



**图 8-8. 系统设置：SR-IOV 的处理器设置**

6. 在 System Setup（系统设置）页面中，选择 **Device Settings**（设备设置）。
7. 在 Device Settings（设备设置）页面中，为 QLogic 适配器选择 **Port 1**（端口 1）。

8. 在 Device Level Configuration（设备级配置）页面（图 8-9）中：
  - a. 将 **Virtualization Mode**（虚拟化模式）设置为 **SR-IOV**。
  - b. 单击 **Back**（后退）。



**图 8-9. SR-IOV 的系统设置：集成设备**

9. 在 Main Configuration Page（主要配置页面）中，单击 **Finish**（完成），保存设置，然后重新引导系统。
10. 要启用并验证虚拟化：
  - a. 打开 `grub.conf` 文件并配置 `iommu` 参数，如图 8-10 中所示。
    - 对于基于 Intel 的系统，请添加 `intel_iommu=on`。
    - 对于基于 AMD 的系统，请添加 `amd_iommu=on`。

## 8-SR-IOV 配置 在 Linux 上配置 SR-IOV

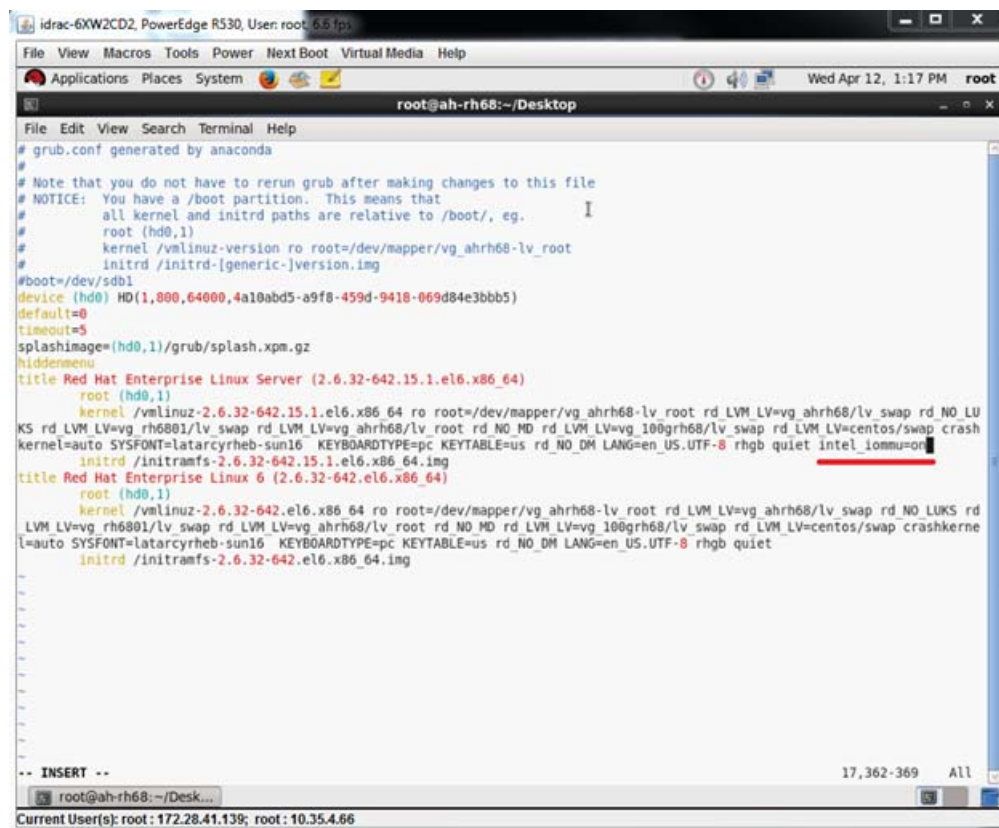


图 8-10. 为 SR-IOV 编辑 grub.conf 文件

b. 保存 grub.conf 文件，然后重新引导系统。

c. 要验证更改是否已生效，请发出以下命令：

```
dmesg | grep -I iommu
```

应显示成功的输入 - 输出内存管理单元 (IOMMU) 命令，例如：

```
Intel-IOMMU: enabled
```

d. 查看 VF 详细信息 (VF 数和 VF 总数)，请发出以下命令：

```
find /sys/|grep -I sriov
```

11. 对于特定端口，启用 VF 的数量。

a. 例如，发出以下命令以启用 PCI 实例 04:00.0 (总线 4，设备 0，功能 0) 上的 8 个 VF：

```
[root@ah-rh68 ~]# echo 8 >
```

```
/sys/devices/pci0000:00/0000:00:02.0/0000:04:00.0/sriov_numvfs
```

- b. 查看命令输出（图 8-11）以确认在总线 4、设备 2（通过 0000:00:02.0 参数）、功能 0 至 7 上创建了实际 VF。请注意，实际的设备 ID 在 PF（此例中为 8070）与 VF（此例中为 8090）上不同。

```
[root@ah-rh68 Desktop]#  
[root@ah-rh68 Desktop]# echo 8 > /sys/devices/pci0000:00/0000:00:02.0/0000:04:00.0/sriov_numvfs  
[root@ah-rh68 Desktop]#  
[root@ah-rh68 Desktop]# lspci -vv|grep -i QLogic  
04:00.0 Ethernet controller: QLogic Corp. Device 8070 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
    Product Name: QLogic 25GE 2P QL45212HxCU-DE Adapter  
    [V4] Vendor specific: NMVQLogic  
04:00.1 Ethernet controller: QLogic Corp. Device 8070 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
    Product Name: QLogic 25GE 2P QL45212HxCU-DE Adapter  
    [V4] Vendor specific: NMVQLogic  
04:02.0 Ethernet controller: QLogic Corp. Device 8090 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
04:02.1 Ethernet controller: QLogic Corp. Device 8090 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
04:02.2 Ethernet controller: QLogic Corp. Device 8090 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
04:02.3 Ethernet controller: QLogic Corp. Device 8090 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
04:02.4 Ethernet controller: QLogic Corp. Device 8090 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
04:02.5 Ethernet controller: QLogic Corp. Device 8090 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
04:02.6 Ethernet controller: QLogic Corp. Device 8090 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
04:02.7 Ethernet controller: QLogic Corp. Device 8090 (rev 02)  
    Subsystem: QLogic Corp. Device 000b  
[root@ah-rh68 Desktop]#
```

图 8-11. sriov\_numvfs 的命令输出

12. 要查看所有 PF 和 VF 接口的列表，请发出以下命令：

```
# ip link show | grep -i vf -b2
```

图 8-12 显示示例输出。

```
[root@localhost ~]# ip link show | grep -i vf -b2  
163-2: em1_1: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc mq state UP mode DEFAULT group default qlen 1000  
271- link/ether f4:e9:d4:ee:54:c2 brd ff:ff:ff:ff:ff:ff  
326: vf 0 MAC 00:00:00:00:00:00, tx rate 10000 (Mbps), max_tx_rate 10000Mbps, spoof checking off, link-state auto  
439: vf 1 MAC 00:00:00:00:00:00, tx rate 10000 (Mbps), max_tx_rate 10000Mbps, spoof checking off, link-state auto  
552: vf 2 MAC 00:00:00:00:00:00, tx rate 10000 (Mbps), max_tx_rate 10000Mbps, spoof checking off, link-state auto  
665: vf 3 MAC 00:00:00:00:00:00, tx rate 10000 (Mbps), max_tx_rate 10000Mbps, spoof checking off, link-state auto  
778: vf 4 MAC 00:00:00:00:00:00, tx rate 10000 (Mbps), max_tx_rate 10000Mbps, spoof checking off, link-state auto  
891: vf 5 MAC 00:00:00:00:00:00, tx rate 10000 (Mbps), max_tx_rate 10000Mbps, spoof checking off, link-state auto  
1004: vf 6 MAC 00:00:00:00:00:00, tx rate 10000 (Mbps), max_tx_rate 10000Mbps, spoof checking off, link-state auto  
1117: vf 7 MAC 00:00:00:00:00:00, tx rate 10000 (Mbps), max_tx_rate 10000Mbps, spoof checking off, link-state auto
```

图 8-12. ip link show 命令的命令输出

13. 分配并验证 MAC 地址：

- a. 要将 MAC 地址分配给 VF，请发出以下命令：

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

- b. 确保 VF 接口正常工作并以分配的 MAC 地址运行。
14. 关闭 VM 的电源并连接 VF。（某些操作系统支持热插拔 VF 到 VM。）
- a. 在 Virtual Machine（虚拟机）对话框（图 8-13）中，单击 **Add Hardware**（添加硬件）。

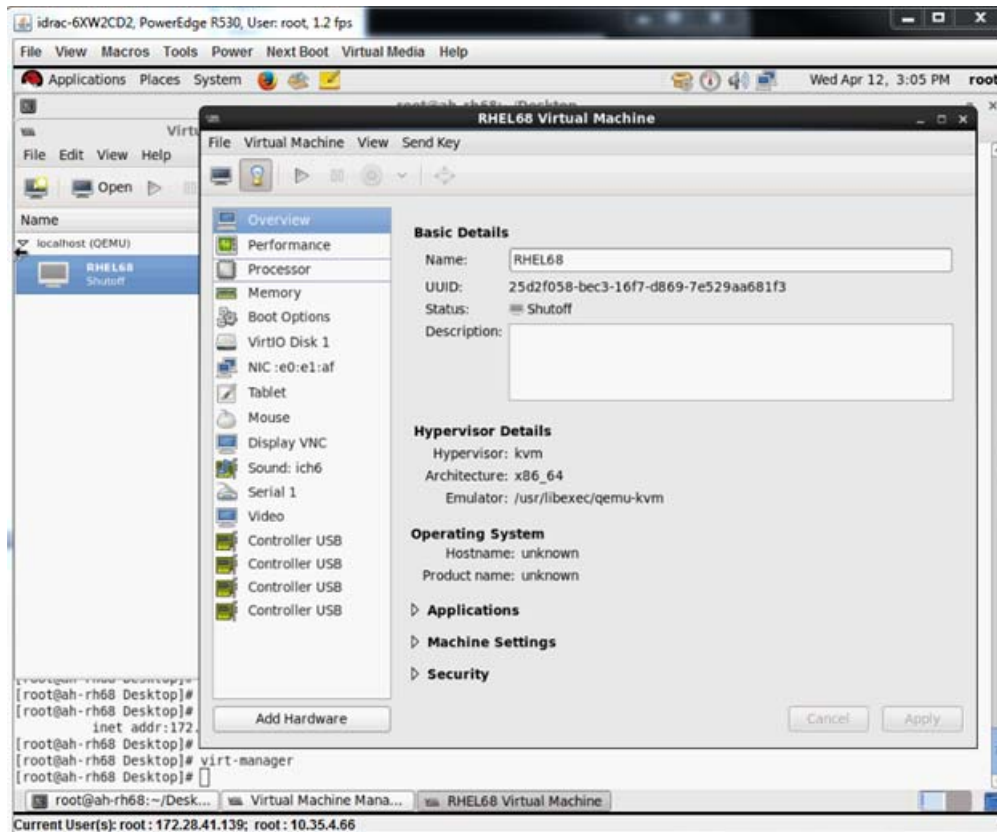


图 8-13. RHEL68 虚拟机

- b. 在 Add New Virtual Hardware（添加新虚拟硬件）对话框（图 8-14）的左侧窗格中，单击 **PCI Host Device**（PCI 主机设备）。
- c. 在右侧窗格中，选择一个主机设备。

- d. 单击 **Finish**（完成）。

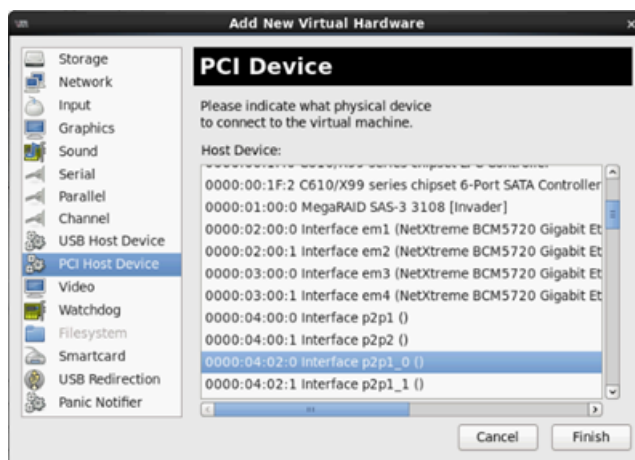


图 8-14. 添加新虚拟硬件

15. 打开 VM 的电源，然后发出以下命令：  

```
check lspci -vv|grep -I ether
```
16. 针对在虚拟机中检测到的适配器来安装驱动程序。针对主机 OS 使用可从供应商获取的最新驱动程序（请勿使用收件箱驱动程序）。在主机和虚拟机上必须安装相同的驱动程序版本。
17. 根据需要，在 VM 中添加更多 VF。

## 在 VMware 上配置 SR-IOV

要在 VMware 上配置 SR-IOV：

1. 访问服务器 BIOS System Setup（BIOS 系统设置），然后单击 **System BIOS Settings**（系统 BIOS 设置）。
2. 在 System BIOS Settings（系统 BIOS 设置）页面中，单击 **Integrated Devices**（集成的设备）。
3. 在 Integrated Devices（集成的设备）页面（请参阅第 91 页上的图 8-1）中：
  - a. 将 **SR-IOV Global Enable**（SR-IOV 全局启用）选项设置为 **Enabled**（已启用）。
  - b. 单击 **Back**（后退）。
4. 在 System Setup（系统设置）窗口中，单击 **Device Settings**（设备设置）。



5. 在 Device Settings（设备设置）页面中，选择用于 25G QL45212 Adapter 的端口。
6. 在 Device Level Configuration（设备级配置）（请参阅第 91 页上的图 8-2）中：
  - a. 将 **Virtualization Mode**（虚拟化模式）设置为 **SR-IOV**。
  - b. 单击 **Back**（后退）。
7. 在 Main Configuration Page（主要配置页面）中，单击 **Finish**（完成）。
8. 保存配置设置并重新引导系统。
9. 要启用每个端口所需的 VF 数量（在此例中，双端口适配器的每个端口上需要 16 个），请发出以下命令：

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

---

**注**

QL45212 Adapter 的每个以太网功能必须具备各自的条目。

---

10. 重新引导主机。
11. 要验证模块级别的更改是否已完成，请发出以下命令：

```
"esxcfg-module -g qedentv"
```

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

12. 要验证是否创建了实际 VF，请发出 `lspci` 命令如下：

```
[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]  
.  
.  
.
```



## 8-SR-IOV 配置

### 在 VMware 上配置 SR-IOV

---

```
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. 请如下将 VF 连接到 VM:
  - a. 关闭 VM 的电源并连接 VF。（某些操作系统支持热插拔 VF 到 VM。）
  - b. 将主机添加到 VMware vCenter Server 虚拟设备 (vCSA)。
  - c. 单击 VM 的 **Edit Settings**（编辑设置）。
14. 请如下填写 Edit Settings（编辑设置）对话框（[图 8-15](#)）:
  - a. 在 **New Device**（新设备）框中，选择 **Network**（网络），然后单击 **Add**（添加）。
  - b. 对于 **Adapter Type**（适配器类型），选择 **SR-IOV Passthrough**（SR-IOV 直通）。
  - c. 对于 **Physical Function**（物理功能），选择 QLogic VF。

- d. 要保存您的配置更改并关闭此对话框，请单击 **OK**（确定）。

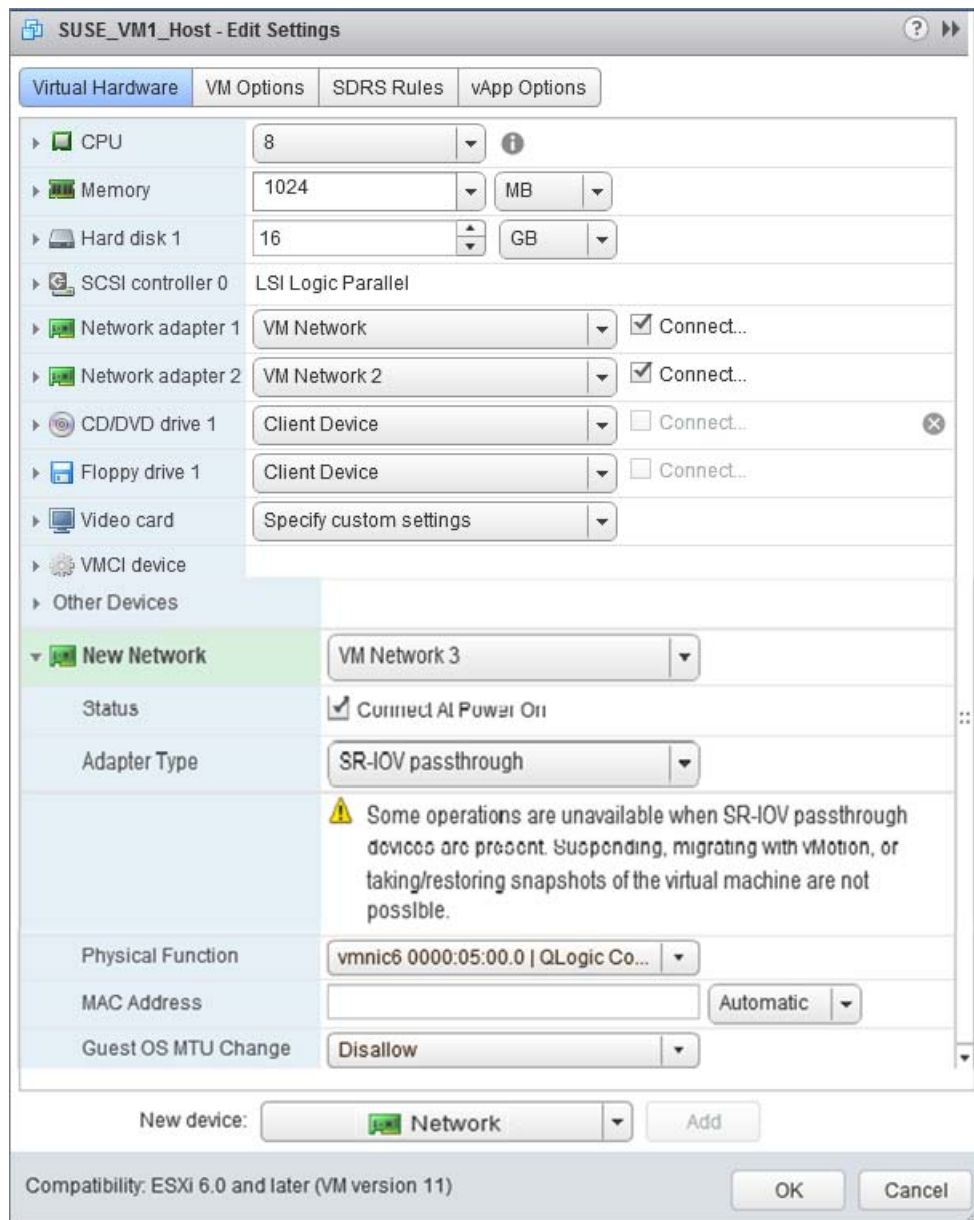


图 8-15. VMware 主机编辑设置

15. 要验证每个端口的 VF，请如下发出 `esxcli` 命令：

```
[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	-

16. 针对在虚拟机中检测到的适配器来安装 QLogic 驱动程序。针对主机操作系统使用可从供应商获取的最新驱动程序（请勿使用收件箱驱动程序）。在主机和虚拟机上必须安装相同的驱动程序版本。
17. 打开 VM 的电源，然后发出 `ifconfig -a` 命令以验证添加的网络接口是否已列出。
18. 根据需要，在 VM 中添加更多 VF。

# 9 Windows Server 2016

本章提供有关 Windows Server 2016 的以下信息：

- [使用 Hyper-V 配置 RoCE 接口](#)
- [第 114 页上的“Switch Embedded Teaming 上的 RoCE”](#)
- [第 115 页上的“为 RoCE 配置 QoS”](#)
- [第 124 页上的“配置 VMMQ”](#)
- [第 131 页上的“配置 VXLAN”](#)
- [第 132 页上的“配置 Storage Spaces Direct”](#)
- [第 138 页上的“部署和管理 Nano Server”](#)

## 使用 Hyper-V 配置 RoCE 接口

在 Windows Server 2016 中，具有网络直接内核提供程序接口 (NDKPI) 模式 2 的 Hyper-V 主机虚拟网络适配器（主机虚拟 NIC）均支持 RDMA。

### 注

Hyper-V 上的 RoCE 需要 DCBX。要配置 DCBX：

- [通过 HII 进行配置](#)（请参阅 [第 54 页上的“准备适配器”](#)）。
- [使用 QoS 进行配置](#)（请参阅 [第 115 页上的“为 RoCE 配置 QoS”](#)）。

本节中的 RoCE 配置步骤包括：

- [创建带 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机](#)
- [将 VLAN ID 添加到主机虚拟 NIC](#)
- [验证 RoCE 是否启用](#)
- [添加主机虚拟 NIC（虚拟端口）](#)
- [映射 SMB 驱动器和运行 RoCE 流量](#)

## 创建带 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机

按照本节中的步骤操作以创建 Hyper-V 虚拟交换机，然后在主机 VNIC 中启用 RDMA。

要创建带 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机：

1. 启动 Hyper-V 管理器。
2. 单击 **Virtual Switch Manager**（虚拟交换机管理器）（请参阅图 9-1）。

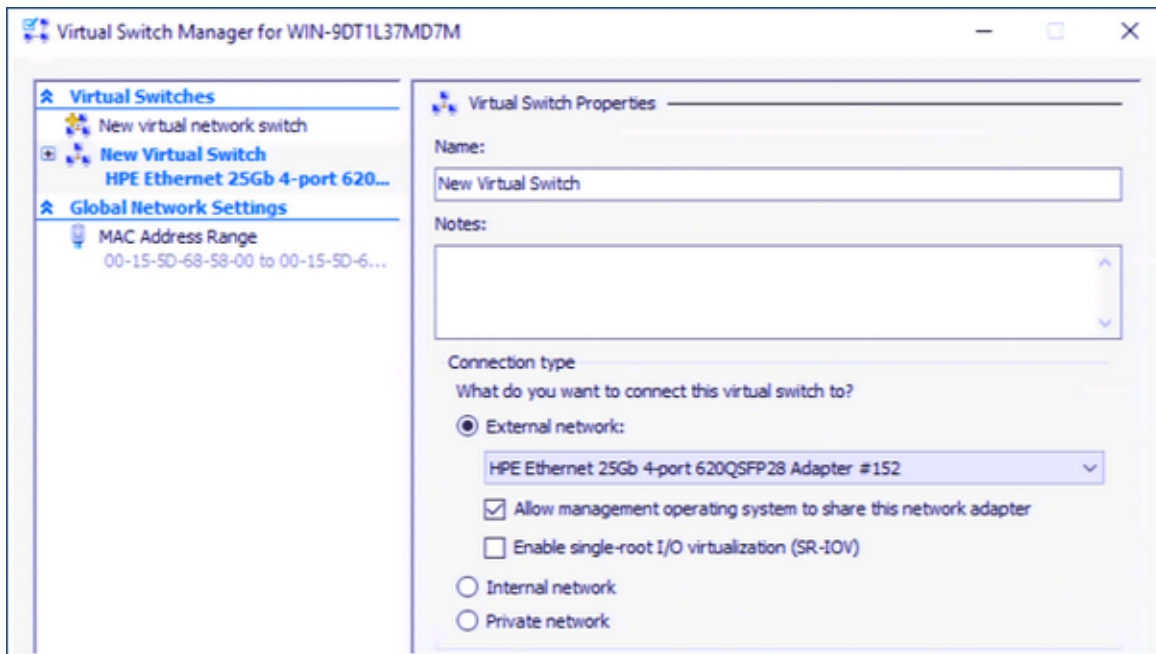


图 9-1. 在主机虚拟 NIC 中启用 RDMA

3. 创建一个虚拟交换机。
4. 选择 **Allow management operating system to share this network adapter**（允许管理操作系统以共享此网络适配器）复选框。

在 Windows Server 2016 中，将在主机虚拟 NIC 中添加一个新参数：Network Direct (RDMA)。

要在主机虚拟 NIC 中启用 RDMA：

1. 打开 Hyper-V Virtual Ethernet Adapter Properties（Hyper-V 虚拟以太网适配器属性）窗口。
2. 单击 **Advanced**（高级）选项卡。

3. 在 Advanced（高级）页面（图 9-2）中：
  - a. 在 **Property**（属性）下，选择 **Network Direct (RDMA)**。
  - b. 在 **Value**（值）下，选择 **Enabled**（已启用）。
  - c. 单击 **OK**（确定）。

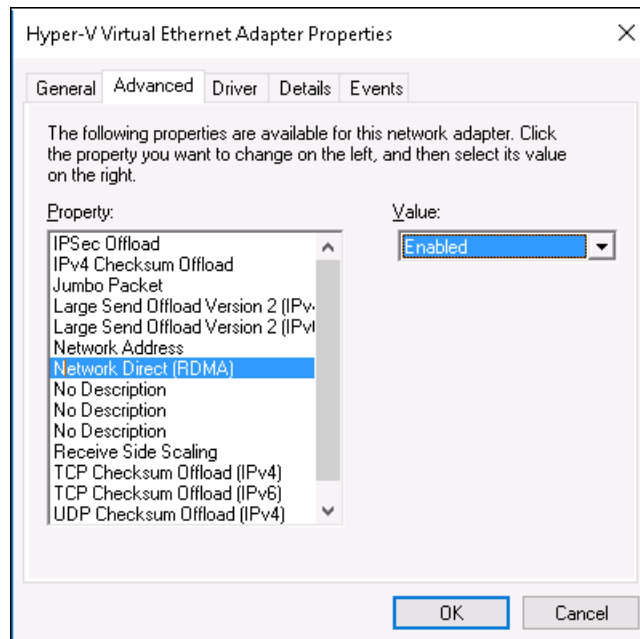


图 9-2. Hyper-V 虚拟以太网适配器属性

4. 要启用 RDMA，请发出以下 Windows PowerShell 命令：

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

## 将 VLAN ID 添加到主机虚拟 NIC

要将 VLAN ID 添加到主机虚拟 NIC：

1. 要查找主机虚拟 NIC 名称，请发出以下 Windows PowerShell 命令：

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

图 9-3 显示命令输出。

```
PS C:\Users\Administrator> Get-VMNetworkAdapter -ManagementOS
Name                IsManagementOs VMName SwitchName                MacAddress Status IPAddresses
-----
New Virtual Switch True                New Virtual Switch 000E1EC41F0B {Ok}
```

图 9-3. Windows PowerShell 命令: `Get-VMNetworkAdapter`

2. 要设置主机虚拟 NIC 的 VLAN ID, 请发出以下 Windows PowerShell 命令:

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

### 注

请注意关于将 VLAN ID 添加到主机虚拟 NIC 的以下事项:

- 必须为主机虚拟 NIC 分配一个 VLAN ID。必须为所有接口和交换机上分配相同的 VLAN ID。
- 将主机虚拟 NIC 用于 RoCE 时, 确保 VLAN ID 没有分配给物理接口。
- 如果创建多个主机虚拟 NIC, 可以为每个主机虚拟 NIC 分配不同的 VLAN。

## 验证 RoCE 是否启用

要验证 RoCE 是否启用:

- 发出以下 Windows PowerShell 命令:

```
Get-NetAdapterRdma
```

命令输出列出 RDMA 支持的适配器, 如图 9-4 中所示。

```
PS C:\Users\Administrator> Get-NetAdapterRdma
Name                InterfaceDescription                Enabled
-----
vEthernet (New Virtual... Hyper-V Virtual Ethernet Adapter True
```

图 9-4. Windows PowerShell 命令: `Get-NetAdapterRdma`

## 添加主机虚拟 NIC（虚拟端口）

### 要添加主机虚拟 NIC：

1. 要添加主机虚拟 NIC，请发出以下命令：  

```
Add-VMNetworkAdapter -SwitchName "New Virtual Switch" -Name  
SMB - ManagementOS
```
2. 在主机虚拟 NIC 上启用 RDMA，如第 109 页上的“要在主机虚拟 NIC 中启用 RDMA：”中所示。
3. 要将 VLAN ID 分配给虚拟端口，请发出以下命令：  

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

## 映射 SMB 驱动器和运行 RoCE 流量

### 要映射 SMB 驱动器和运行 RoCE 流量：

1. 启动性能监视器 (Perfmon)。
2. 按如下方式填写 Add Counters（添加计数器）对话框（图 9-5）：
  - a. 在 **Available counters**（可用计数器）下，选择 **RDMA Activity**（RDMA 活动）。
  - b. 在 **Instances of selected object**（所选对象的实例）下，选择适配器。



c. 单击添加。

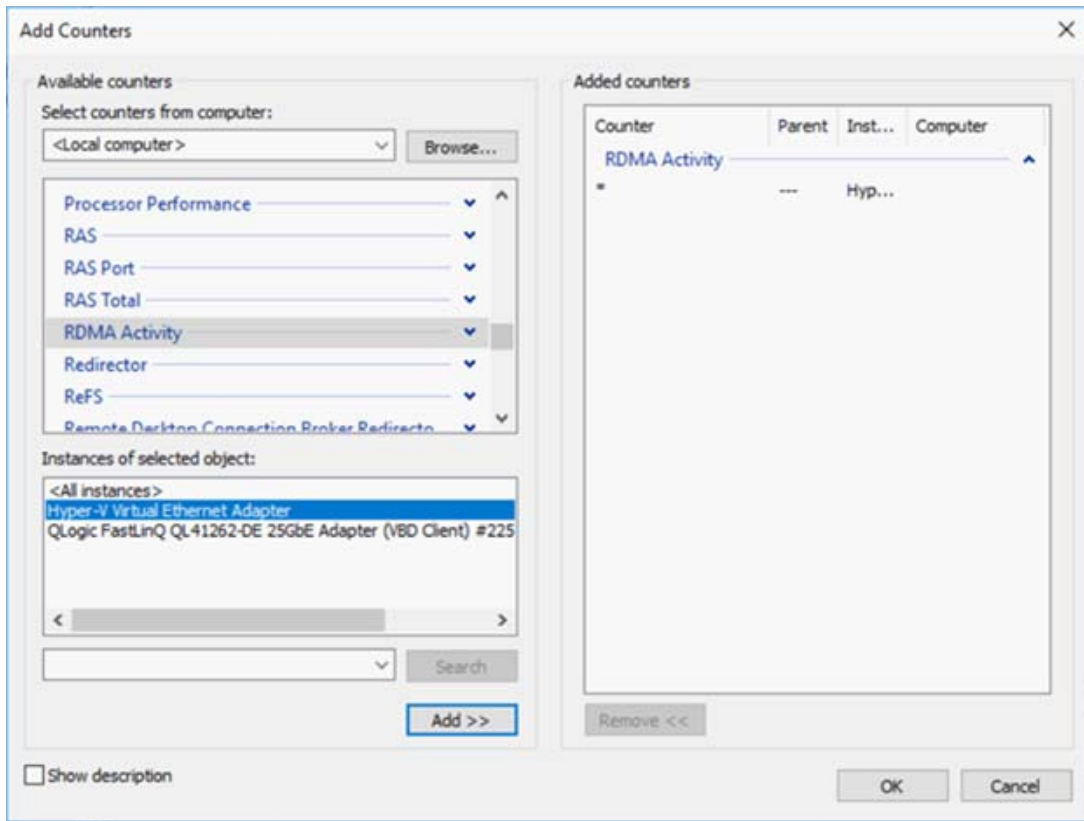


图 9-5. 添加计数器对话框

如果 RoCE 流量正在运行，计数器将显示为如图 9-6 中所示。

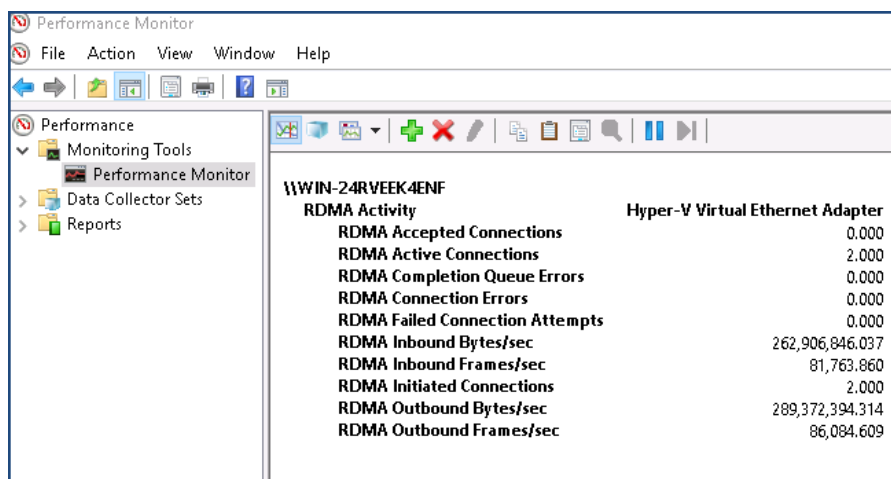


图 9-6. 显示 RoCE 流量的性能监视器

## Switch Embedded Teaming 上的 RoCE

Switch Embedded Teaming (SET) 是 Microsoft 的备选 NIC 组合解决方案，可在 Windows Server 2016 Technical Preview 中包括 Hyper-V 和软件定义网络 (SDN) 堆栈的环境中使用。SET 将受限的 NIC 组合功能集成到 Hyper-V 虚拟交换机中。

使用 SET 将一到八个物理以太网网络适配器分组为一个或多个基于软件的虚拟网络适配器。如果网络适配器发生故障，这些适配器可提供快速性能和容错。SET 成员网络适配器必须均安装在同一物理 Hyper-V 主机中方可放在一个组中。

本节中包括以下 SET 上的 RoCE 步骤：

- 创建带 SET 和 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机
- 在 SET 上启用 RDMA
- 在 SET 上分配 VLAN ID
- 在 SET 上运行 RDMA 流量

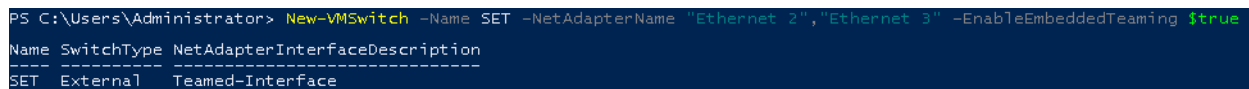
### 创建带 SET 和 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机

要创建带 SET 和 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机：

- 要创建 SET，请发出以下 Windows PowerShell 命令：

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

图 9-7 显示命令输出。



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

图 9-7. Windows PowerShell 命令：New-VMSwitch

### 在 SET 上启用 RDMA

要在 SET 上启用 RDMA：

1. 要查看适配器上的 SET，请发出以下 Windows PowerShell 命令：

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

图 9-8 显示命令输出。

```
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

图 9-8. Windows PowerShell 命令: Get-NetAdapter

2. 要在 SET 上启用 RDMA，请发出以下 Windows PowerShell 命令：

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

## 在 SET 上分配 VLAN ID

要在 SET 上分配 VLAN ID：

- 要在 SET 上分配 VLAN ID，请发出以下 Windows PowerShell 命令：

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

### 注

请注意将 VLAN ID 添加到主机虚拟 NIC 时的以下事项：

- 将主机虚拟 NIC 用于 RoCE 时，确保 VLAN ID 没有分配给物理接口。
- 如果创建多个主机虚拟 NIC，可以为每个主机虚拟 NIC 分配不同的 VLAN。

## 在 SET 上运行 RDMA 流量

有关在 SET 上运行 RDMA 流量的信息，请访问：

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

## 为 RoCE 配置 QoS

配置服务质量 (QoS) 的两种方法包括：

- [通过在适配器上禁用 DCBX 配置 QoS](#)
- [通过在适配器上启用 DCBX 配置 QoS](#)

## 通过在适配器上禁用 DCBX 配置 QoS

通过在适配器上禁用 DCBX 配置服务质量之前，必须完成所有使用中系统的所有配置。基于优先级的流控制 (PFC)、增强的转换服务 (ETS) 和流量类配置在交换机和服务器上必须相同。

### 要通过禁用 DCBX 配置 QoS:

1. 在适配器上禁用 DCBX。
2. 使用 HII，将 **RoCE Priority**（RoCE 优先级）设置为 0。
3. 要在主机中安装 DCB 角色，请发出以下 Windows PowerShell 命令：

```
PS C:\Users\Administrators> Install-WindowsFeature  
Data-Center-Bridging
```
4. 要将 **DCBX Willing**（DCBX 愿意）模式设置为 **False**（假），请发出以下 Windows PowerShell 命令：

```
PS C:\Users\Administrators> set-NetQosDcbxSetting -Willing 0
```
5. 请如下在微型端口中启用 QoS:
  - a. 打开微型端口窗口，然后单击 **Advanced**（高级）选项卡。
  - b. 在适配器的 Advanced Properties（高级属性）页面（[图 9-9](#)）的 **Property**（属性）下，选择 **Quality of Service**（服务质量），然后将值设置为 **Enabled**（已启用）。

- c. 单击 **OK**（确定）。

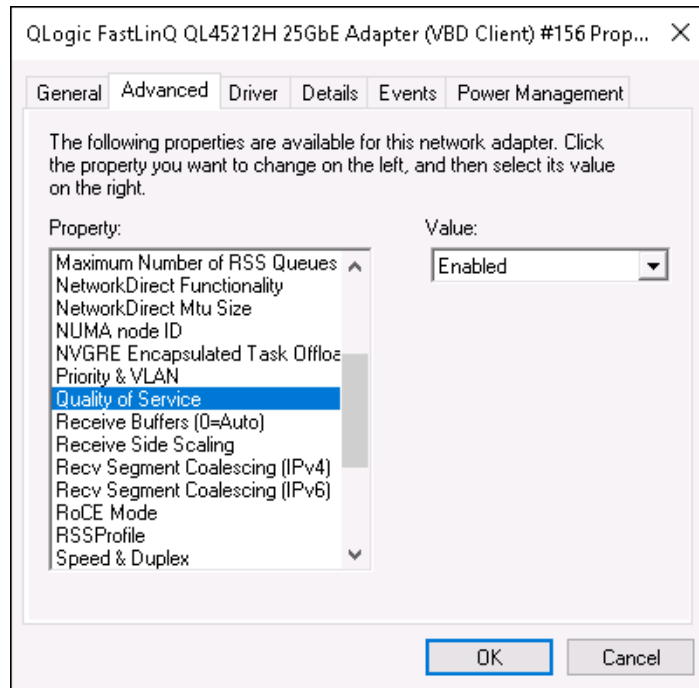
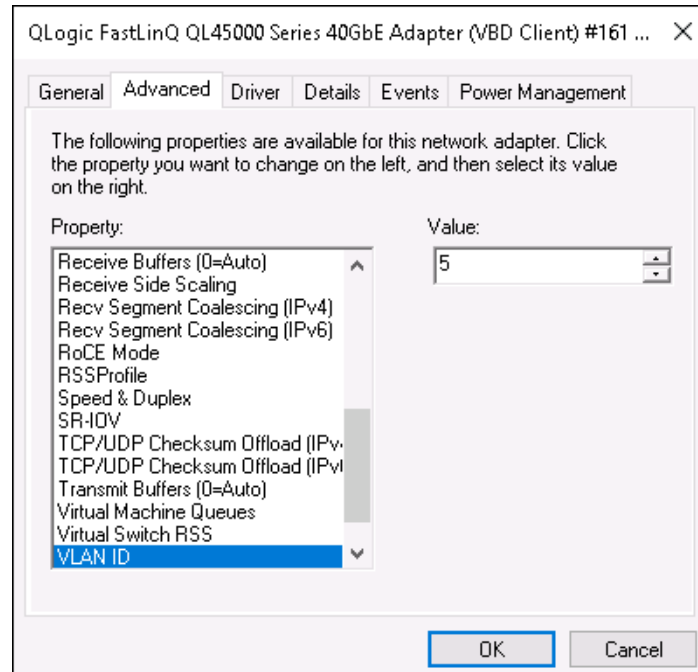


图 9-9. 高级属性：启用 QoS

6. 请如下将 VLAN ID 分配给接口：
- 打开微型端口窗口，然后单击 **Advanced**（高级）选项卡。
  - 在适配器的 Advanced Properties（高级属性）页面（图 9-10）的 **Property**（属性）下，选择 **VLAN ID**，然后设置值。
  - 单击 **OK**（确定）。

**注**

优先级流控制 (PFC) 需要上述步骤。



**图 9-10. 高级属性：设置 VLAN ID**

7. 要基于特定优先级启用 RoCE 的优先级流控制，请发出以下命令：

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

**注**

如果通过 Hyper-V 配置 RoCE，请勿将 VLAN ID 分配给物理接口。

8. 要基于任何其他优先级禁用优先级流控制，请发出以下命令：

```
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. 要为每种类型的流量配置 QoS 和分配相关的优先级，请发出以下命令（其中，优先级 4 标记用于 RoCE 而优先级 0 标记用于 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. 要为之前步骤中定义的所有流量类配置 ETS，请发出以下命令：

```
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. 要查看之前配置的网络适配器 QoS，请发出以下 Windows PowerShell 命令：

```
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. 创建启动脚本，以使设置在系统重新引导后保持不变。

13. 运行 RDMA 流量并验证，如第 52 页上的“RoCE 配置”中所述。



## 通过在适配器上启用 DCBX 配置 QoS

必须在所有使用中的系统上完成所有配置。PFC、ETS 和流量类配置在交换机和服务器上必须相同。

### 要通过启用 DCBX 配置 QoS：

1. 启用 DCBX（IEEE、CEE 或动态）。
2. 使用 HII，将 **RoCE Priority**（RoCE 优先级）设置为 0。
3. 要在主机中安装 DCB 角色，请发出以下 Windows PowerShell 命令：

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

---

### 注

对于此配置，将 **DCBX Protocol**（DCBX 协议）设置为 **CEE**。

---

4. 要将 **DCBX Willing**（DCBX 愿意）模式设置为 **True**（真），请发出以下命令：

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

5. 请如下在微型端口中启用 QoS:
  - a. 在适配器的 Advanced Properties（高级属性）页面（图 9-11）的 **Property**（属性）下，选择 **Quality of Service**（服务质量），然后将值设置为 **Enabled**（已启用）。
  - b. 单击 **OK**（确定）。

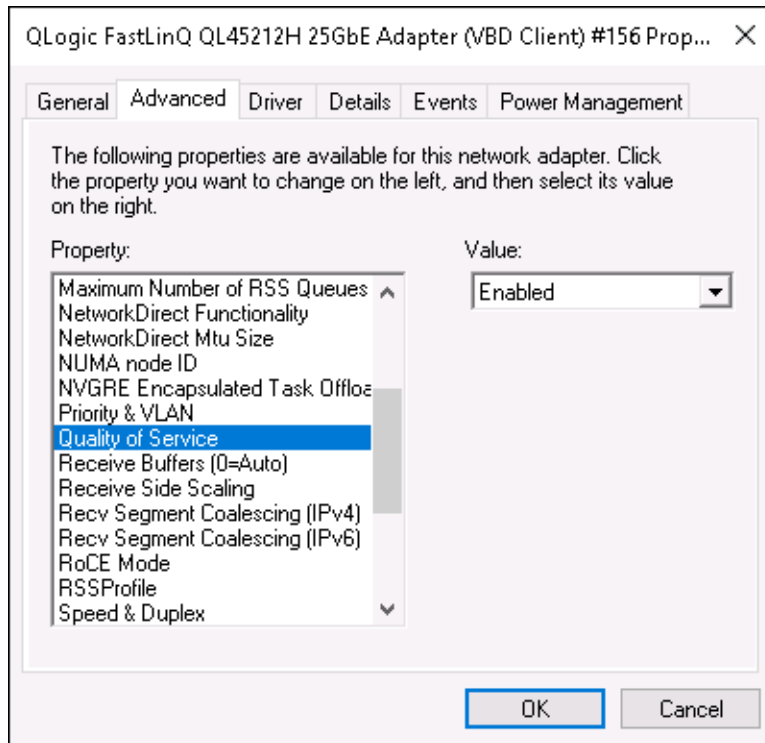


图 9-11. 高级属性：启用 QoS

6. 请如下将 VLAN ID 分配给接口（PFC 需要）:
  - a. 打开微型端口窗口，然后单击 **Advanced**（高级）选项卡。
  - b. 在适配器的 Advanced Properties（高级属性）页面（图 9-12）的 **Property**（属性）下，选择 **VLAN ID**，然后设置值。
  - c. 单击 **OK**（确定）。

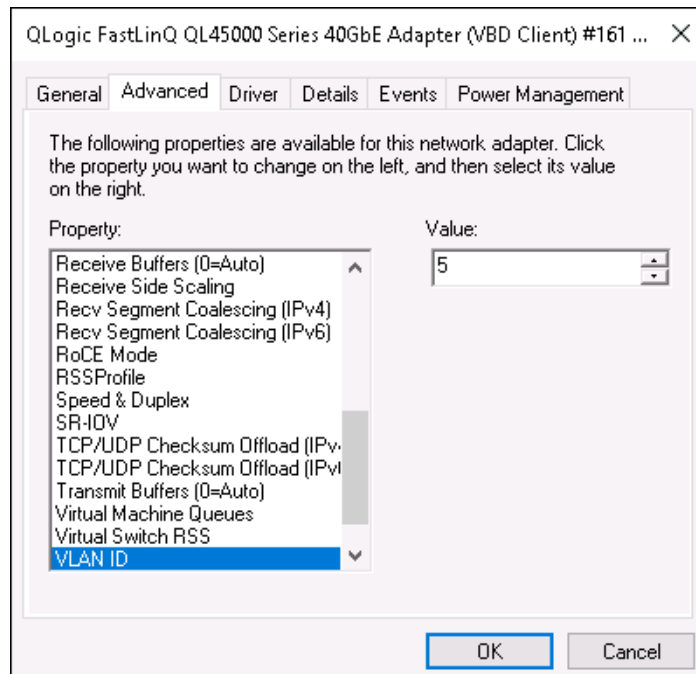


图 9-12. 高级属性：设置 VLAN ID

7. 要配置交换机，请发出以下 Windows PowerShell 命令：

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

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

```

**注**

上例为适配器端口连接到 Arista® 7060X 交换机时采用。在此示例中，交换机 PFC 启用为优先级 4。RoCE App TLV 已定义。两个流量类定义为 TC0 和 TC1，其中 TC1 定义用于 RoCE。**DCBX Protocol**（DCBX 协议）模式设置为 **CEE**。有关 Arista 交换机配置，请参阅第 54 页上的“准备以太网交换机”。适配器处于 **Willing**（愿意）模式时，它会接受远程配置并将其显示为 **Operational Parameters**（操作参数）。

## 配置 VMMQ

虚拟机多队列 (VMMQ) 配置信息包括：

- 在适配器上启用 VMMQ
- 设置 VMMQ 最大 QP 默认和非默认 VPort
- 创建带或不带 SR-IOV 的虚拟机交换机
- 在虚拟机交换机上启用 VMMQ
- 获取虚拟机交换机功能
- 创建 VM 并在 VM 中的 VMNetworkadapter 上启用 VMMQ
- 默认和最大 VMMQ 虚拟 NIC
- 在管理 NIC 上启用和禁用 VMMQ
- 监测流量统计信息

## 在适配器上启用 VMMQ

要在适配器上启用 VMMQ：

1. 打开微型端口窗口，然后单击 **Advanced**（高级）选项卡。
2. 在 Advanced Properties（高级属性）页面（图 9-13）的 **Property**（属性）下，选择 **Virtual Switch RSS**（虚拟交换机 RSS），然后将值设置为 **Enabled**（已启用）。
3. 单击 **OK**（确定）。

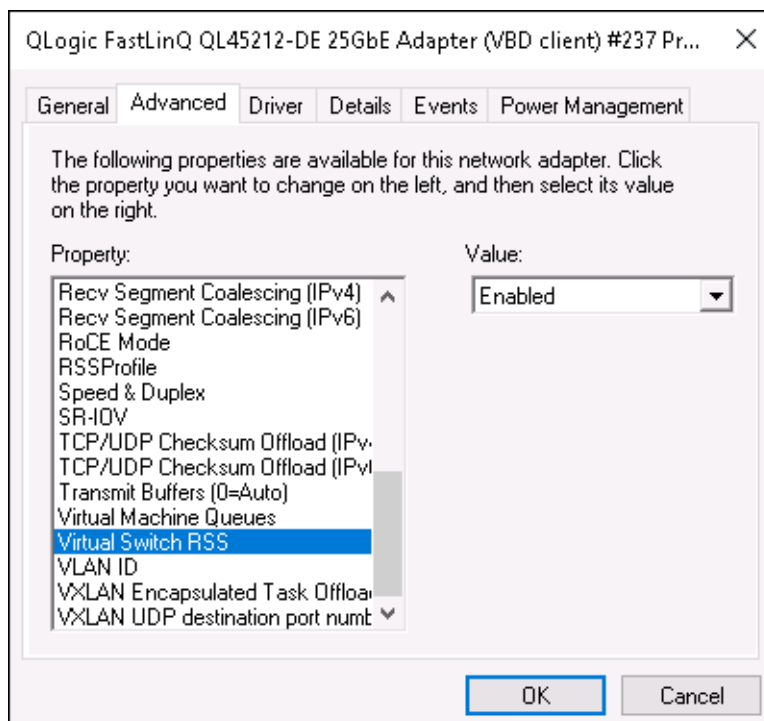


图 9-13. 高级属性：启用虚拟交换机 RSS

## 设置 VMMQ 最大 QP 默认和非默认 VPort

要设置 VMMQ 最大 QP 默认和非默认 VPort：

1. 打开微型端口窗口，然后单击 **Advanced**（高级）选项卡。
2. 在 Advanced Properties（高级属性）页面的 **Property**（属性）下，选择以下项之一：
  - VMMQ Max QPs - Default VPort**（VMMQ 最大 QP - 默认 VPort）

- VMMQ Max QPs - Non-Default VPort** (VMMQ 最大 QP - 非默认 VPort)

3. 如果适用, 调整所选属性的 **Value** (值)。
4. 单击 **OK** (确定)。

## 创建带或不带 SR-IOV 的虚拟机交换机

要创建带或不带 SR-IOV 的虚拟机交换机:

1. 启动 Hyper-V 管理器。
2. 选择 **Virtual Switch Manager** (虚拟交换机管理器) (请参阅 [图 9-14](#))。
3. 在 **Name** (名称) 框中, 键入虚拟交换机的名称。

4. 在 **Connection type**（连接类型）下：
  - a. 单击 **External network**（外部网络）。
  - b. 选择 **Allow management operating system to share this network adapter**（允许管理操作系统共享此网络适配器）复选框。

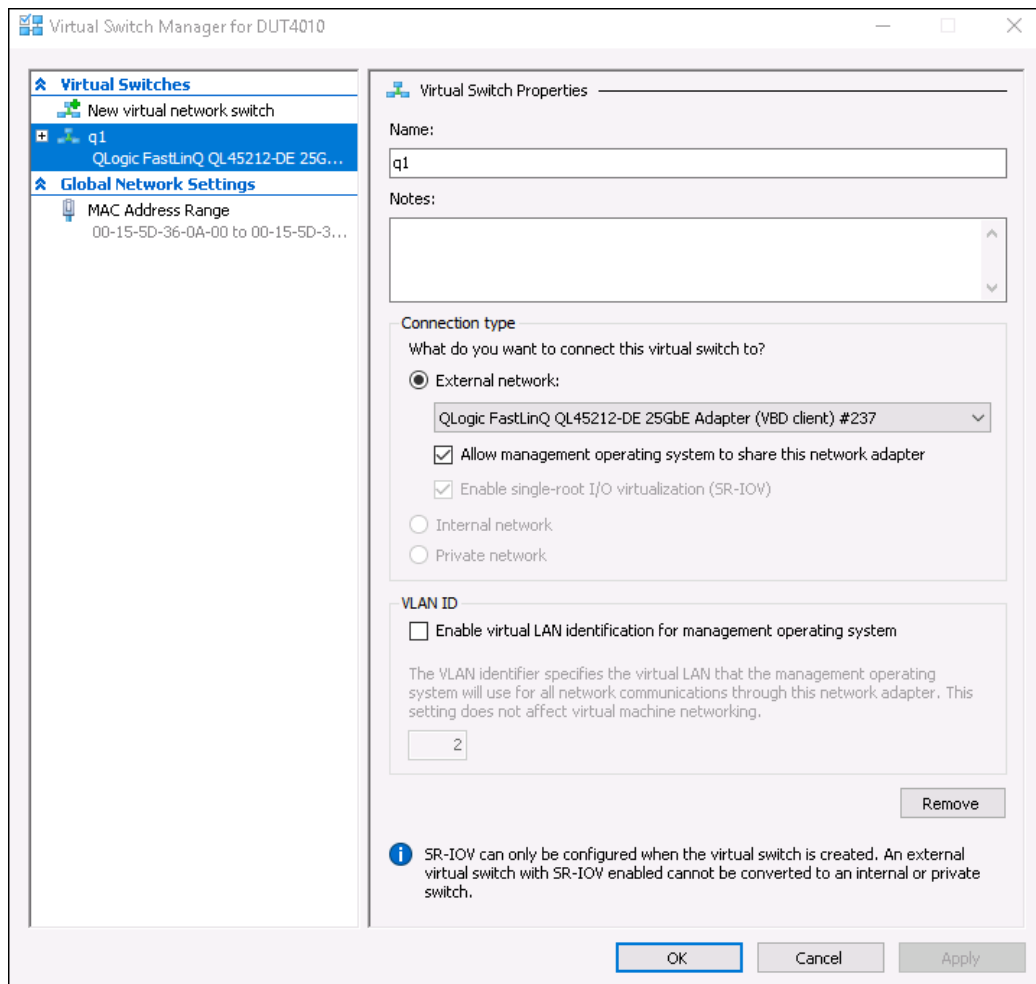


图 9-14. 虚拟交换机管理器

5. 单击 **OK**（确定）。

## 在虚拟机交换机上启用 VMMQ

要在虚拟机交换机上启用 VMMQ:

- 发出以下 Windows PowerShell 命令:

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

## 获取虚拟机交换机功能

要获取虚拟机交换机功能:

- 发出以下 Windows PowerShell 命令:

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

图 9-15 显示示例输出。

```
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  
NumberIPSecSAA1located : 0  
AvailableVMQueues : 16  
NumberVmqA1located : 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
```

图 9-15. Windows PowerShell 命令: Get-VMSwitch



## 创建 VM 并在 VM 中的 VMNetworkadapter 上启用 VMMQ

要创建虚拟机 (VM) 并在 VM 中的 VMNetworkadapter 上启用 VMMQ:

1. 创建 VM。
2. 将 VMNetworkadapter 添加到 VM。
3. 将虚拟交换机分配给 VMNetworkadapter。
4. 要在 VM 上启用 VMMQ，请发出以下 Windows PowerShell 命令:

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

### 注

对于支持 SR-IOV 的虚拟交换机：如果 VM 交换机和硬件加速启用了 SR-IOV，您必须创建 10 个 VM（每个带有 8 个虚拟 NIC）以利用 VMMQ。此要求是因为 SR-IOV 优先于 VMMQ。

64 个虚拟功能和 16 个 VMMQ 的示例输出如下所示:

```
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-netadaptervmmq
```

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

## 默认和最大 VMMQ 虚拟 NIC

根据当前实施，每个虚拟 NIC 最多 4 个可用的 VMMQ；也就是说，最多 16 个虚拟 NIC。

如之前使用 Windows PowerShell 命令所设置，有四个可用的默认队列。最大默认队列当前可设置为 8。要验证最大默认队列，请使用 VMswitch 功能。

## 在管理 NIC 上启用和禁用 VMMQ

要在管理 NIC 上启用或禁用 VMMQ：

- 要在管理 NIC 上启用 VMMQ，请发出以下命令：

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

MOS VNIC 拥有四个 VMMQ。

- 要在管理 NIC 上禁用 VMMQ，请发出以下命令：

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

VMMQ 也将可用于多播开放式最短路径优先 (MOSPF)。

## 监测流量统计信息

要监测虚拟机中的虚拟功能流量，请发出以下 Windows PowerShell 命令：

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

## 配置 VXLAN

VXLAN 配置信息包括：

- 在适配器上启用 VXLAN 卸载
- 部署软件定义网络

### 在适配器上启用 VXLAN 卸载

要在适配器上启用 VXLAN 卸载：

1. 打开微型端口窗口，然后单击 **Advanced**（高级）选项卡。
2. 在 Advanced Properties（高级属性）页面（图 9-16）的 **Property**（属性）下，选择 **VXLAN Encapsulated Task Offload**（VXLAN 封装任务卸载）。

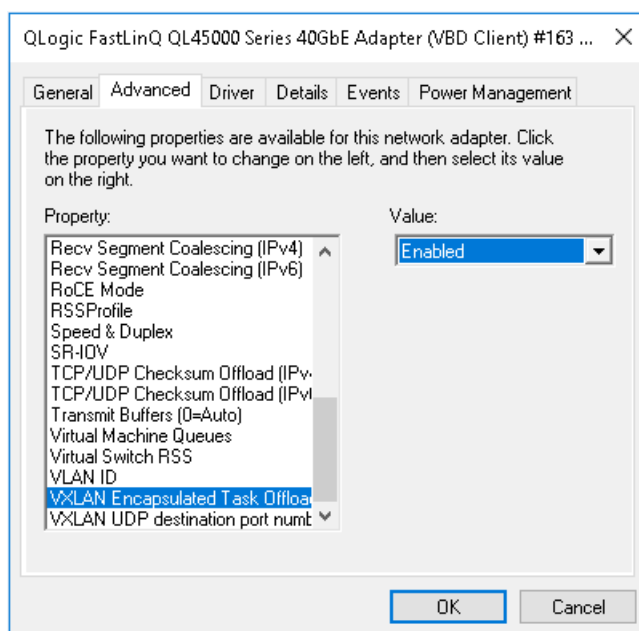


图 9-16. 高级属性：启用 VXLAN

3. 将 **Value**（值）设置为 **Enabled**（已启用）。
4. 单击 **OK**（确定）。

## 部署软件定义网络

要利用虚拟机上的 VXLAN 封装任务卸载，必须部署使用 Microsoft 网络控制器的软件定义网络 (SDN) 堆栈。

有关更多详细信息，请参阅有关软件定义网络的以下 Microsoft TechNet 链接：

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

## 配置 Storage Spaces Direct

Windows Server 2016 引入了 Storage Spaces Direct，这样便可通过本地存储构建高度可用和可扩展的存储系统。

有关更多信息，请参阅以下 Microsoft TechnNet 链接：

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

## 配置硬件

图 9-17 显示硬件配置的示例。

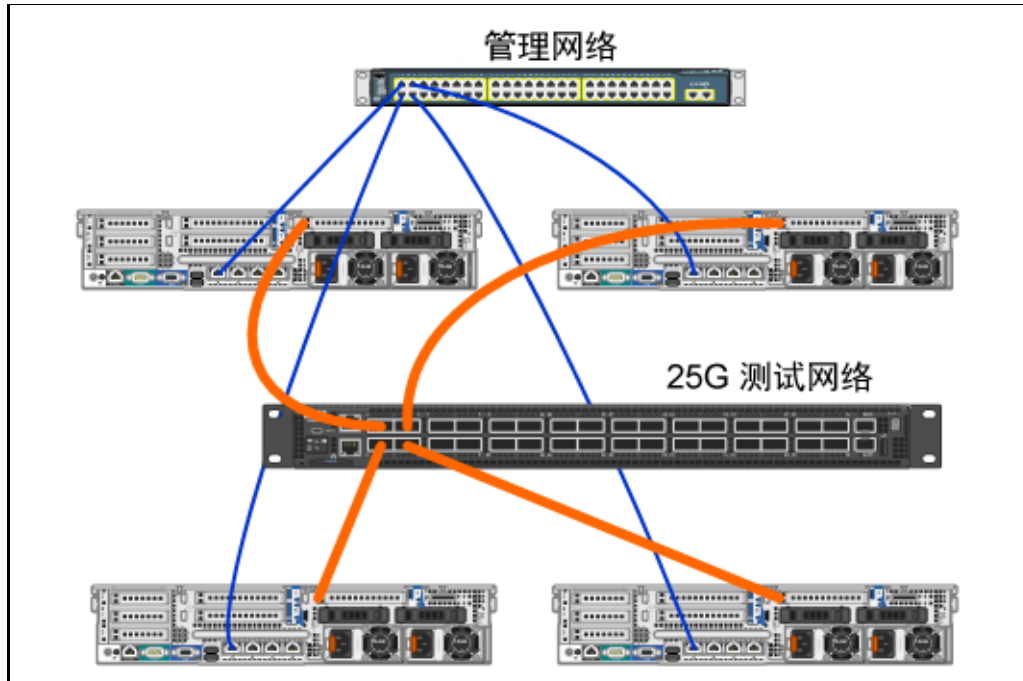


图 9-17. 示例性硬件配置

### 注

本例中使用的磁盘为 4 × 400G NVMe™ 和 12 × 200G SSD 磁盘。

## 部署超聚合系统

本节包括使用 Windows Server 2016 安装和配置超聚合系统组件的简介。部署超聚合系统的操作可以分为以下三个高级阶段：

- 部署操作系统
- 配置网络
- 配置 Storage Spaces Direct

### 部署操作系统

要部署操作系统：

1. 安装操作系统。
2. 安装 Windows 服务器角色 (Hyper-V)。

3. 安装以下功能：
  - 故障转移
  - 群集
  - 数据中心桥接 (DCB)
4. 将节点连接到域并添加域帐户。

## 配置网络

要部署 Storage Spaces Direct，Hyper-V 交换机必须部署为带有启用 RDMA 的主机虚拟 NIC。

---

### 注

以下步骤假设有四个 RDMA NIC 端口。

---

### 要在每个服务器上配置网络：

1. 请如下配置物理网络交换机：
  - a. 将所有适配器 NIC 连接到交换机端口。

---

### 注

如果测试适配器有多个 NIC 端口，您必须将两个端口连接到同一交换机。

---

- b. 启用交换机端口，确保交换机端口支持与交换机无关的组合模式，同时是多个 VLAN 网络的组成部分。

示例 Dell 交换机配置：

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

2. 启用 **Network Quality of Service**（网络服务质量）。

---

### 注

网络服务质量用于确保软件定义的存储系统有足够的带宽在节点之间通信，以确保弹性和性能。要在适配器上配置 QoS，请参阅[第 115 页上的“为 RoCE 配置 QoS”](#)。

---

3. 请如下创建带 SET 和 RDMA 虚拟 NIC 的 Hyper-V 虚拟交换机:
- a. 要标识网络适配器, 请发出以下命令:

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

- b. 要创建连接到所有物理网络适配器的虚拟交换机, 然后启用 Switch Embedded Teaming, 请发出以下命令:

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

- c. 要将主机虚拟 NIC 添加到虚拟交换机, 请发出以下命令:

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

---

### 注

上述命令从您刚刚为要使用的管理操作系统配置的虚拟交换机配置虚拟 NIC。

---

- d. 要配置主机虚拟 NIC 以使用 VLAN, 请发出以下命令:

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

---

### 注

这些命令可位于相同或不同的 VLAN 上。

---

- e. 要验证 VLAN ID 是否已设置, 请发出以下命令:

```
Get-VMNetworkAdapterVlan -ManagementOS
```

- f. 要禁用和启用每个主机虚拟 NIC 适配器以使 VLAN 处于活动状态, 请发出以下命令:

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

- g. 要在主机虚拟 NIC 适配器上启用 RDMA，请发出以下命令：
- ```
Enable-NetAdapterRdma "SMB1","SMB2"
```
- h. 要验证 RDMA 功能，请发出以下命令：
- ```
Get-SmbClientNetworkInterface | where RdmaCapable -EQ $true
```

## 配置 Storage Spaces Direct

在 Windows Server 2016 中配置 Storage Spaces Direct 包括以下步骤：

- [步骤 1. 运行群集验证工具](#)
- [步骤 2. 创建群集](#)
- [步骤 3. 配置群集见证](#)
- [步骤 4. 清理用于 Storage Spaces Direct 的磁盘](#)
- [步骤 5. 启用 Storage Spaces Direct](#)
- [步骤 6. 创建虚拟磁盘](#)
- [步骤 7. 创建或部署虚拟机](#)

### 步骤 1. 运行群集验证工具

运行群集验证工具以确保服务器节点正确配置，从而使用 Storage Spaces Direct 创建群集。

发出以下 Windows PowerShell 命令，验证用作 Storage Spaces Direct 群集的一组服务器：

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

### 步骤 2. 创建群集

创建 [步骤 1. 运行群集验证工具](#) 中具有四个节点的群集（群集创建已验证）。

要创建群集，请发出以下 Windows PowerShell 命令。

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

`-NoStorage` 参数是必需的。如果未包括该参数，磁盘将自动添加到群集，您必须移除磁盘，然后才能启用 Storage Spaces Direct。否则，磁盘将不包括在 Storage Spaces Direct 存储池中。



### 步骤 3. 配置群集见证

您应当配置群集的见证，使得该四节点系统承受两个节点发生故障或脱机。对于这些系统，您可配置文件共享见证或云见证。

有关详细信息，请访问：

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

### 步骤 4. 清理用于 Storage Spaces Direct 的磁盘

旨在用于 Storage Spaces Direct 的磁盘必须为空，且不带分区或其他数据。如果磁盘有分区或其他数据，该磁盘将不会包括在 Storage Spaces Direct 系统中。

可在 Windows PowerShell 脚本 (.PS1) 文件中放入以下 Windows PowerShell 命令，并从管理系统打开的 Windows PowerShell（或 Windows PowerShell ISE）控制台以管理员权限执行：

---

#### 注

运行此脚本可帮助识别每个节点上可用于 Storage Spaces Direct 的磁盘，并从这些磁盘移除所有数据和分区。

---

```
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
```

### 步骤 5. 启用 Storage Spaces Direct

创建群集后，发出 `Enable-ClusterStorageSpacesDirect` Windows PowerShell cmdlet。该 cmdlet 会将存储系统置于 Storage Spaces Direct 模式并自动执行以下操作：

- 创建一个名为如 *S2D on Cluster1* 的大型池。
- 配置 Storage Spaces Direct 高速缓存。如果有多种介质类型可供 Storage Spaces Direct 使用，它会配置最有效的类型作为高速缓存设备（在大多数情况下，读取和写入）。
- 创建两个层作为默认层：**Capacity**（容量）和 **Performance**（性能）。该 cmdlet 会分析设备并以混合的设备类型和弹性配置每个层。

### 步骤 6. 创建虚拟磁盘

如果 Storage Spaces Direct 已启用，它会使用所有磁盘创建一个池。它还会命名该池（例如 *S2D on Cluster1*），并在名称中指定群集名称。

以下 Windows PowerShell 命令在存储池上创建具有镜像和奇偶校验弹性的虚拟磁盘：

```
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>
```

### 步骤 7. 创建或部署虚拟机

您可以在超聚合 S2D 群集的节点上部署虚拟机。将虚拟机的文件存储在系统的 CSV 名称空间（例如，`c:\ClusterStorage\Volume1`）中，与故障转移群集上的群集虚拟机类似。

## 部署和管理 Nano Server

Windows Server 2016 提供 Nano Server 作为新的安装选项。Nano Server 是远程管理的服务器操作系统，优化用于私有云和数据中心。它类似于服务器核心模式下的 Windows Server，但明显更小，没有本地登录功能，并且仅支持 64 位应用程序、工具和代理。与 Windows Server 相比，Nano Server 占用较少的磁盘空间，设置更快，并且需要较少的更新和重新启动次数。重新启动时，它会重新启动得更快。

## 角色和功能

表 9-1 显示此版本 Nano Server 中可用的角色和功能，以及将为其安装软件包的 Windows PowerShell 选项。某些软件包直接通过自己的 Windows PowerShell 选项（例如 `-Compute`）进行安装。其他软件包作为 `-Packages` 选项的扩展进行安装，您可在逗号分隔的列表中进行组合。

**表 9-1. Nano Server 的角色和功能**

角色或功能	选项
Hyper-V 角色	<code>-Compute</code>
故障转移群集	<code>-Clustering</code>
用于将 Nano Server 托管为虚拟机的 Hyper-V 来宾驱动程序	<code>-GuestDrivers</code>
用于各种网络适配器和存储控制器的基本驱动程序。这与 Windows Server 2016 Technical Preview 服务器核心安装中包括的驱动程序集相同。	<code>-OEMDrivers</code>
文件服务器角色和其他存储组件	<code>-Storage</code>
Windows Defender 反恶意软件，包括默认的签名文件	<code>-Defender</code>
应用程序兼容性的反向转发器；例如，Ruby、Node.js 及其他常用应用程序框架。	<code>-ReverseForwarders</code>
DNS 服务器角色	<code>-Packages Microsoft-NanoServer-DNS-Package</code>
Desired State Configuration (DSC)	<code>-Packages Microsoft-NanoServer-DSC-Package</code>
Internet Information Server (IIS)	<code>-Packages Microsoft-NanoServer-IIS-Package</code>
Windows 容器的主机支持	<code>-Containers</code>

表 9-1. Nano Server 的角色和功能 (续)

角色或功能	选项
System Center Virtual Machine Manager Agent	<p>-Packages Microsoft-Windows-Server-SCVMM-Package</p> <p>-Packages Microsoft-Windows-Server-SCVMM-Compute-Package</p> <p>注：仅当您监测 Hyper-V 时使用此软件包。如果您安装此软件包，请勿对 Hyper-V 角色使用 -Compute 选项；而是使用 -Packages 选项来安装 -Packages Microsoft-NanoServer-Compute-Package、Microsoft-Windows-Server-SCVMM-Compute-Package。</p>
网络性能诊断服务 (NPDS)	<p>-Packages Microsoft-NanoServer-NPDS-Package</p>
数据中心桥接	<p>-Packages Microsoft-NanoServer-DCB-Package</p>

接下来的各节介绍如何使用所需的软件包配置 Nano Server 映像，以及如何添加特定于 QLogic 设备的附加设备驱动程序。它们还说明如何使用 Nano Server Recovery Console，如何远程管理 Nano Server，以及如何从 Nano Server 运行 Ntttcp 流量。

## 在物理服务器上部署 Nano Server

按照以下步骤操作，使用预安装的设备驱动程序创建将在物理服务器上运行的 Nano Server 虚拟硬盘 (VHD)。

### 要部署 Nano Server：

1. 下载 Windows Server 2016 操作系统映像。
2. 装载 ISO。
3. 将以下文件从 NanoServer 文件夹复制到您的硬盘驱动器上的文件夹：
  - NanoServerImageGenerator.psml
  - Convert-WindowsImage.ps1
4. 以管理员身份启动 Windows PowerShell。
5. 将目录更改为步骤 3 中您粘贴文件的文件夹。
6. 通过发出以下命令，导入 NanoServerImageGenerator 脚本：

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

7. 要创建设置计算机名称并包括 OEM 驱动程序和 Hyper-V 的 VHD，请发出以下 Windows PowerShell 命令：

### 注

此命令将提示您为新 VHD 输入管理员密码。

---

```
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>"
```

示例：

```
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"
```

在上例中，C:\Nano\Drivers 是 QLogic 驱动程序的路径。此命令需要大约 10 到 15 分钟来创建 VHD 文件。此命令的示例输出如下所示：

```
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. 在您要运行 Nano Server VHD 的物理服务器上以管理员身份登录。
9. 要将 VHD 复制到物理服务器并将其配置为从新 VHD 引导：
  - a. 转至 **Computer Management**（计算机管理）> **Storage**（存储）> **Disk Management**（磁盘管理）。
  - b. 右键单击 **Disk Management**（磁盘管理），然后选择 **Attach VHD**（附加 VHD）。
  - c. 提供 VHD 文件路径。
  - d. 单击 **OK**（确定）。
  - e. 运行 `bcdboot d:\windows`。

---

### 注

在本例中，VHD 附加在 D:\ 下。

---

- f. 右键单击 **Disk Management**（磁盘管理），然后选择 **Detach VHD**（分离 VHD）。
10. 将物理服务器重新引导到 Nano Server VHD 中。
11. 使用在 [步骤 7](#) 中运行脚本时您提供的管理员和密码登录到恢复控制台。
12. 获取 Nano Server 计算机的 IP 地址。
13. 使用 Windows PowerShell 远程处理（或其他远程管理）工具连接并远程管理服务器。

## 在虚拟机中部署 Nano Server

要创建 Nano Server 虚拟硬盘驱动器 (VHD) 以在虚拟机中运行：

1. 下载 Windows Server 2016 操作系统映像。
2. 转至 [步骤 1](#) 中下载的文件所在的 NanoServer 文件夹。

3. 将以下文件从 NanoServer 文件夹复制到您的硬盘驱动器上的文件夹：
  - ❑ NanoServerImageGenerator.psml
  - ❑ Convert-WindowsImage.ps1
4. 以管理员身份启动 Windows PowerShell。
5. 将目录更改为 [步骤 3](#) 中您粘贴文件的文件夹。
6. 通过发出以下命令，导入 NanoServerImageGenerator 脚本：  
**Import-Module .\NanoServerImageGenerator.psml -Verbose**
7. 发出以下 Windows PowerShell 命令，创建设置计算机名称并包括 Hyper-V 来宾驱动程序的 VHD：

---

### 注

以下命令将提示您针对新 VHD 的管理员密码。

---

```
New-NanoServerImage -DeploymentType Guest -Edition  
<Standard/Datacenter> -MediaPath <path to root of media>  
-BasePath  
.\Base -TargetPath .\NanoServerPhysical\NanoServer.vhd  
-ComputerName  
<computer name> -GuestDrivers
```

示例：

```
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
```

上述命令需要大约 10 到 15 分钟来创建 VHD 文件。此命令的示例输出如下：

```
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. 在 Hyper-V 管理器中创建新的虚拟机，并使用[步骤 7](#)中创建的 VHD。
9. 引导虚拟机。
10. 连接到 Hyper-V 管理器中的虚拟机。
11. 使用在[步骤 7](#)中运行脚本时您提供的管理员和密码登录到恢复控制台。
12. 获取 Nano Server 计算机的 IP 地址。
13. 使用 Windows PowerShell 远程处理（或其他远程管理）工具连接并远程管理服务器。



## 远程管理 Nano Server

远程管理 Nano Server 的选项包括 Windows PowerShell、Windows Management Instrumentation (WMI)、Windows 远程管理和紧急管理服务 (EMS)。本节介绍如何使用 Windows PowerShell 远程处理访问 Nano Server。

### 使用 Windows PowerShell 远程处理管理 Nano Server

**要使用 Windows PowerShell 远程处理管理 Nano Server：**

1. 将 Nano Server 的 IP 地址添加到管理计算机的受信任的主机列表。

---

#### 注

使用恢复控制台查找服务器 IP 地址。

---

2. 添加用于 Nano Server 管理员的帐户。
3. (可选) 如果适用, 启用 **CredSSP**。

### 将 Nano Server 添加到受信任的主机列表

在提升权限的 Windows PowerShell 提示符下, 通过发出以下命令, 将 Nano Server 添加到受信任的主机列表:

```
Set-Item WSMAN:\localhost\Client\TrustedHosts "<IP address of Nano Server>"
```

**示例:**

```
Set-Item WSMAN:\localhost\Client\TrustedHosts "172.28.41.152"  
Set-Item WSMAN:\localhost\Client\TrustedHosts "*"
```

---

#### 注

上述命令将所有主机服务器设置为受信任的主机。

---

### 启动远程 Windows PowerShell 会话

在提升权限的本地 Windows PowerShell 会话中, 通过发出以下命令, 启动远程 Windows PowerShell 会话:

```
$ip = "<IP address of Nano Server>"  
$user = "$ip\Administrator"  
Enter-PSSession -ComputerName $ip -Credential $user
```

现在，您可以在 Nano Server 上如常运行 Windows PowerShell 命令。不过，并非所有 Windows PowerShell 命令在此版本的 Nano Server 中均可用。要查看哪些命令可用，请发出命令 `Get-Command -CommandType Cmdlet`。要停止远程会话，请发出命令 `Exit-PSSession`。

有关 Nano Server 的更多详细信息，请访问：

<https://technet.microsoft.com/en-us/library/mt126167.aspx>

## 在 Windows Nano Server 上管理 QLogic 适配器

为了在 Nano Server 环境中管理 QLogic 适配器，请参阅 Windows QConvergeConsole GUI 和 Windows QLogic Control Suite CLI 管理工具及关联的说明文件，这些内容可从 Cavium 网站下载。

## RoCE 配置

**要使用 Windows PowerShell 远程处理管理 Nano Server：**

1. 从另一台计算机通过 Windows PowerShell 远程处理连接到 Nano Server。  
例如：

```
PS C:\Windows\system32> $ip="172.28.41.152"  
PS C:\Windows\system32> $user="172.28.41.152\Administrator"  
PS C:\Windows\system32> Enter-PSSession -ComputerName $ip  
-Credential $user
```

---

### 注

在上例中，Nano Server IP 地址为 172.28.41.152，用户名为 Administrator。

---

如果 Nano Server 成功连接，将返回以下内容：

```
[172.28.41.152]: PS C:\Users\Administrator\Documents>
```

2. 要确定驱动程序是否已安装和链路是否正常工作，请发出以下 Windows PowerShell 命令：

```
[172.28.41.152]: PS C:\Users\Administrator\Documents>  
Get-NetAdapter
```

图 9-18 显示示例输出。

```
[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

图 9-18. Windows PowerShell 命令: *Get-NetAdapter*

- 要验证在适配器上是否已启用 RDMA，请发出以下 Windows PowerShell 命令:

```
[172.28.41.152]: PS C:\Users\Administrator\Documents>  
Get-NetAdapterRdma
```

图 9-19 显示示例输出。

```
[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

图 9-19. Windows PowerShell 命令: *Get-NetAdapterRdma*

- 要将 IP 地址和 VLAN ID 分配至适配器的所有接口，请发出以下 Windows PowerShell 命令:

```
[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
```

- 要在 Nano Server 上创建 SMBShare，请发出以下 Windows PowerShell 命令:

```
[172.28.41.152]: PS C:\Users\Administrator\Documents>  
New-Item -Path c:\ -Type Directory -Name smbshare -Verbose
```

图 9-20 显示示例输出。

```
[172.28.41.152]: PS C:\Users\Administrator\Documents> New-Item -Path c:\ -Type Directory -Name smbshare -Verbose
VERBOSE: Performing the operation "Create Directory" on target "Destination: C:\smbshare".

Directory: C:\

Mode                LastWriteTime         Length Name
----                -
d-----          4/25/2016  1:34 AM             smbshare
```

图 9-20. Windows PowerShell 命令: New-Item

```
[172.28.41.152]: PS C:\> New-SMBShare -Name "smbshare" -Path c:\smbshare -FullAccess Everyone
```

图 9-21 显示示例输出。

```
[172.28.41.152]: PS C:\> New-SMBShare -Name "smbshare" -Path c:\smbshare -FullAccess Everyone

Name      ScopeName Path      Description
-----
smbshare *      c:\smbshare
```

图 9-21. Windows PowerShell 命令: New-SMBShare

6. 要将 SMBShare 映射为客户端计算机中的网络驱动器，请发出以下 Windows PowerShell 命令：

**注**

Nano Server 上接口的 IP 地址为 192.168.10.10。

```
PS C:\Windows\system32> net use z: \\192.168.10.10\smbshare
This command completed successfully.
```

7. 要在 SMBShare 上执行读取 / 写入操作，并检查 Nano Sever 上的 RDMA 统计信息，请发出以下 Windows PowerShell 命令：

```
[172.28.41.152]: PS C:\>
(Get-NetAdapterStatistics).RdmaStatistics
```

图 9-22 显示命令输出。

```
[172.28.41.152]: PS C:\> (Get-NetAdapterStatistics).RdmaStatistics  
AcceptedConnections      : 2  
ActiveConnections       : 2  
CompletionQueueErrors   : 0  
ConnectionErrors        : 0  
FailedConnectionAttempts : 0  
InboundBytes             : 403913290  
InboundFrames            : 4110373  
InitiatedConnections     : 0  
OutboundBytes            : 63902433706  
OutboundFrames           : 58728133  
PSComputerName           :
```

图 9-22. Windows PowerShell 命令: *Get-NetAdapterStatistics*

# 10 故障排除

本章提供了以下故障排除信息：

- [故障排除核查表](#)
- [第 151 页上的“验证是否已加载最新驱动程序”](#)
- [第 152 页上的“测试网络连接”](#)
- [第 153 页上的“使用 Hyper-V 的 Microsoft Virtualization”](#)
- [第 153 页上的“Linux 特定问题”](#)
- [第 153 页上的“其他问题”](#)
- [第 154 页上的“收集调试数据”](#)

## 故障排除核查表

### 小心

在打开服务器机箱以添加或拆卸适配器之前，请先查阅[第 6 页上的“安全预防措施”](#)。

以下核查表提供了一些建议的操作，以解决在系统中安装或运行 QL45212 Adapter 时可能遇到的问题。

- 检查所有电缆和连接。验证网络适配器和交换机上的电缆正确连接。
- 对照[第 7 页上的“安装适配器”](#)验证适配器安装。确保适配器正确插入插槽中。检查是否有特定的硬件问题，如插卡组件或 PCI 边缘连接器明显损坏。
- 验证配置设置，如果与其他设备冲突，则进行更改。
- 验证服务器在使用最新的 BIOS。
- 尝试将适配器插入另一插槽。如果在新位置没有问题，则系统中的原插槽可能有缺陷。
- 用已知工作正常的适配器替换故障适配器。如果第二个适配器在第一个适配器无法运行的插槽中可运行，则原适配器可能有缺陷。

- 将适配器安装在另一台运行正常的系统中，然后再次运行测试。如果适配器在新系统中通过测试，则原系统可能有缺陷。
- 从该系统卸下所有其他适配器，然后再次运行测试。如果适配器通过测试，则其他适配器可能导致了争用。

## 验证是否已加载最新驱动程序

确保已为您的 Windows、Linux 或 VMware 系统加载最新驱动程序。

### 在 Windows 中验证驱动程序

请参阅设备管理器，查看有关适配器、链路状态和网络连接的重要信息。

### 在 Linux 中验证驱动程序

要验证 qed.ko 驱动程序是否已正确加载，请发出以下命令：

```
# lsmod | grep -i <module name>
```

如果驱动程序已加载，此命令的输出将显示驱动程序大小（以字节为单位）。以下示例显示为 qed 模块加载的驱动程序：

```
# lsmod | grep -i qed
qed                199238  1
qede               1417947  0
```

如果加载新的驱动程序后重新引导，可以发出以下命令验证当前加载的驱动程序版本是否正确：

```
modinfo qede
```

也可以发出以下命令：

```
[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
```

如果已加载新驱动程序，但尚未重新引导，则 `modinfo` 命令不会显示更新的驱动程序信息。相反地，发出以下 `dmesg` 命令可查看日志。在此示例中，最后一个条目用于标识将在重新引导时激活的驱动程序。

```
# dmesg | grep -i "Cavium" | 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
```

## 在 VMware 中验证驱动程序

要验证 VMware ESXi 驱动程序已加载，请发出以下命令：

```
# esxcli software vib list
```

## 测试网络连接

本节提供在 Windows 和 Linux 环境中测试网络连接的步骤。

---

### 注

在使用强制链路速度时，验证适配器和交换机均被强制为同一速度。

---

## 测试 Windows 的网络连接

使用 `ping` 命令测试网络连接。

要确定网络连接是否工作：

1. 单击 **Start**（开始），然后单击 **Run**（运行）。
2. 在 **Open**（打开）框中，键入 `cmd`，然后单击 **OK**（确定）。
3. 要查看被测试的网络连接，发出以下命令：

```
ipconfig /all
```

4. 发出以下命令，然后按 ENTER。

```
ping <ip_address>
```

显示的 ping 统计信息指示网络连接是否在工作。



## 测试 Linux 的网络连接

**要验证以太网接口正常工作并运行：**

1. 要检查以太网接口的状态，请发出 `ifconfig` 命令。
2. 要检查以太网接口的统计信息，请发出 `netstat -i` 命令。

**要验证连接是否已建立：**

1. 对网络上的 IP 主机执行 Ping 操作。从命令行，发出以下命令：

```
ping <ip_address>
```

2. 按 ENTER 键。

显示的 ping 统计信息指示网络连接是否在工作。

使用操作系统 GUI 工具或 `ethtool` 命令 `ethtool -s ethX speed SSSS` 可以将适配器链路速度强制为 10 Gbps 或 25 Gbps。将 QL45212 Adapter 上的两个端口设置为相同的速度（10Gbps 或 25Gbps）。

## 使用 Hyper-V 的 Microsoft Virtualization

Microsoft Virtualization 是适用于 Windows Server 2012 R2 的虚拟机监控程序虚拟化系统。有关 Hyper-V 的更多信息，请访问：

<https://technet.microsoft.com/en-us/library/Dn282278.aspx>

## Linux 特定问题

**问题：** 汇编驱动程序源码时出错。

**解决办法：** Linux 分发版的有些安装没有默认安装开发工具和内核源。汇编驱动程序源码之前，确保所使用的 Linux 分发版的开发工具已安装。

## 其他问题

**问题：** QL45212 Adapter 已关闭并出现错误消息，提示适配器上的风扇发生故障。

**解决办法：** 可人为关闭 QL45212 Adapter，以防止永久损坏。联系 Cavium 技术支持以获取帮助。

## 收集调试数据

使用 [表 10-1](#) 中的命令收集调试数据。

**表 10-1. 收集调试数据命令**

调试数据	说明
dmesg-T	内核日志
ethtool-d	寄存器转储
sys_info.sh	系统信息；在驱动程序包中提供

# A 适配器 LED

表 A-1 列出适配器端口链路和活动状态的 LED 指示灯。

**表 A-1. 适配器端口链路和活动 LED**

端口 LED	LED 外观	网络状态
链路 LED	不亮 持续发亮	无链接（电缆断开） 链接
活动 LED	不亮 闪烁	无端口活动 端口活动

# B 电缆和光学模块

本附录提供关于所支持的电缆和光学模块的以下信息：

- 支持的规格
- 第 156 页上的“测试的电缆和光学模块”
- 第 158 页上的“经测试的交换机”

## 支持的规格

QL45212 Adapters 支持符合 SFF8024 的各种电缆和光学模块。具体的形状因数符合性如下：

- SFP:
  - SFF8472（用于内存映射）
  - SFF8419 或 SFF8431（低速信号和电源）
- 四通道小型可插拔接口 (QSFP):
  - SFF8636（用于内存映射）
  - SFF8679 或 SFF8436（低速信号和电源）
- 光学模块电气输入 / 输出、有源铜缆 (ACC) 和有源光缆 (AOC):
  - 10G - SFF8431 限制接口
  - 25G - IEEE802.3by Annex 109B (25GAUI)

## 测试的电缆和光学模块

Cavium 不保证满足合规要求的每个电缆或光学模块均可与 QL45212 Adapters 一起使用。QL45212 适配器支持 3m DAC 和 30m AOC，但不支持 25G SR 光纤。Cavium 已经测试了 QL45212 适配器与 [表 B-1](#) 所列电缆的互操作性。

表 B-1. 测试的电缆和光学模块

速度 / 形状因素	生产商	部件号	类型	电缆长度 <sup>a</sup>	标准尺寸
10G DAC <sup>b</sup>	Dell	407-BBBK	SFP 双轴 10G	1	30
		407-BBBI	SFP 双轴 10G	3	26
	Cisco	COPQAA4JAA	SFP 双轴 10G	1	30
		COPQAA6JAA	SFP 双轴 10G	3	26
	HP®	2074260-2	SFP 双轴 10G	1	26
		AP784A	SFP 双轴 10G	3	26
25G DAC	Amphenol®		SFP28 至 SFP28	1	30
		NDCCGF0003	SFP28 至 SFP28	3	30
		NDCCGJ0003	SFP28 至 SFP28	3	26
	HP		SFP28 至 SFP28	1	26
		844477-B21	SFP28 至 SFP28	3	26
<b>光学解决方案</b>					
10G AOC <sup>c</sup>	Dell	470-ABLV	SFP 10G AOC	2	N/A
		470-ABLZ	SFP 10G AOC	3	N/A
		470-ABLT	SFP 10G AOC	5	N/A
		470-ABML	SFP 10G AOC	7	N/A
		470-ABLU	SFP 10G AOC	10	N/A
		470-ABMD	SFP 10G AOC	15	N/A
		470-ABMJ	SFP 10G AOC	15	N/A
25G AOC	InnoLight		SFP28 AOC	3	N/A
		TF-PY020-N00	SFP28 AOC	20	N/A

<sup>a</sup> 电缆长度以米为单位。

<sup>b</sup> DAC 为直连电缆。

<sup>c</sup> AOC 为有源光缆。

## 经测试的交换机

表 B-2 列出经测试与 QL45212 Adapters 的互操作性的交换机。此列表是基于产品发布时的交换机的，并随着新交换机进入市场或停产而随时间变化。

**表 B-2. 经测试互操作性的交换机**

制造商	以太网交换机型号
Arista	
	7160
Cisco	Nexus 3132
	Nexus 3232C
	Nexus 5548
	Nexus 5596T
	Nexus 6000
Dell EMC	
	Z9100
HPE	FlexFabric 5950
Mellanox®	
	SN2700

# C Dell Z9100 交换机配置

QL45212 Adapters 支持与 Dell Z9100 以太网交换机的连接。但是，在标准化自动协商过程之前，必须将交换机明确地配置为按 25 Gbps 速率连接到适配器。

## 要配置 Dell Z9100 交换机端口以便按 25 Gbps 速率连接 QL45212 Adapter：

1. 在您的管理工作站与交换机之间建立串行端口连接。
2. 打开命令行会话，然后如下登录到交换机：

```
Login: admin
Password: admin
```

3. 启用交换机端口的配置：

```
Dell> enable
Password: xxxxxxx
Dell# config
```

4. 标识要配置的模块和端口。以下示例使用模块 1，端口 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
```

有关更改适配器链路速度的信息，请参阅[第 152 页上的“测试网络连接”](#)。

5. 验证端口的运行速率是否为 25 Gbps：

```
Dell# Dell#show running-config | grep "port 5"
stack-unit 1 port 5 portmode quad speed 25G
```

6. 要在交换机端口 5 上禁用自动协商，请按照以下步骤操作：

- a. 标识交换机端口接口（模块 1，端口 5，接口 1），并确认自动协商状态：

```
Dell(conf)#interface tw 1/5/1

Dell(conf-if-tf-1/5/1)#intf-type cr4 ?
autoneg                               Enable autoneg
```

- b. 禁用自动协商：

```
Dell(conf-if-tf-1/5/1)#no intf-type cr4 autoneg
```

- c. 验证是否已禁用自动协商。

```
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
```

有关配置 Dell Z9100 交换机的更多信息，请参阅 Dell 支持网站上的 *Dell Z9100 Switch Configuration Guide*（Dell Z9100 交换机配置指南），网址为：

[support.dell.com](http://support.dell.com)



# 词汇表

## ACPI

*高级配置和电源接口 (ACPI)* 规格提供了统一的、以操作系统为中心的设备配置和电源管理的开放标准。ACPI 定义了与平台无关的接口，用于硬件查找、配置、电源管理和监测。该规格以操作系统导向的配置和电源管理（OSPM，一个用于描述实施 ACPI 的系统的术语）为中心，从而移除了旧版固件接口的设备管理责任。

## 适配器

用于联接主机系统与目标设备的板卡。适配器的其他表示形式包括主机总线适配器、主机适配器和板。

## 适配器端口

适配器板上的端口。

## 高级配置和电源接口

请参阅 [ACPI](#)。

## 交替路由 ID 解析

请参阅 [ARI](#)。

## ARI

交替路由 ID 解析。PCI-SIG ARI 提供了一种机制，允许单个 PCIe 设备支持最多 256 个物理功能 (PF)。如果在系统 BIOS 中启用了 ARI，则当前操作系统允许使用最多 16 个 PF（每设备），如果禁用了 ARI，则允许使用最多 8 个 PF（每设备）。

## 带宽

以特定传输速率可传输的数据量的度量。1Gbps 或 2Gbps 光纤信道端口可传输或接收的标称速率为 1Gbps 或 2Gbps，具体视所连接的设备而定。这分别对应于实际带宽值 106MB 和 212MB。

## BAR

基本地址寄存器。用于保留设备所使用的内存地址，或端口地址的偏移。通常，内存地址 BAR 必须位于物理 RAM 中，而 I/O 空间 BAR 可以驻留在任意内存地址（即使超出物理内存范围）。

## 基本地址寄存器

请参阅 [BAR](#)。

## 基本输入输出系统

请参阅 [BIOS](#)。

## BIOS

基本输入输出系统。通常位于 Flash PROM 中，用于充当硬件和操作系统间接口的程序（或公用程序），并且允许在启动时从适配器引导。

## 数据中心桥接

请参阅 [DCB](#)。

## 数据中心桥接交换

请参阅 [DCBX](#)。

## DCB

数据中心桥接。提供对现有 802.1 桥接规范的增强，以满足数据中心内的协议和应用程序的需求。由于现有高性能数据中心通常包含在不同链路层技术上运行的多个应用特定网络（光纤信道用于存储，以太网用于网络管理和 LAN 连接），DCB 允许使用 802.1 桥接来部署聚合网络，从而使所有应用可以在单个物理基础结构上运行。

## DCBX

数据中心桥接交换。DCB 由设备用来与直连对等方交换配置信息的协议。该协议也可用于误配置检测和对等方配置。

## 设备

**目标**，通常为磁盘驱动器。安装或连接到系统的磁盘驱动器、磁带驱动器、打印机或键盘等硬件。在光纤信道中，为**目标设备**。

## DHCP

动态主机配置协议。仅当请求后，允许 IP 网络上的计算机从具有该计算机相关信息的服务器提取其配置。

## 驱动程序

用于充当文件系统与物理数据存储设备或网络介质之间的接口的软件。

## 动态主机配置协议

请参阅 [DHCP](#)。

## eCore

操作系统与硬件和固件之间的层。它是设备特定的，与操作系统无关。当 eCore 代码需要操作系统服务时（例如用于内存分配、PCI 配置空间访问等），它将调用操作系统特定层中实施的抽象操作系统功能。eCore 流可以由硬件（例如通过中断）或驱动程序的操作系统的特定分区（例如加载与卸载程序的加载和卸载）驱动。

## EEE

节能以太网。计算机联网标准的双绞线和背板以太网家族的一套增强功能允许在低数据活动时间段期间的降低的功耗。其目标是将功耗降低 50% 或更多，同时保持与现有设备的完全兼容。电气和电子工程师协会 (IEEE) 通过 IEEE802.3az 工作组而开发了该标准。

## EFI

可扩展固件接口。此规范定义操作系统与平台固件之间的软件接口。EFI 取代了在所有 IBM PC 兼容个人计算机中提供的旧版 BIOS 固件接口。

## 节能以太网

请参阅 [EEE](#)。

## 增强的传输选择

请参阅 [ETS](#)。

## 以太网

最为广泛使用的 LAN 技术，用于在计算机之间发送信息，通常速度为每秒 10 兆比特和 100 兆比特 (Mbps)。

## ETS

增强的传输选择。用于指定传输选择增强的标准，以支持流量类型之间的带宽分配。当某个流量类型中提供的负载不为其分配的带宽时，增强的传输选择允许其他流量类型使用可用带宽。带宽分配优先级与严格优先级共存。ETS 包含管理对象以支持带宽分配。有关更多信息，请参阅：  
<http://ieee802.org/1/pages/802.1az.html>

## 可扩展固件接口

请参阅 [EFI](#)。

## FCoE

以太网光纤信道。一种由 T11 标准机构定义的新技术，通过在第二层以太网帧内封装光纤信道帧，允许传统的光纤信道存储网络流量在以太网链路上传输。有关更多信息，请访问 [www.fcoe.com](http://www.fcoe.com)。

## 以太网光纤信道

请参阅 [FCoE](#)。

## 文件传输协议

请参阅 [FTP](#)。

## FTP

文件传输协议。一种标准网络协议，用于将文件通过基于 TCP 的网络（如互联网）从一台主机传送到另一台主机。进行带外固件上传（比带内固件上传速度更快）时需使用 FTP。

## 人机界面基础设施

请参阅 [HII](#)。

## HII

人机界面基础设施。此规范（UEFI 2.1 的一部分）用于管理用户输入、本地化字符串、字体和窗体，允许 OEM 开发用于预引导配置的图形界面。

## IEEE

电气和电子工程师协会。一个旨在促进电气相关技术进步的国际化非营利组织。

## 互联网协议

请参阅 [IP](#)。

## 互联网小型计算机系统接口

请参阅 [iSCSI](#)。

## IP

互联网协议。一种通过互联网将数据从一台计算机发送到另一台计算机的方法。IP 指定数据包的格式（也称为 *数据报*）和寻址方案。

## IQN

iSCSI 限定名称。iSCSI 节点名称基于启动器制造商和唯一的设备名称部分。

## iSCSI

互联网小型计算机系统接口。一种将数据封装到 IP 数据包以通过以太网连接来发送的协议。

## iSCSI 限定名称

请参阅 [IQN](#)。

## 巨型帧

在高性能网络中使用以提高长距离性能的大型 IP 帧。对于千兆位以太网，巨型帧通常意味着 9,000 字节，但可指超出 IP MTU（在以太网上为 1,500 字节）的任意大小。

## 大量发送卸载

请参阅 [LSO](#)。

## 第二层

指多层通信模型开放系统互连 (OSI) 的数据链路层。数据链路层的功能是跨网络中的物理链路移动数据，其中交换机使用目标 MAC 地址在第二层级别重定向数据消息以确定消息目标。

## 链路层发现协议

请参阅 [LLDP](#)。

## LLDP

供应商中立二层协议允许网络设备在本地网络上通告其身份和功能。该协议取代专有协议，如 Cisco 发现协议、Extreme 发现协议、和 Nortel 发现协议（也被称为 SONMP）。

利用 LLDP 收集的信息被存储在设备中并可使用 SNMP 而被查询。基于 LLDP 的网络的拓扑可通过爬取主机并查询该数据库而被发现。

## LSO

大量发送卸载。LSO 以太网适配器功能，允许 TCP/IP 网络堆栈构建大型（最多 64KB）TCP 消息，然后将其发送给适配器。适配器硬件将消息分段为可通过线路发送的较小数据包（帧）：对于标准以太网帧最多为 1,500 字节，对于巨型以太网帧最多为 9,000 字节。分段过程可释放服务器 CPU，从而不必将大型 TCP 消息分段为装入所支持帧大小的较小数据包。

## 最大传输单元

请参阅 [MTU](#)。

## 消息信号中断

请参阅 [MSI](#)，[MSI-X](#)。

## MSI，MSI-X

消息信号中断。两个由 PCI 定义的扩展之一，用于支持 PCI 2.2 及更高版本和 PCI Express 中的消息信号中断 (MSI)。MSI 是另一种通过特殊消息来生成中断的方式，可模拟引脚有效或无效置位。

MSI-X（在 PCI 3.0 中定义）允许设备分配介于 1 到 2,048 之间任意数量的中断，并为每个中断提供单独的数据和地址寄存器。MSI 中可选的功能（64 位寻址和中断屏蔽）对 MSI-X 为强制的。

## MTU

最大传输单元。是指指定通信层协议可以传输的最大数据包（IP 数据报）大小（以字节为单位）。

## 网络接口卡

请参阅 [NIC](#)。

## NIC

网络接口卡。安装以启用专用网络连接的计算机卡。

## NIC 分区

请参阅 [NPAR](#)。

## 非易失性随机存取存储器

请参阅 [NVRAM](#)。

## NPAR

**NIC 分区**。将一个 NIC 端口分成多个物理功能或分区，每个均有用户可配置的带宽和个性设置（接口类型）。个性设置包括 [NIC](#)、[FCoE](#) 和 [iSCSI](#)。

## NVRAM

非易失性随机存取存储器。一种在即使断电时仍可保留数据（配置设置）的存储器。您可以手动配置或从文件还原 NVRAM 设置。

## OFED™

OpenFabrics 企业发行。RDMA 和内核旁路应用程序的开源软件。

## PCI™

外围组件接口。一种由 Intel® 引入的 32 位本地总线规格。

## PCI Express (PCIe)

第三代 I/O 标准，除支持旧式外围组件互连 (PCI) 和 PCI 扩展 (PCI-X) 台式机和服务器的插槽外，还支持增强型以太网网络性能。

## QoS

服务质量。在通过虚拟端口传输数据时，此方法用于通过设置优先级和分配带宽来预防瓶颈并确保业务连续性。

## 服务质量

请参阅 [QoS](#)。

## PF

物理功能。

## RDMA

远程直接内存访问。此功能允许一个节点通过网络直接写入另一个节点的内存（使用地址和大小语义）。此功能是 VI 网络的一项重要功能。

## 精简指令集计算机

请参阅 [RISC](#)。

## 远程直接内存访问

请参阅 [RDMA](#)。

## RISC

精简指令集计算机。一种计算机微处理器，它执行较少类型的计算机指令，从而以更高的速度运行。

## 基于聚合以太网的 RDMA

请参阅 [RoCE](#)。

## RoCE

基于聚合以太网的 RDMA。一种网络协议，允许通过聚合或非聚合以太网进行远程直接内存访问 (RDMA)。RoCE 是链路层协议，允许位于同一以太网广播域中的任意两个主机之间进行通信。

## SCSI

小型计算机系统接口。一种用于将硬盘驱动器、CD 驱动器、打印机和扫描仪等设备连接到计算机的高速接口。SCSI 可使用一个控制器连接许多设备。每个设备均可通过 SCSI 控制器总线上单独的标识号进行访问。

## SerDes

串行化器 / 并行化器。一对功能块，通常在高速通信中用于弥补受限的输入 / 输出。这些功能块在串行数据接口和并行接口之间沿每个方向转换数据。

## 串行化器 / 并行化器

请参阅 [SerDes](#)。

## 单根输入 / 输出虚拟化

请参阅 [SR-IOV](#)。

## 小型计算机系统接口

请参阅 [SCSI](#)。

## SR-IOV

单根输入 / 输出虚拟化。一种 PCI SIG 的规格，使单个 PCIe 设备能够显示为多个单独的物理 PCIe 设备。SR-IOV 允许隔离 PCIe 资源以实现性能、互操作性和可管理性。

## 目标

SCSI 会话的存储设备端点。启动器从目标请求数据。目标通常为磁盘驱动器、磁带驱动器或其他媒体设备。通常，SCSI 外围设备是目标，但某些情况下，适配器也可能是目标。目标可包含许多 LUN。

目标是响应启动器（主机系统）请求的设备。外围设备是目标，但对于某些命令（例如，SCSI COPY 命令）来说，外围设备可充当启动器。

## TCP

传输控制协议。一组通过互联网协议发送数据包中数据的规则。

## TCP/IP

传输控制协议 / 互联网协议。互联网的基本通信语言。

## TLV

类型长度值。可编码为协议内元素的可选信息。类型和长度字段为固定大小（通常 1-4 字节），而值字段为可变大小。这些字段用法如下：

- 类型 - 表示此部分消息所代表字段类型的数字代码。
- 长度 - 值字段的大小（通常以字节为单位）。

- 值 - 一组包含此部分消息数据的、可变大小的字节。

### 传输控制协议

请参阅 [TCP](#)。

### 传输控制协议 / 互联网协议

请参阅 [TCP/IP](#)。

### 类型长度值

请参阅 [TLV](#)。

### UDP

用户数据报协议。一种不保证数据包顺序或传递的无连接传输协议。它直接在 IP 之上工作。

### UEFI

统一可扩展固件接口。此规范详细说明了一个接口，可帮助将预引导环境（即从系统开机之后到操作系统启动之前）的系统控制移交给操作系统（如 Windows 或 Linux）。在引导时，UEFI 提供操作系统与平台固件之间的干净接口，并支持用于初始化添加式卡的独立于体系结构的机制。

### 统一可扩展固件接口

请参阅 [UEFI](#)。

### 用户数据报协议

请参阅 [UDP](#)。

### VF

虚拟功能。

### VI

虚拟接口。用于跨光纤信道和其他通信协议进行远程直接内存访问的方案。用于群集和消息传递。

### 虚拟接口

请参阅 [VI](#)。

### 虚拟逻辑区域网络

请参阅 [VLAN](#)。

### 虚拟机

请参阅 [VM](#)。

### VLAN

虚拟逻辑区域网络 (LAN)。具有一组通用要求的主机组，如同连接到同一线路一样通信，无论其物理位置如何。尽管 VLAN 具有与物理 LAN 相同的属性，但它允许终端站组合在一起，即使其不在同一 LAN 网段上。VLAN 能够通过软件重新配置网络，而不必物理重新定位设备。

### VM

虚拟机。机器（计算机）的软件实现，如同真实机器一样执行程序。

### LAN 唤醒

请参阅 [WoL](#)。

### WoL

LAN 唤醒。一种以太网计算机联网标准，允许通过发送的网络消息远程开启或唤醒计算机，这些消息通常由在网络中另一台计算机上执行的简单程序发送。





**公司总部** Cavium, Inc. 2315 N. First Street San Jose, CA 95131 408-943-7100

**国际办事处** 英国 | 爱尔兰 | 德国 | 法国 | 印度 | 日本 | 中国 | 中国香港 | 新加坡 | 中国台湾 | 以色列

---

版权所有 © 2015–2018 Cavium, Inc. 在世界范围内保留所有权利。Qlogic 公司是 Cavium 公司的全资子公司。Cavium、FastLinQ、QConvergeConsole、QLogic 和 SmartAN 是 Cavium 公司的注册商标。所有其他品牌和产品名称均为其各自所有者的商标或注册商标。

本文档提供的信息仅供参考，并且可能包含错误。Cavium 保留在不另行通知的情况下对本文档或者产品设计或规格进行更改的权利。Cavium 不提供任何明示或暗示的任何担保，不保证您可获得本文档中描述的任何结果或性能。所有关于 Cavium 未来发展方向和意图的声明，均有可能在不另行通知的情况下进行更改或者取消，且仅表示目标与目的。

