用户手册

NetXtreme-E

Broadcom® NetXreme-C 和 NetXtreme-E

用户手册

NetXtreme-E-UG100

2018 年 2 月 26 日

修订历史

修订版	*日期*	*修改说明*
NetXtreme-E-UG100	2018 年 2 月 26 日	20.6 初始发行。

目录

法规和安全许可 7
法规 7
安全 7
电磁兼容性 (EMC) 8
静电抗扰性 (ESD) 合规 8
FCC 声明 8
功能描述 9
网络链路和活动指示 14
BCM957402AXXXX/BCM957412AXXXX 14
BCM957404AXXXX/BCM957414AXXXX 15
BCM957406AXXXX/BCM957416AXXXX 16
BCM957414M4140D 17
BCM957412M4120D 18
BCM957416M4160 19
特性 20
软件和硬件特性 20
虚拟化特性 21
VXLAN 22
NVGRE/GRE/IP-in-IP/Geneve 22
无状态卸载 22
RSS 22
TPA 22
标头 - 承载数据分离 22
UDP 分片卸载 22
无状态传输隧道卸载 22
对操作系统的多队列支持 23
NDIS VMQ 23
VMWare NetQueue 23
KVM/Xen 多队列 23
SR-IOV 配置支持矩阵 23
SR-IOV 23
网络分区 (NPAR) 24
融合以太网上的 RDMA - RoCE 24
支持的组合 24
NPAR、 SR-IOV 和 RoCE 25
NPAR、 SR-IOV 和 DPDK 25
不支持的组合 25

安装硬件 26
安全预防措施 26
系统要求 26
硬件要求 26
预安装检查表 26
安装适配器 27
连接网络电缆 27
支持的电缆和模块 27
铜缆 28
SFP+ 28
SFP28 28
软件包和安装 29
支持的操作系统 29
安装驱动程序 29
Windows 29
Dell DUP 29
GUI 安装 29
静默安装 29
INF 安装 29
Linux 30
模块安装 30
Linux Ethtool 命令 31
VMware 32
固件更新 33
Dell Update Package 33
Windows 33
Linux 33
Windows 驱动程序高级属性和事件日志消息 34
驱动程序高级属性 34
事件日志消息 35
组合 37
Windows 37
Linux 37
系统级配置 38
UEFI HII 菜单 38
主配置页面 38
固件图像属性 38
设备级别配置 38
NIC 配置 38

iSCSI 配置 38
Comprehensive Configuration Management 39
设备硬件配置 39
MBA 配置菜单 39
iSCSI 引导主菜单 39
自动协商配置 40
操作链路速度 43
固件链路速度 43
自动协商协议 43
Windows 驱动程序设置 43
Linux 驱动程序设置 43
ESXi 驱动程序设置 44
FEC 自动协商 45
链路训练 46
介质自动检测 47
ISCSI 引导 49
适用于 iSCSI 引导的支持的操作系统 49
设置 iSCSI 引导 49
配置 iSCSI 目标 49
配置 iSCSI 引导参数 50
MBA 引导协议配置 50
iSCSI 引导配置 51
静态 iSCSI 引导配置 51
动态 iSCSI 引导配置 52
启用 CHAP 身份验证 53
配置 DHCP 服务器以支持 iSCSI 引导 54
IPv4 的 DHCP iSCSI 引导配置 54
DHCP 选项 17，根路径 54
DHCP 选项 43，供应商特定信息 54
配置 DHCP 服务器 55
IPv6 的 DHCP iSCSI 引导配置 55
DHCPv6 选项 16， Vendor Class 选项 55
DHCPv6 选项 17，供应商特定信息 55
配置 DHCP 服务器 56
VXLAN：配置和使用案例示例 56
SR-IOV：配置和使用案例示例 57
Linux 使用案例 57
Windows 情况下 58
VMWare SRIOV 情况下 59

NPAR – 配置和使用案例示例 62
功能和要求 62
限制 62
配置 62
降低 NIC 内存消耗注意事项 65
RoCE - 配置和使用案例示例 66
Linux 配置 66
要求 66
BNXT_RE 驱动程序依存关系 66
安装 67
限制 67
已知问题 67
Windows 68
内核模式 68
验证 RDMA 68
用户模式 69
VMware ESX 70
限制 70
BNXT RoCE 驱动程序要求 70
安装 70
配置 Paravirtualized RDMA 网络适配器 71
为 PVRDMA 配置虚拟中心 71
为 ESX 主机上的 PVRDMA 标记 vmknic 71
为 PVRDMA 设置防火墙规则 71
将 PVRDMA 设备添加到 VM 71
在 Linux 客户机操作系统上配置 VM 72
DCBX - 数据中心桥接 73
QoS 配置文件 – 默认 QoS 列队配置文件 73
DCBX 模式 = 启用 （仅 IEEE） 74
DCBX Willing 位 74
常见问题 77

法规和安全许可

以下各节详细介绍 NetXtreme-E 网卡所遵循的法规、安全性、电磁兼容性 (EMC) 和静电抗扰性 (ESD) 标准。

法规

表 1：法规许可

项目	*适用标准*	*许可/认证*
CE/欧盟	EN 62368-1:2014	CB 报告和证书
UL/美国	IEC 62368-1 ed. 2	CB 报告和证书
CSA/加拿大	CSA 22.2 第 950 号	CSA 报告和证书
台湾	CNS14336 B 类	–

安全

表 2：安全许可

国家或地区	*认证类型/标准*	*合规*
国际	CB 体系 ICES 003 – 数字设备 UL 1977（连接器安全） UL 796（PCB 接线安全） UL 94（部件可燃性）	是

电磁兼容性 (EMC)

表 3：电磁兼容性

标准/国家或地区	*认证类型*	*合规*
CE/欧盟	EN 55032:2012/AC:2013 B 类 EN 55024:2010 EN 61000-3-2:2014 EN 61000-3-3:2013	CE 报告和 CE DoC
FCC/美国	CFR47，第 15 部分 B 类	FCC/IC DoC 和 EMC 报告（参照 FCC 和 IC 标准）
IC/加拿大	ICES-003 B 类	FCC/IC DoC 和报告（参照 FCC 和 IC 标准）
ACA/澳大利亚、新西兰	AS/NZS CISPR 22:2009 +A1:2010 / AS/NZS CISPR 32:2015	ACA 证书 RCM 标志
BSMI/台湾	CNS13438 B 类	BSMI 证书
BSMI/台湾	CNS15663	BSMI 证书
MIC/S.Korea	KN32 B 类 KN35	韩国证书 MSIP 标志
VCCI/日本	V-3/2014/04（2015 年 3 月 31 日生效）	VCCI 在线证书的副本

静电抗扰性 (ESD) 合规

表 4：ESD 合规性摘要

标准	*认证类型*	*合规*
EN55024:2010 (EN 61000-4-2)	空气/直接放电	是

FCC 声明

此设备经过测试，符合 FCC 规则第 15 部分规定的 B 类数字设备限制。这些限制旨在提供合理的保护，防止在住宅区安装时产生有害干扰。此设备会产生、使用并可能放射无线电频率能量，如果未按照说明安装和使用，则可能导致严重干扰无线电通信。但是，并不能保证在特殊安装时不会发生干扰。如果该设备确实对射频电视接收造成有害干扰（可以通过关闭并打开设备来确定），建议用户尝试采取以下一种或多种措施来消除干扰：

- 调整接收天线的方向或位置。
- 增大该设备与接收器之间的距离。
- 向经销商或有经验的无线电/电视技术人员咨询以获得帮助。

注：未经负责合规性的制造方明确准许的更改或修改可能造成用户操作该设备的权力失效。

功能描述

Dell 支持 10GBase-T、10G SFP+ 和 25G SFP28 网卡 (NIC)。表 5 中描述了这些 NIC。

表 5：功能描述

网卡	*描述*
BCM957402A4020DLPC/BCM957402A4020DC/BCM957412A4120D/BCM957412M4120D	
速度	双端口 10 Gbps 以太网
PCI-E	第 3 代 x8[a]
接口	SFP+ 支持 10 Gbps
设备	Broadcom BCM57402/BCM57412 10 Gbps MAC 控制器，集成了双通道 10 Gbps SFI 收发器。
NDIS 名称	Broadcom NetXtreme E 系列双端口 10Gb SFP+ 以太网 PCIe 适配器
UEFI 名称	Broadcom 双 10Gb SFP+ 以太网
BCM57404A4041DLPC/BCM57404A4041DC/BCM957414A4141D/BCM957414M4140D	
速度	双端口 25 Gbps 或 10 Gbps 以太网
PCI-E	第 3 代 x8[a]
接口	SFP28 支持 25 Gbps 和 SFP+ 支持 10 Gbps
设备	Broadcom BCM57404/BCM57414 25 Gbps MAC 控制器，集成了双通道 25 Gbps SFI 收发器。
NDIS 名称	Broadcom NetXtreme E 系列双端口 25 Gb SFP28 以太网 PCIe 适配器
UEFI 名称	Broadcom 双 25 Gb SFP 28 以太网
BCM957406A4060DLPC/BCM957406A4060DC/BCM957416A4160D/BCM957416M4160	
速度	双端口 10GBase-T 以太网
PCI-E	第 3 代 x8[a]
接口	RJ45 支持 10 Gbps 和 1 Gbps
设备	Broadcom BCM57406/BCM57416 10 Gbps MAC 控制器，集成了双通道 10GBase-T 收发器。
NDIS 名称	Broadcom NetXtreme E 系列双端口 10GBASE-T 以太网 PCIe 适配器
UEFI 名称	Broadcom 双 10GBASE-T 以太网

a. 该 NIC 支持 PCI-E 第 3 代、第 2 代和第 1 代速度，但是，当 2 个端口的 25G 链路同时传送和接收通信量时，推荐 PCI 第 3 代，以便达到标称吞吐量。

图 1：BCM957402A4020DC、BCM957412A4120D 网卡

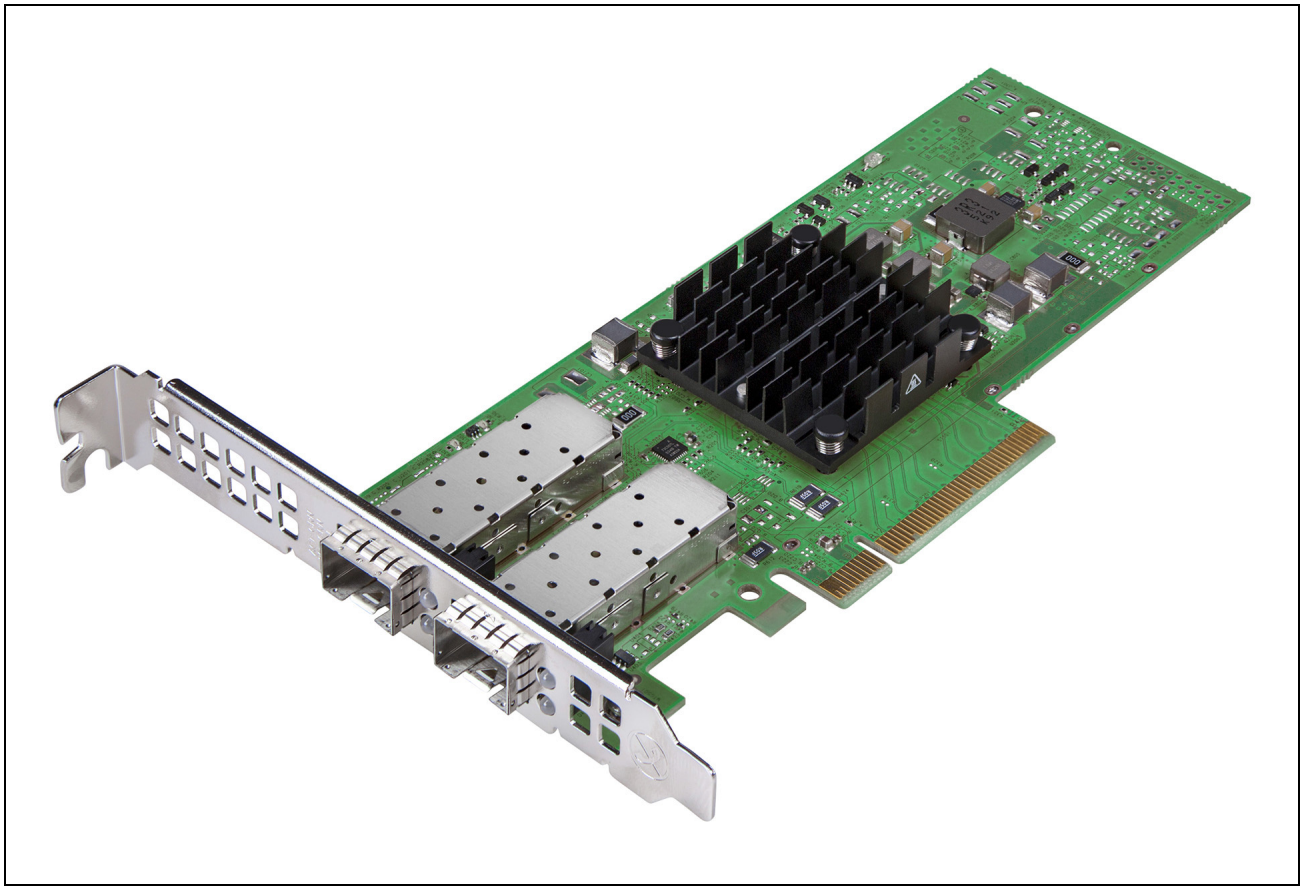


图 2：BCM957404A4041DLPC、BCM957414A4141D 网卡

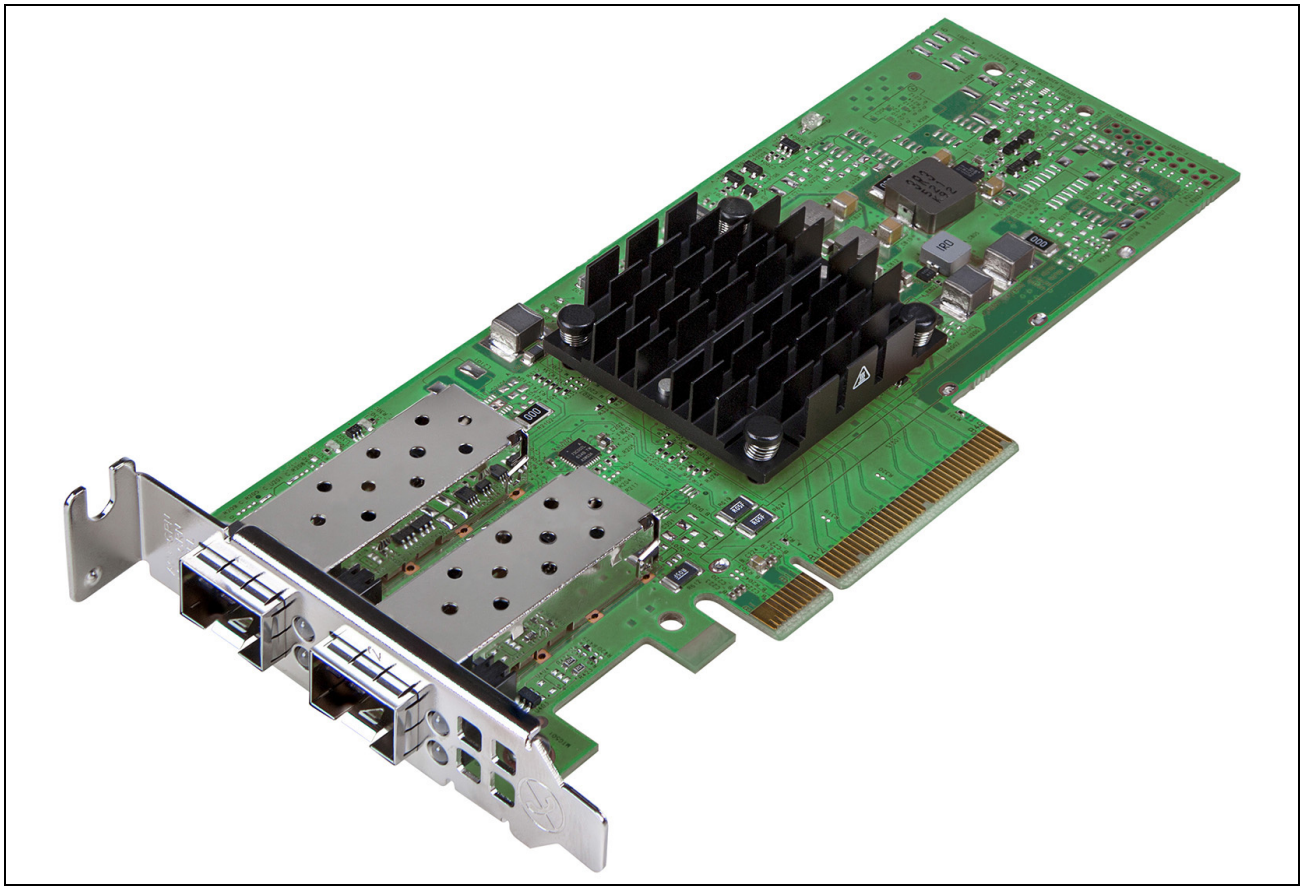


图 3：BCM957406A4060DLPC、BCM957416A4160D 网卡

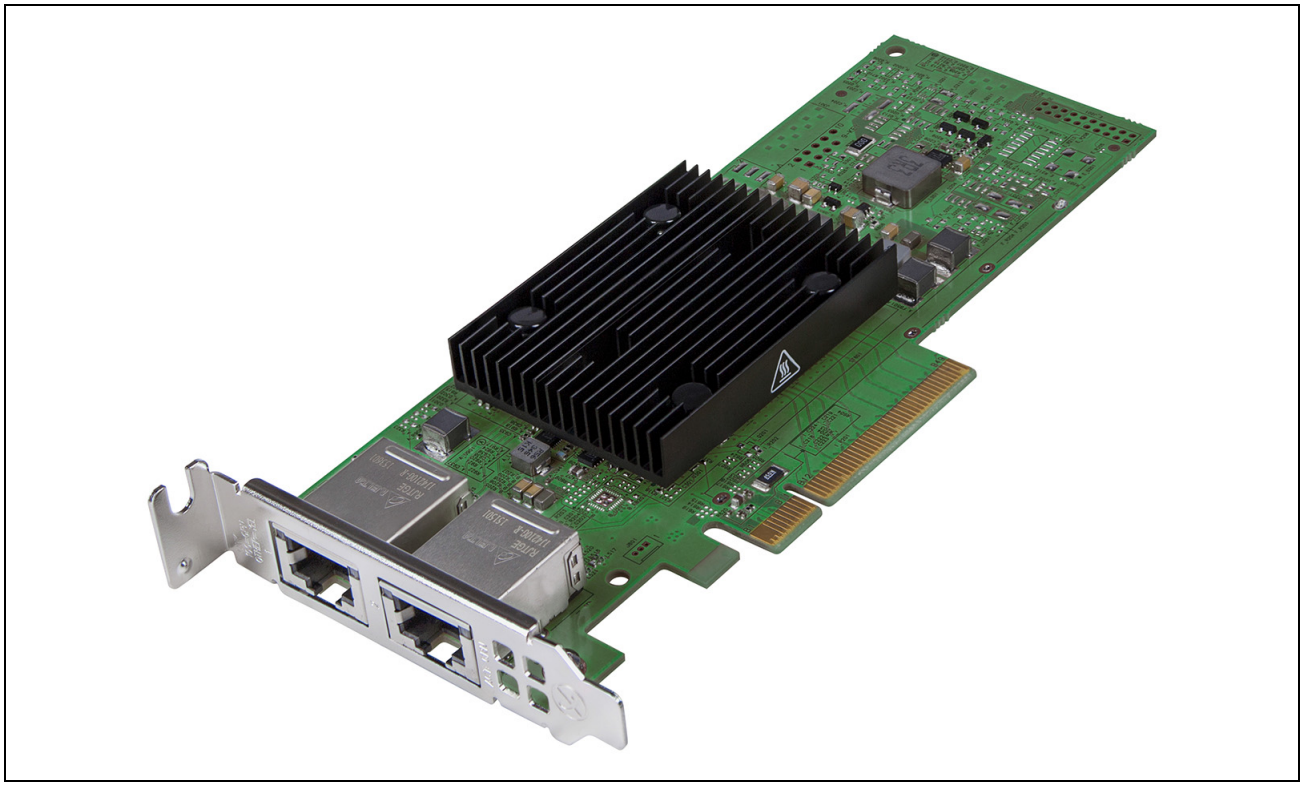


图 4：BCM957414M4140D 网络子卡 (rNDC)

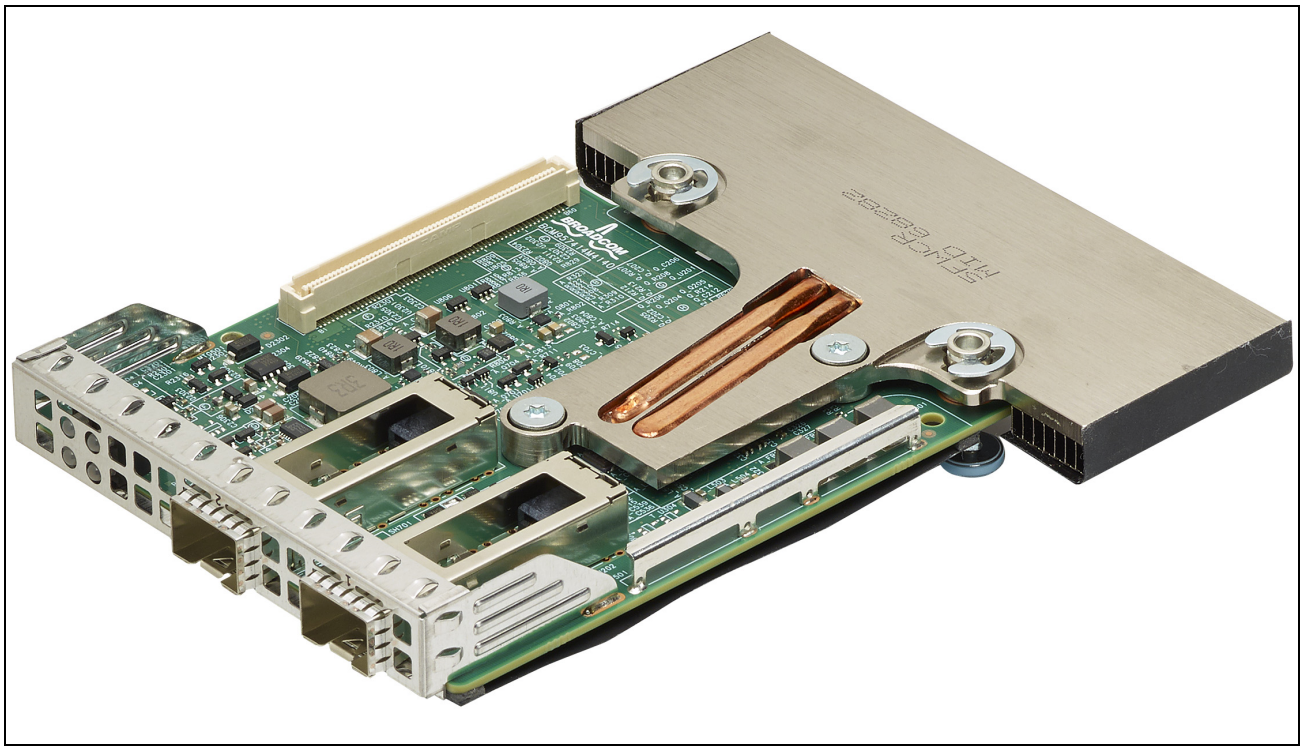


图 5：BCM957412M4120D 网络子卡 (rNDC)

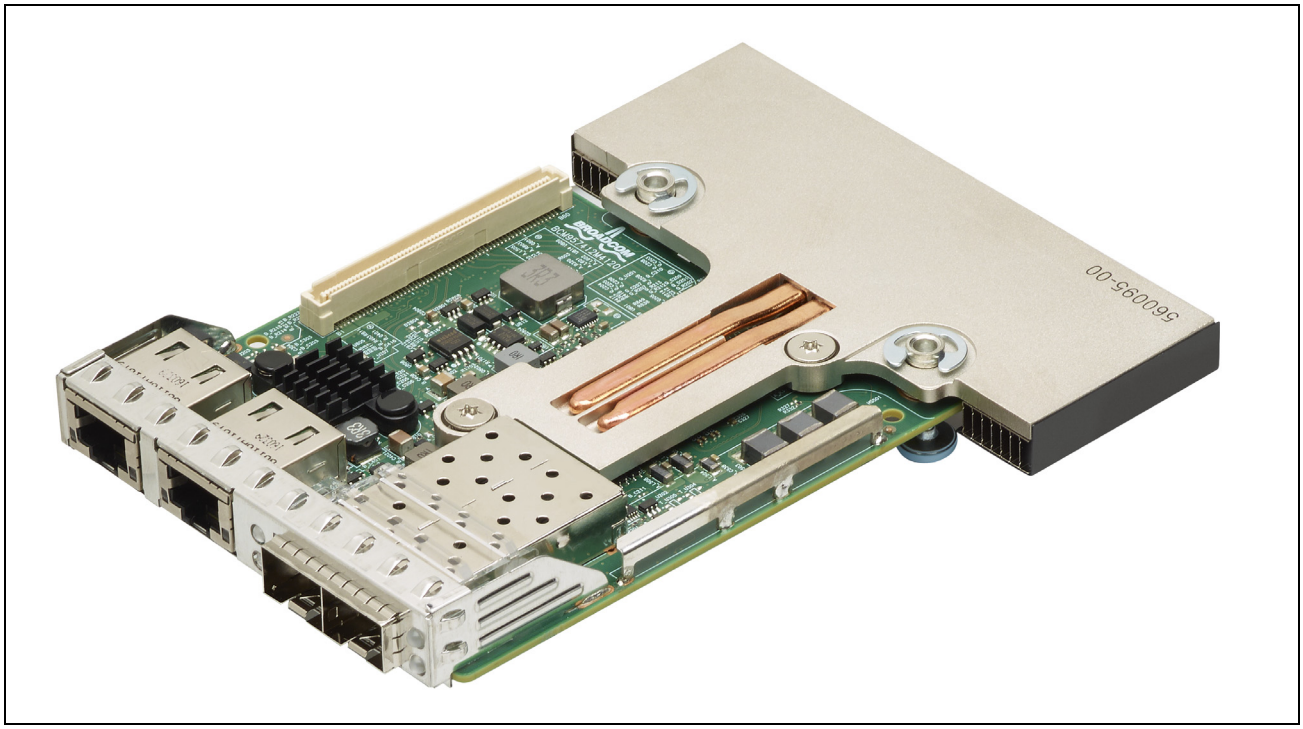
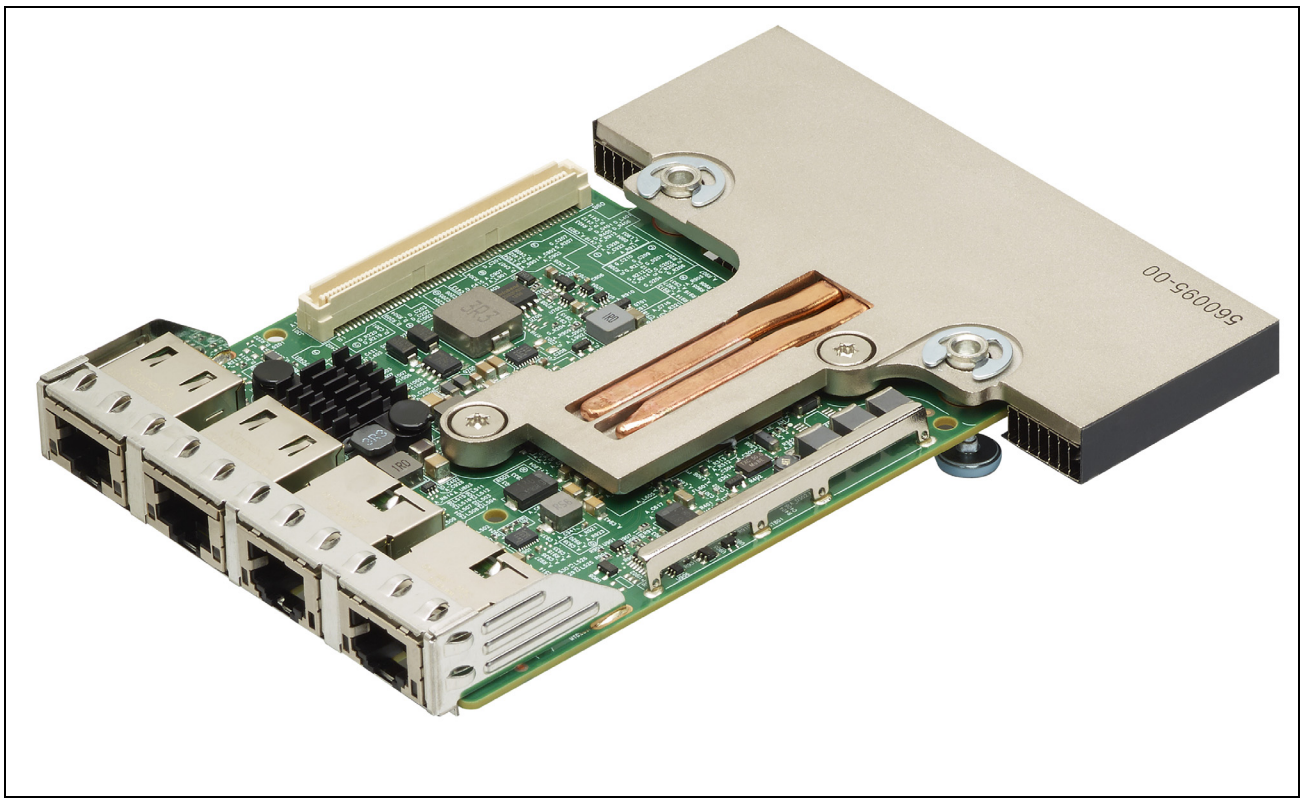


图 6：BCM957416M4160 网络子卡 (rNDC)



网络链路和活动指示

BCM957402AXXXX/BCM957412AXXXX

SFP+ 端口有两个 LED，用来指示通讯活动和链路速度。通过支架上的开孔可以看到这些 LED，如图 7 中所示。LED 功能在表 6 中介绍。

图 7：BCM957402AXXXX/BCM957412AXXXX 活动和链路 LED 位置

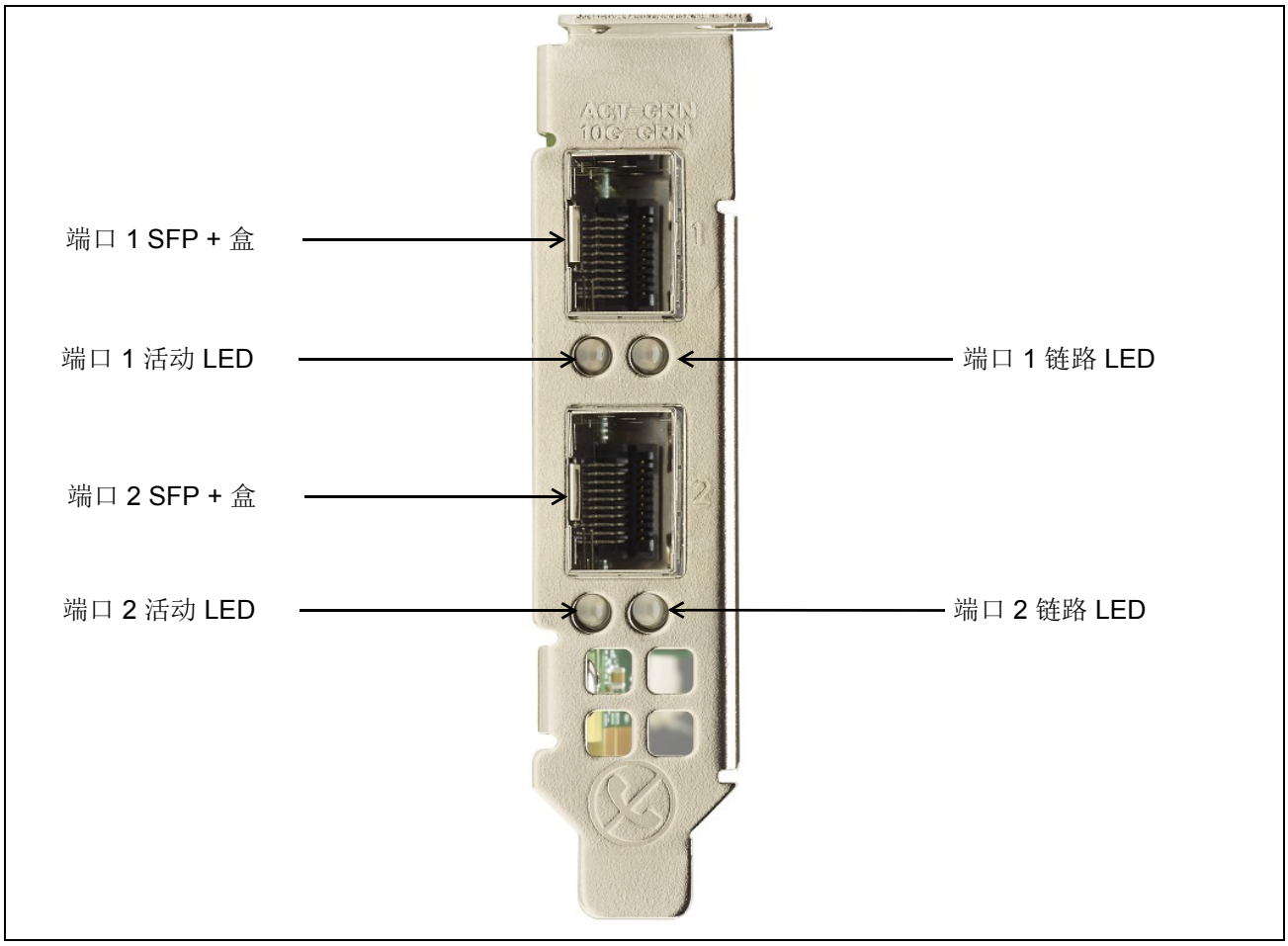


***表 6：BCM957402AXXXX/BCM957412AXXXX* 活动和链路 LED 位置**

LED 类型	*颜色/行为*	*注释*
活动	熄灭	无活动
	绿色闪烁	通讯流活动
链路	熄灭	无链路
	绿色	链路速度为 10 Gbps

BCM957404AXXXX/BCM957414AXXXX

SFP28 端口有两个 LED，用来指示通讯活动和链路速度。通过支架上的开孔可以看到这些 LED，如图 8 中所示。LED 功能在表 7 中介绍。

图 8：BCM957404AXXXX/BCM957414AXXXX 活动和链路 LED 位置

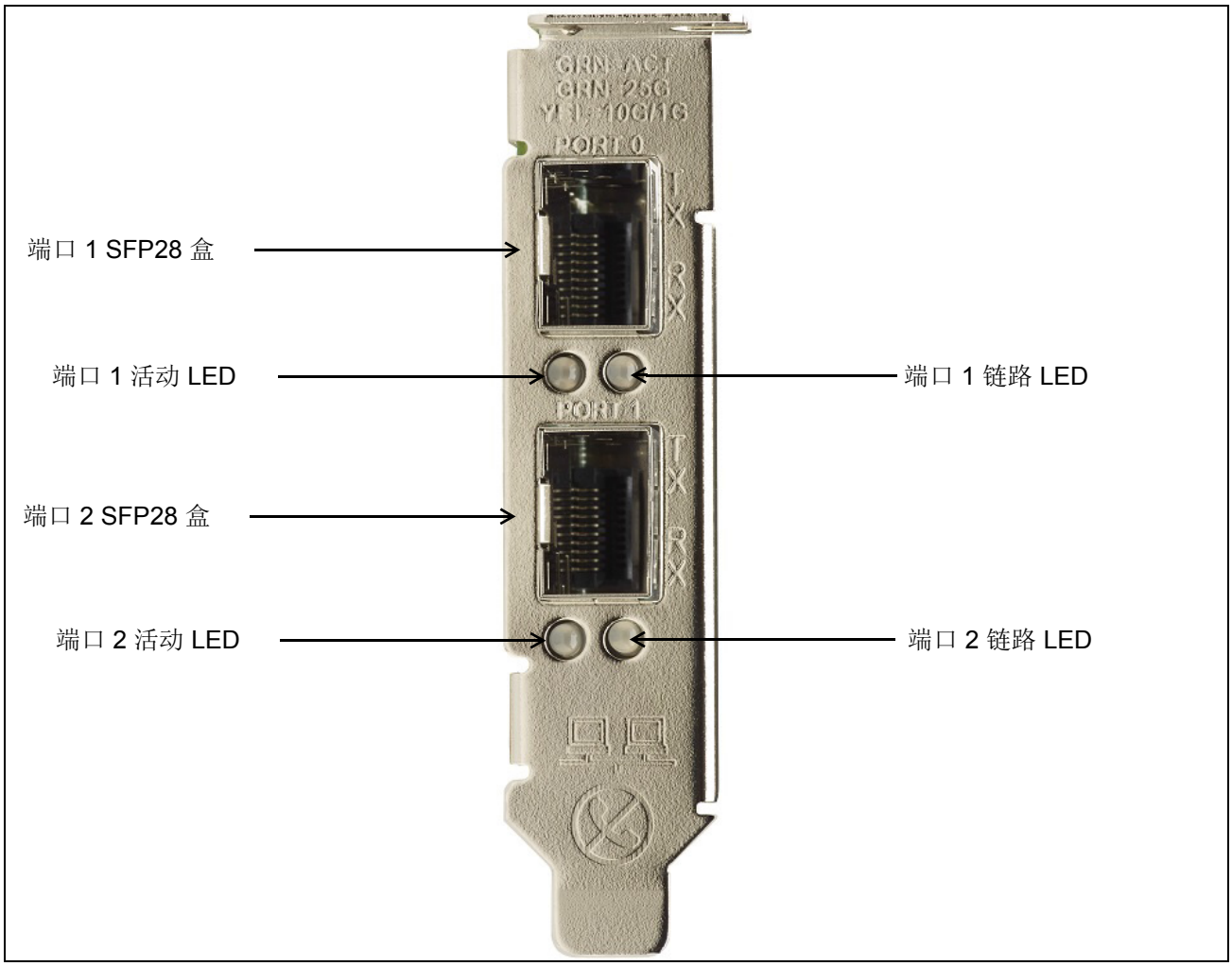


表 7：BCM957404AXXXX/BCM957414AXXXX 活动和链路 LED 位置

LED 类型	*颜色/行为*	*注释*
活动	熄灭	无活动
	绿色闪烁	通讯流活动
链路	熄灭	无链路
	绿色	链路速度为 25 Gbps
	黄色	链路速度为 10 Gbps

BCM957406AXXXX/BCM957416AXXXX

RJ-45 端口有两个 LED，用来指示通讯活动和链路速度。通过支架上的开孔可以看到这些 LED，如图 9 中所示。LED 功能在表 8 中介绍。

图 9：BCM957406AXXXX/BCM957416AXXXX 活动和链路 LED 位置

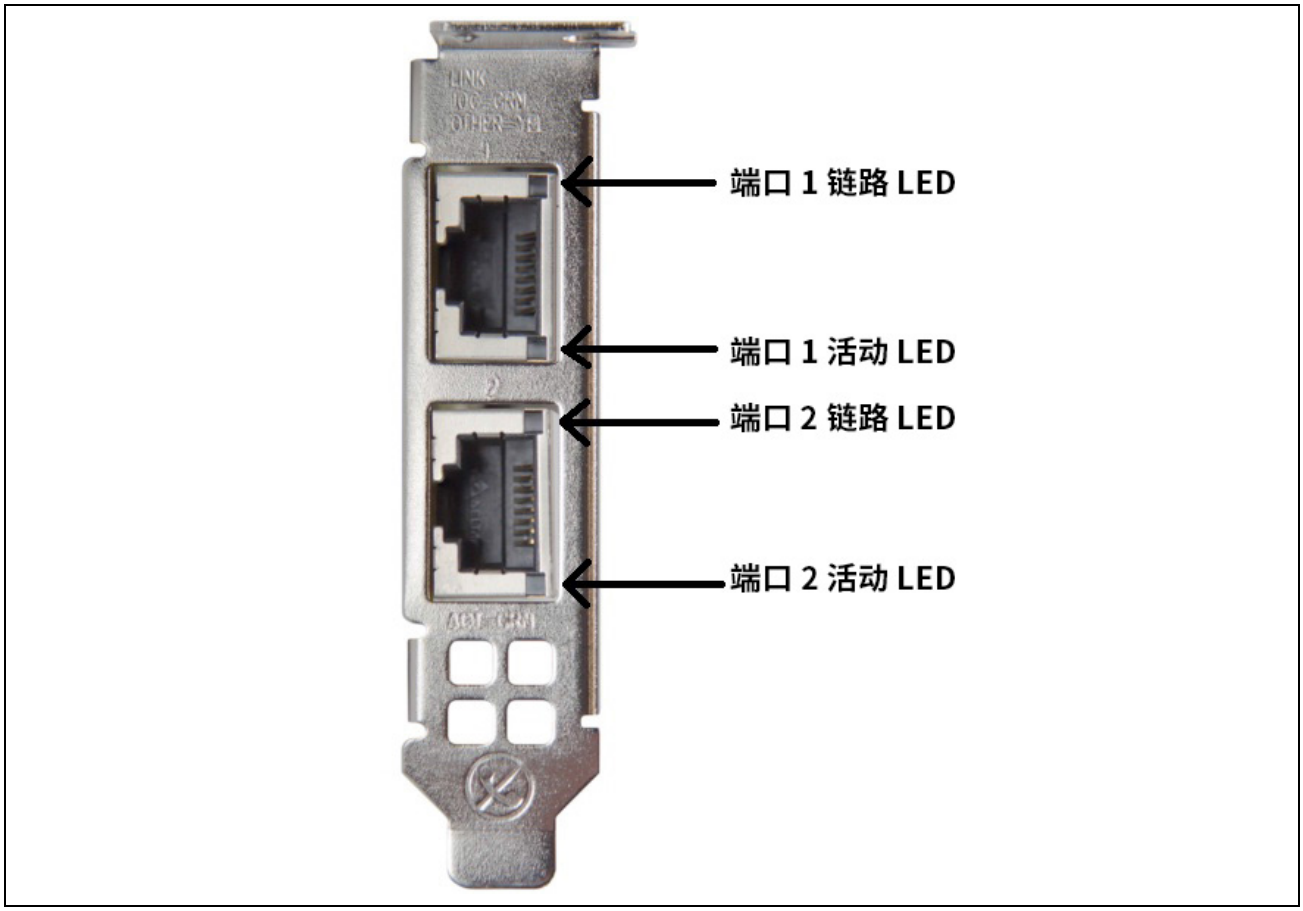


***表 8：BCM957406AXXXX/BCM957416AXXXX* 活动和链路 LED 位置**

LED 类型	*颜色/行为*	*便笺*
活动	熄灭	无活动
	绿色闪烁	通讯流活动
链路	熄灭	无链路
	绿色	链路速度为 10 Gbps
	琥珀色	链路速度为 1 Gbps

BCM957414M4140D

SFP28 端口有两个 LED，用来指示通讯活动和链路速度。通过支架上的开孔可以看到这些 LED，如图 10 中所示。

图 10：BCM957414M4140D 网络子卡 (rNDC) 活动和链路 LED 位置

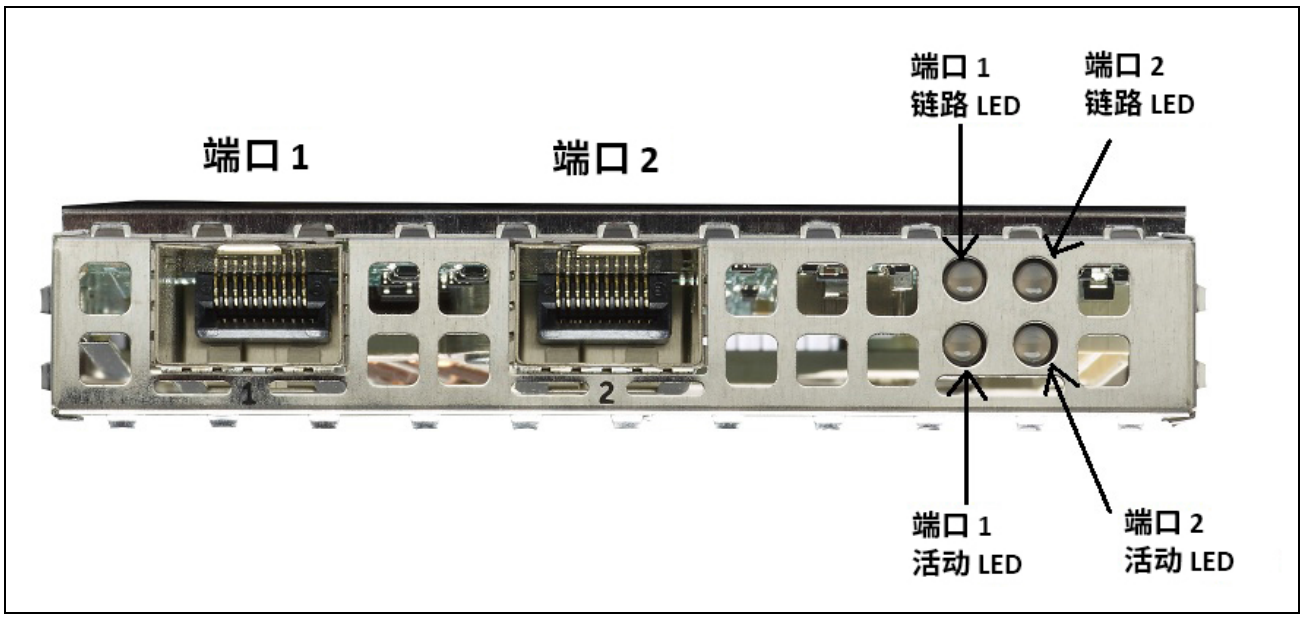


表 9：BCM957414M4140D 网络子卡 (rNDC) 活动和链路 LED 位置

LED 类型	*颜色/行为*	*便笺*
活动	熄灭	无活动
	绿色闪烁	通讯流活动
链路	熄灭	无链路
	绿色	链路速度为 25 Gbps
	黄色	链路速度为 10 Gbps

BCM957412M4120D

此 rNDC 具有 SFP+ 和 RJ-45 端口，每个端口都有两个 LED，用来指示通讯活动和链路速度。LED 如图 11 中所示。

图 11：BCM957412M4120D 网络子卡 (rNDC) 活动和链路 LED 位置

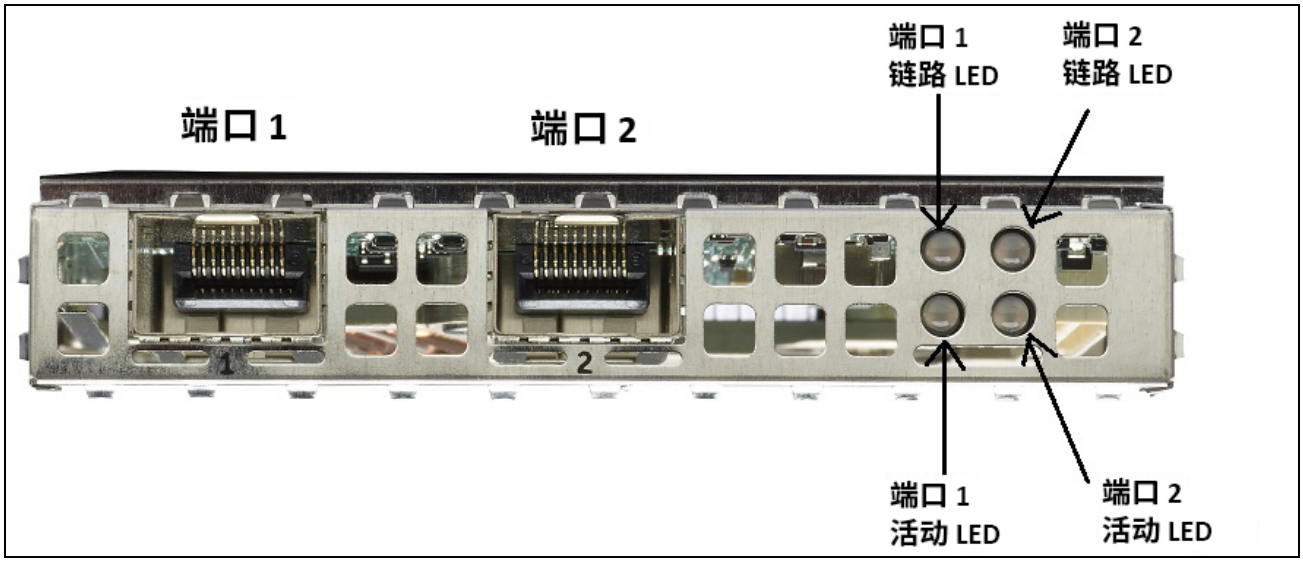


表 10：BCM957412M4120D 网络子卡 (rNDC) 活动和链路 LED 位置 SFP+ 端口 1 和 2

LED 类型	*颜色/行为*	*便笺*
活动	熄灭	无活动
	绿色闪烁	通讯流活动
链路	熄灭	无链路
	绿色	链路速度为 10 Gbps

表 11：1000BaseT 端口 3 和 4

LED 类型	*颜色/行为*	*便笺*
活动	熄灭	无活动
	绿色闪烁	通讯流活动
链路	熄灭	无链路
	绿色	链路速度为 1 Gbps
	琥珀色	链路速度为 10/100 Gbps

BCM957416M4160

此 rNDC 具有 10GBaseT 和 1000BaseT RJ-45 端口，每个端口都有两个 LED，用来指示通讯活动和链路速度。LED 如图 12 中所示。

图 12：BCM957416M4160 网络子卡 (rNDC) 活动和链路 LED 位置

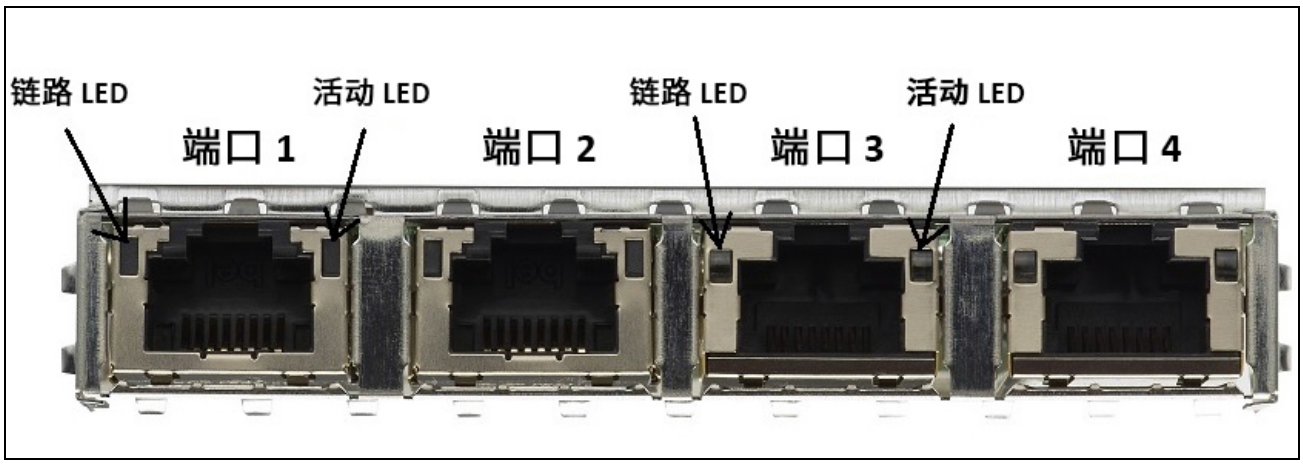


表 12：BCM957416M4160 网络子卡 (rNDC) 活动和链路 LED 位置 10GBaseT 端口 1 和 2

LED 类型	*颜色/行为*	*便笺*
活动	熄灭	无活动
	绿色闪烁	通讯流活动
链路	熄灭	无链路
	绿色	链路速度为 10 Gbps
	琥珀色	链路速度为 1 Gbps

特性

参阅以下各节，了解设备特性。

软件和硬件特性

表 13 提供主机接口特性列表。

表 13：主机接口特性

特性	*详情*
主机接口	PCIe v3.0（第 3 代：8 GT/s；第 2 代：5 GT/s；第 1 代：2.5 GT/s）。
PCIe 数据通路数	PCI-E 边缘接驳器：x8。
关键产品数据 (VPD)	支持。
替换工艺路线 ID (ARI)	支持。
功能级的复位 (FLR)	支持。
高级错误报告	支持。
PCIe ECN	支持 TLP 处理提示 (TPH)、延迟容忍报告 (LTR) 和优化的缓存刷新/填充 (OBFF)。
每队列 MSI-X 中断矢量	每个 RSS 队列 1 个，每个 NetQueue 1 个，每个虚拟机队列 (VMQ) 1 个。
IP Checksum Offload	发送端和接收端支持。
TCP Checksum Offload	发送端和接收端支持。
UDP Checksum Offload	发送端和接收端支持。
NDIS TCP Large Send Offload	LSOV1 和 LSOV2 支持。
NDIS 接收段合并 (RSC)	Windows 环境支持。
TCP Segmentation Offload (TSO)	Linux 和 VMware 环境支持。
Large Receive Offload (LRO)	Linux 和 VMware 环境支持。
Generic Receive Offload (GRO)	Linux 和 VMware 环境支持。
接收端伸缩 (RSS)	Windows、Linux 和 VMware 环境支持。RSS 支持最多 8 个队列/端口。
标头-承载数据分离	使软件 TCP/IP 堆栈能够接收 TCP/IP 数据包并将标头和承载数据分离到不同的缓存中。支持 Windows、Linux 和 VMware 环境。

表 13：主机接口特性（待续）

特性	*详情*
Jumbo 帧	支持。
iSCSI 引导	支持。
NIC 分区 (NPAR)	支持每端口最多 8 个物理功能 (PF)，或每芯片最多 16 个 PF。该选项可配置于 NVRAM。
融合以太网上的 RDMA (RoCE)	BCM5741X 支持 Windows、Linux 和 VMware 环境下的 RoCE v1/v2。
数据中心桥接 (DCB)	BCM5741X 支持 DCBX（IEEE 和 CEE 规范）、PFC 和 AVB。
NCSI（网络控制器边带接口）	支持。
Wake on LAN (WOL)	带 10GBase-T、SFP+ 和 SFP28 接口的 rNDC 上受支持。
PXE 引导	支持。
UEFI 引导	支持。
流控制（暂停）	支持。
自动协商	支持。
802.1q VLAN	支持。
中断调整	支持。
MAC/VLAN 过滤器	支持。

虚拟化特性

表 14 列出了 NetXtreme-E 的虚拟化特性。

表 14：虚拟化特性

特性	*详情*
Linux KVM 多队列	支持。
VMware NetQueue	支持。
NDIS 虚拟机队列 (VMQ)	支持。
虚拟可扩展局域网 (VXLAN) - 知悉无状态卸载（IP/UDP/TCP 校验和卸载）	支持。
通用路由封装 (GRE) - 知悉无状态卸载（IP/UDP/TCP 校验和卸载）	支持。
使用通用路由封装的网络虚拟化 (NVGRE) - 知悉无状态卸载	支持。
IP-in-IP 知悉无状态卸载（IP/UDP/TCP 校验和卸载）	支持
SR-IOV v1.0	每设备来宾操作系统 (GOS) 的 128 个虚拟功能 (VF)。每个 VF 的 MSI-X 矢量设为 16。
MSI-X 矢量端口	每端口默认值 74（两端口配置）。每个 VF 为 16 个，且可配置于 HII 和 CCM。

VXLAN

IETF RFC 7348 中定义的虚拟可扩展局域网 (VXLAN) 用来应对容纳多租户的虚拟化数据中心内的重叠网络的需要。VXLAN 是 3 层网络上的 2 层重叠或隧道技术方案。只有相同 VXLAN 段内的 VM 才能相互通信。

NVGRE/GRE/IP-in-IP/Geneve

IETF RFC 7637 中定义的使用 GRE 的网络虚拟化 (NVGRE) 类似于 VXLAN。

无状态卸载

RSS

接收端伸缩 (RSS) 使用一种 Toeplitz 算法，该算法在接收的帧上使用 4 元组匹配，并将其转发到确定性 CPU 进行帧处理。这实现了流线化的帧处理并且平衡 CPU 利用率。使用一个间接表，将流映射到一个 CPU。

对称 RSS 允许给定 TCP 或 UDP 流的数据包映射到同一接收队列。

TPA

Transparent Packet Aggregation (TPA) 是一个技法，对于接收的帧，相同 4 元组匹配的帧聚集在一起，然后指示给网络堆栈。TPA 上下文中的每个条目都由以下 4 元组标识：源 IP、目的 IP、源 TCP 端口和目的 TCP 端口。TPA 通过减少网络通信中断和降低 CPU 开销提高了系统性能。

标头-承载数据分离

标头-承载数据分离是一个特性，使软件 TCP/IP 堆栈能够接收 TCP/IP 数据包并将标头和承载数据分离到不同的缓存中。Windows 和 Linux 环境中均提供对此特性的支持。以下是标头-承载数据分离的潜在好处：

- 标头-承载数据分离使数据包标头紧凑而高效地缓存到主机 CPU 缓冲区。这可能导致接收端 TCP/IP 性能提升。
- 标头-承载数据分离通过主机 TCP/IP 堆栈实现翻页和零复制操作。这可能进一步提高接收路径的性能。

UDP 分片卸载

UDP 分片卸载 (UFO) 是一个特性，使软件堆栈能够将 UDP/IP 数据报的分片卸载到 UDP/IP 数据包中。只有 Linux 环境中提供对此特性的支持。以下是 UFO 的潜在好处：

- UFO 使 NIC 能够将 UDP 数据报的分片处理到 UDP/IP 数据包中。这可能导致发送端 UDP/IP 处理的 CPU 开销的降低。

无状态传输隧道卸载

无状态传输隧道卸载 (STT) 是在虚拟化数据中心内实现重叠网络的一种隧道封装。STT 使用基于 IP 的封装，有类似 TCP 的标头。没有与隧道相关的 TCP 连接状态，这就是 STT 无状态的原因。Open Virtual Switch (OVS) 使用 STT。

STT 帧包含 STT 帧标头和承载数据。STT 帧的承载数据是未标记的以太网帧。STT 帧标头和封装的承载数据被视为 TCP 承载数据和类似 TCP 的标头。为传送的每个 STT 段创建 IP 标头（IPv4 或 IPv6）和以太网标头。

对操作系统的多队列支持

NDIS VMQ

NDIS 虚拟机队列 (VMQ) 是受 Microsoft 支持的一个特性，用来提高 Hyper-V 网络性能。VMQ 特性支持基于目的地 MAC 地址的数据包分类，在不同的完成队列上返回收到的数据包。此数据包分类与 DMA 数据包直接进入虚拟机内存的能力相结合，允许跨多个处理器伸缩虚拟机。

参阅第 34 页上的“Windows 驱动程序高级属性和事件日志消息”，了解有关 VMQ 的信息。

VMWare NetQueue

VMware NetQueue 是与 Microsoft 的 NDIS VMQ 特性类似的特性。NetQueue 特性支持基于目的地 MAC 地址和 VLAN 的数据包分类，在不同的 NetQueue 上返回收到的数据包。此数据包分类与 DMA 数据包直接进入虚拟机内存的能力相结合，允许跨多个处理器伸缩虚拟机。

KVM/Xen 多队列

KVM/多队列通过处理已收到数据包的目的地 MAC 地址和/或 802.1Q VLAN 标记对传入帧进行分类，将这些帧返回到不同的主机栈队列。该分类与 DMA 数据帧直接进入虚拟机内存的能力相结合，允许跨多个处理器伸缩虚拟机。

SR-IOV 配置支持矩阵

- Windows 管理程序上的 Windows VF
- VMware 管理程序上的 Windows VF 和 Linux VF
- Linux KVM 上的 Linux VF

SR-IOV

PCI-SIG 定义对单域根 IO 虚拟化 (SR-IOV) 的可选支持。SR-IOV 旨在允许 VM 使用虚拟功能 (VF) 直接访问设备。NIC 物理功能 (PF) 划分为多个虚拟功能，每个 VF 表示为 VM 的一个 PF。

SR-IOV 使用 IOMMU 功能，借助转换表将 PCI-E 虚拟地址转换为物理地址。

物理功能 (PF) 和虚拟功能 (VF) 的数量通过 UEFI HII 菜单、CCM 和通过 NVRAM 配置进行管理。SRIOV 可在与 NPAR 模式组合时受到支持。

网络分区 (NPAR)

网络分区 (NPAR) 功能允许单个物理网络接口的端口在系统中实现多个网络设备的功能。当 NPAR 模式已启用时，NetXtreme-E 设备被枚举为多个 PCIe 物理功能 (PF)。在初始开机时会为每个 PF 或“分区”分配一个单独的 PCIe 功能 ID。原始的 PCIe 定义允许为每台设备定义 8 个 PF。对于具有替换工艺路线-ID (ARI) 功能的系统，Broadcom NetXtreme-E 适配器支持每台设备最多 16 个 PF。每个分区都被分配一个单独的配置空间、BAR 地址和 MAC 地址，以使其可以独立操作。与任何其他物理接口一样，分区支持直接分配给 VM、VLAN 等。

注： 在**系统设置 > 设备设置 > [Broadcom 5741x 设备] > 设备级配置**页面，用户可以启用 **NParEP** 以允许 NXE 适配器支持每台设备最多 16 个 PF。对于 2 端口设备，表示每个端口最多 8 个 PF。

融合以太网上的 RDMA - RoCE

融合以太网上的远程直接内存访问 (RDMA) (RoCE) 是 BCM5741X（支持以太网上的 RDMA 功能）中的完整硬件卸载功能。RoCE 功能适用于用户模式和内核模式应用。RoCE 物理功能 (PF) 和 SRIOV 虚拟功能 (VF) 适用于单功能模式和多功能模式（NIC 分区模式）。Broadcom 支持 Windows、Linux 和 VMware 环境下的 RoCE。

请参阅以下链接以获取每个操作系统的 RDMA 支持：

Windows

https://technet.microsoft.com/en-us/library/jj134210(v=ws.11).aspx

Redhat Linux

https://access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/7/html/Networking_Guide/ch-Configure_InfiniBand_and_RDMA_Networks.html

VMware

https://pubs.vmware.com/vsphere-65/index.jsp?topic=%2Fcom.vmware.vsphere.networking.doc%2FGUID-4A5EBD44-FB1E-4A83-BB47-BBC65181E1C2.html

支持的组合

以下各节描述了此设备支持的功能组合。

NPAR、SR-IOV 和 RoCE

表 15 提供了 NPAR、SR-IOV 和 RoCE 支持的功能组合。

表 15：NPAR、SR-IOV 和 RoCE

SW 功能	*便笺*
NPAR	最多 8 个或 16 个 PF
SR-IOV	最多 128 个 VF（每芯片合计）
基于 PF 的 RoCE	最多 4 个 PF
基于 VF 的 RoCE	VF 连接到已启用 RoCE 的 PF 时有效
主机操作系统	Linux、Windows、ESXi（不支持 vRDMA）
客户机操作系统	Linux 和 Windows
DCB	每端口最多 2 个 COS，带非共享保留内存

NPAR、SR-IOV 和 DPDK

表 16 提供了 NPAR、SR-IOV 和 DPDK 支持的功能组合。

表 16：NPAR、SR-IOV 和 DPDK

SW 功能	*便笺*
NPAR	最多 8 个或 16 个 PF
SR-IOV	最多 128 个 VF（每芯片合计）
DPDK	仅支持作为 VF
主机操作系统	Linux
客户机操作系统	DPDK (Linux)

不支持的组合

不支持 NPAR、SR-IOV、RoCE 和 DPDK 的组合。

安装硬件

安全预防措施

注意！ 适配器安装在致命高压系统中。在卸下系统机盖之前，请仔细阅读以下预防措施以保护自己并避免损坏系统组件：

- 从手上和手腕取下所有金属物体或珠宝。
- 确保只使用绝缘或不导电工具。
- 接触内部组件之前，请验证系统电源已关闭并已拔下插头。
- 在无静电环境中安装或卸下适配器。强烈建议您使用正确接地的腕带或其它个人防静电设备与防静电垫。

系统要求

在安装 Broadcom NetXtreme-E 以太网适配器之前，请验证系统符合对操作系统列出的要求。

硬件要求

参阅下面列出的硬件要求：

- 符合操作系统要求的 Dell 13G 系统。
- 支持 NetXtreme-E 以太网卡的 Dell 13G 系统。
- 带 rNDC 封装的 NIC 适配器有一个未使用的 PCI-E 第 3 代 x8 插槽或一个未使用的 PCIe 第 3 代 rNDC 插槽。
- 4 GB 或更大内存（推荐 32 GB 或更大内存，以利于虚拟化应用和标称网络吞吐性能）。

预安装检查表

在安装 NetXtreme-E 设备之前，参阅下面的列表。

1. 验证服务器符合“系统要求”中列出的硬件和软件要求。
2. 验证服务器在使用最新的 BIOS。
3. 如果系统处于活动状态，请将其关闭。
4. 系统关机完成后，关闭电源并拔下电源线。
5. 抓住适配器卡的边缘，将它从其运输包装中取出并放在防静电表面上。
6. 检查适配器，特别是在插卡边缘的连接器上是否有明显的损坏痕迹。切勿尝试安装任何损坏的适配器。

安装适配器

下面说明适用于在大多数服务器中安装 Broadcom NetXtreme-E 以太网适配器（外插 NIC）。参阅随服务器提供的手册，了解有关在此特定服务器上执行这些任务的详细信息。

1. 在安装适配器之前，查阅第 26 页上的“安全预防措施”和“预安装检查表”。确保系统电源已关闭并已从电源插座上断开，并且已执行了适当的电接地步骤。
2. 打开系统机箱并选择空闲的 PCI Express 第 3 代 x8 插槽。
3. 从插槽中取出封盖板。
4. 将适配器的连接器边缘与系统中的连接器插槽对齐。
5. 使用适配器夹或螺丝固定适配器。
6. 合上系统机箱，并断开任何个人防静电设备。

注：对于带 rNDC 封装的 NIC 适配器，找到一个未使用的 RNDC 插槽或将现有默认 rNDC 移除并替换为 NetXtreme rNDC。

连接网络电缆

Dell 以太网交换机采用支持最高 100 Gbps 的 SFP+/SFP28/QSFP28 端口实现产品化。这些 100 Gbps 端口可以分成 4 x 25 Gbps SFP28 端口。QSFP 端口可以使用 4 x 25G SFP28 分支电缆连接到 SFP28 端口。

支持的电缆和模块

表 17：支持的电缆和模块

光纤模块	*Dell 部件号*	*适配器*	*描述*
FTLX8571D3BCL-DL	3G84K	BCM957404A4041DLPC、 BCM957404A4041DC BCM957402A4020DLPC、 BCM957402A4020DC	10 Gbps 850 纳米多模 SFP+ 收发器
FCLF-8521-3	8T47V	BCM957406A4060DLPC、 BCM957406A4060DC	1000Base-T 铜缆 SFP 收发器
FTLF8536P4BCL	P7D7R	BCM957404A4041DLPC、 BCM957404A4041DC	25 Gbps 短波长 SFP+ 收发器
Methode DM7051	PGYJT	BCM57402X、 BCM57404X、 BCM57412X、BCM57414X	SFP+ 到 10GBASE-T 收发器
FTLX1471D3BCL-FC	RN84N	BCM57402X、 BCM57404X、 BCM57412X、BCM57414X	25 Gpbs SFP28 收发器
FTLX8574D3BNL	N8TDR	BCM57402X、 BCM57404X、 BCM57412X、BCM57414X	85C 扩展温度范围 10 Gbps SFP+ 收发器

表 17：支持的电缆和模块（待续）

光纤模块	*Dell 部件号*	*适配器*	*描述*
FTLF8536P4BNL-FC	HHHHC	BCM57402X、BCM57404X、BCM57412X、BCM57414X	85C 扩展温度范围 10 Gbps SFP+ 收发器
FTLX8574D3BCL-FC 或 PLRXPLSCS43811	WTRD1	BCM57402X、BCM57404X、BCM57412X、BCM57414X	10 Gbps-SR SFP+ 收发器

注：

1. 符合 IEEE 标准的直连电缆 (DAC) 可以连接到适配器。

2. BCM957414M4140D 需要 Dell 部件 HHHHC 和 N8TDR。

铜缆

BCM957406AXXXX、BCM957416AXXXX 和 BCM957416XXXX 适配器有两个 RJ-45 连接器，用于将系统连接到 CAT 6E 以太网铜线段。

SFP+

BCM957402AXXXX、BCM957412AXXXX 和 BCM957412MXXXX 适配器有两个 SFP+ 连接器，用于将系统连接到 10 Gbps 以太网交换机。

SFP28

BCM957404AXXXX、BCM957414XXXX 和 BCM957414AXXXX 适配器有两个 SFP28 连接器，用于将系统连接到 100 Gbps 以太网交换机。

软件包和安装

参阅以下各节，了解有关软件包和安装的信息。

支持的操作系统

表 18 提供支持的操作系统列表。

表 18：支持的操作系统列表

操作系统品类	*分发版*
Windows	Windows 2012 R2 或更高
Linux	Redhat 6.9、Redhat 7.1 或更高 SLES 11 SP 4、SLES 12 SP 2 或更高
VMware	ESXi 6.0 U3 或更高

安装驱动程序

参阅以下各节，了解驱动程序安装。

Windows

Dell DUP

Broadcom NetXtreme E 系列控制器驱动程序可以使用驱动程序 DUP 进行安装。该安装程序以 x64 可执行文件格式提供。

GUI 安装

当文件执行时，会出现对话框，要求用户输入。该安装程序支持仅驱动程序选项。

静默安装

可执行文件可以使用下面所示的命令静默地执行。

示例：

```
Network_Driver_<version>.EXE /s /driveronly
```

INF 安装

Dell DUP 用来安装 Broadcom NetXreme-E 以太网控制器的驱动程序。使用以下命令，从 Dell DUP 中提取驱动程序 INF 文件：

```
Network_Driver_<version>.EXE /s /v"EXTRACTDRIVERS=c:\dell\drivers\network"
```

提取文件后，使用设备管理器 (devmgmt.msc) 通过“升级驱动程序”功能执行 INF 安装。打开设备管理器，选择所需的 NIC，右键单击并选择升级驱动程序来进行更新。

Linux

Linux 驱动程序以 RPM、KMP 和源代码格式提供。要使用 Linux 从源代码生成设备驱动程序，请参考以下示例：

1. 以根用户身份登录到 Linux 系统。
2. scp 或 cp 驱动程序 tar ball 到 Linux 系统上。典型的示例为：

   ```
   cp /var/run/media/usb/bnxt_en-<version>.tar.gz /root/
   ```

3. 执行以下命令：

   ```
   tar –zxvf /root/bnxt_en-<version>.tar.gz
   ```

4. 执行以下命令：

   ```
   cd bnxt_en-<version>
   ```

5. 执行以下命令：

   ```
   make; make install; modprobe -r bnxt_en; modprobe bnxt_en
   ```

对于 RDMA 功能，需要安装 `bnxt_en` 和 `bnxt_re` 驱动程序。使用 `netxtreme-bnxt_en-<version>.tar.gz`，而不使用 `bnxt_en-<version>.tar.gz`。

模块安装

RHEL

驱动程序映像可使用下列选项之一进行安装：

- 使用 Dell iDRAC 虚拟控制台挂载 bnxt_en-x.x.x-rhelYuZ-x86_64-dd.iso 映像。
- 从 CD/DVD 挂载 bnxt_en-x.x.x-rhelYuZ-x86_64-dd.iso 映像。
- 将 bnxt_en-x.x.x-rhelYuZ-x86_64-dd.iso 映像复制到 USB 设备并挂载该设备。

启动操作系统安装，按 tab 键，然后输入“linux dd”。继续安装，直到要求提供驱动程序磁盘，选择 bnxt_en 驱动程序。

SLES

驱动程序映像可使用下列选项之一进行安装：

- 使用 Dell iDRAC 虚拟控制台挂载 bnxt_en-x.x.x-rhelYuZ-x86_64-dd.iso 映像。
- 从 CD/DVD 挂载 bnxt_en-x.x.x-rhelYuZ-x86_64-dd.iso 映像。
- 提取 bnxt_en-x.x.x-rhelYuZ-x86_64-dd.iso 映像，将内容复制到 USB 设备，然后挂载该 USB 设备。

Linux Ethtool 命令

注：在表 19 中，ethX 应替换为实际接口名称。

表 19：Linux Ethtool 命令

命令	*描述*
ethtool -s ethX speed 25000 autoneg off	设置速度。如果链路在一个端口上启动，驱动程序不会允许其他端口设置为不相容的速度。
ethtool -i ethX	输出包括 Package 版本、NIC BIOS 版本（引导代码）。
ethtool -k ethX	显示卸载特性。
ethtool -K ethX tso off	关闭 TSO。
ethtool -K ethX gro off lro off	关闭 GRO / LRO。
ethtool -g ethX	显示环大小。
ethtool -G ethX rx N	设置环大小。
ethtool -S ethX	获得统计信息。
ethtool -l ethX	显示环的数量。
ethtool -L ethX rx 0 tx 0 combined M	设置环的数量。
ethtool -C ethX rx-frames N	设置中断结合。其他支持的参数有：rx-usecs、rx-frames、rx-usecs-irq、rx-frames-irq、tx-usecs、tx-frames、tx-usecs-irq、tx-frames-irq。
ethtool -x ethX	显示 RSS 流散列间接表和 RSS 密钥。
ethtool -s ethX autoneg on speed 10000 duplex full	启用 Autoneg（请参阅 第 40 页上的“自动协商配置”了解更多信息）
ethtool --show-eee ethX	显示 EEE 状态。
ethtool --set-eee ethX eee off	禁用 EEE。
ethtool --set-eee ethX eee on tx-lpi off	启用 EEE，但禁用 LPI。
ethtool -L ethX combined 1 rx 0 tx 0	禁用 RSS。将组合通道设为 1。
ethtool -K ethX ntuple off	通过禁用 ntuple 过滤器禁用加速 RFS。
ethtool -K ethX ntuple on	启用加速 RFS。
Ethtool -t ethX	执行各种诊断自测。
echo 32768 > /proc/sys/net/core/rps_sock_flow_entries echo 2048 > /sys/class/net/ethX/queues/rx-X/rps_flow_cnt	启用环 X 的 RFS。
sysctl -w net.core.busy_read=50	这将忙着读取设备的接收环的时间设为 50 微秒。对于等待数据到达的套接字应用，使用此方法通常可以减少延迟时间 2 或 3 微秒，代价是较高的 CPU 利用率。
echo 4 > /sys/bus/pci/devices/0000:82:00.0/sriov_numvfs	使用总线 82 上的四个 VF、设备 0 和 功能 0 启用 SR-IOV。
ip link set ethX vf 0 mac 00:12:34:56:78:9a	设置 VF MAC 地址。

表 19：Linux Ethtool 命令（待续）

命令	*描述*
ip link set ethX vf 0 state enable	设置 VF 0 的 VF 链路状态。
ip link set ethX vf 0 vlan 100	使用 VLAN ID 100 设置 VF 0。

VMware

ESX 驱动程序以 VMware 标准 VIB 格式提供。

1. 要安装 Ethernet 和 RDMA 驱动程序，请使用以下命令：

```
$ esxcli software vib install --no-sig-check -v <bnxtnet>-<driver version>.vib

$ esxcli software vib install --no-sig-check -v <bnxtroce>-<driver version>.vib
```

2. 系统需要重新启动，新驱动程序才会生效。

表 20 中显示了其他有用的 VMware 命令。

注：在表 20 中，vmnicX 应替换为实际接口名称。

注：`$ kill -HUP $(cat /var/run/vmware/vmkdevmgr.pid)`，此命令需要在 `vmkload_mod bnxtnet` 将模块启用后执行。

表 20：VMware 命令

命令	*描述*
esxcli software vib list |grep bnx	列出安装的 VIB 以查看已成功安装的 bnxt 驱动程序。
esxcfg-module –l bnxtnet	将模块信息打印到屏幕上。
esxcli network get –n vmnicX	获得 vmnicX 属性。
esxcfg-module –g bnxtnet	打印模块参数。
esxcfg-module –s ‘multi_rx_filters=2 disable_tap=0 max_vfs=0,0 RSS=0’	设置模块参数。
vmkload_mod –u bnxtnet	卸载 bnxtnet 模块。
vmkload_mod bnxtnet	加载 bnxtnet 模块。
esxcli network nic set –n vmnicX –D full –S 25000	设置 vmnicX 的速度和双工。
esxcli network nic down –n vmnicX	禁用 vmnicX。
esxcli network nic up –n vmnic6	启用 vmnicX。
bnxtnetcli –s –n vmnic6 –S “25000”	设置链路速度。旧版本的 ESX 中需要使用 Bnxtnetcli 以支持 25G 速度设置。

固件更新

NIC 固件可以使用下列方法之一进行更新：

- 系统处于操作系统已引导状态时使用 Dell Update Package (DUP)。此方法只适用于 Windows 和 Linux 操作系统。
- 使用 Dell iDRAC – Lifecycle Controller。无论是什么操作系统，均可使用此方法。如果系统在运行 VMware，请使用 Lifecycle Controller 升级固件。

参阅产品支持页，网址是 http://www.dell.com/support

Dell Update Package

参阅以下各节，使用 Dell Update Package (DUP)：

Windows

Broadcom NetXtreme-E 系列控制器固件可以使用 Dell DUP 包进行升级。可执行文件以标准 Windows x64 可执行文件格式提供。双击该文件予以执行。

DUP 包可从 http://support.dell.com 下载

Linux

Dell Linux DUP 以 `x86_64` 可执行文件格式提供。使用标准 Linux `chmod` 更新执行权限并运行可执行文件。参阅以下示例：

1. 登录 Linux。
2. `scp` 或 `cp` DUP 可执行文件到文件系统上。典型的示例为：

   ```
   cp /var/run/media/usb/Network_Firmware_<version>.BIN /root/
   ```

3. 执行以下命令：

   ```
   chmod 755 Network_Firmware_<version>.BIN
   ```

4. 执行以下命令：

   ```
   ./Network_Firmware_<version>.BIN
   ```

   需要重新启动才能激活新固件。

Windows 驱动程序高级属性和事件日志消息

驱动程序高级属性

表 21 中显示了 Windows 驱动程序高级属性。

表 21：Windows 驱动程序高级属性

驱动程序键	*参数*	*描述*
封装的任务卸载	启用或禁用	用于配置 NVGRE 封装的任务卸载。
节能以太网	启用或禁用	对铜质端口启用 EEE，对 SFP+ 或 SFP28 端口禁用 EEE。此特性仅对 BCM957406A4060 适配器启用。
流控制	TX 或 RX 或 TX/RX 启用	在 RX 或 TX 或两端都配置流控制。
中断调整	启用或禁用	默认启用。通过节省 CPU 时间允许帧进行批处理。
Jumbo 数据包	1514 或 4088 或 9014	Jumbo 数据包大小。
Large Send offload V2 (IPV4)	启用或禁用	对 IPV4 的 LSO。
Large Send offload V2 (IPV6)	启用或禁用	对 IPV6 的 LSO。
Locally Administered Address	用户输入的 MAC 地址。	在操作系统启动后覆盖默认硬件 MAC 地址。
RSS 队列的最大数量	2、4 或 8。	默认值为 8。允许用户配置接收端伸缩队列。
优先级和 VLAN	优先级和 VLAN 禁用、优先级启用、VLAN 启用、优先级和 VLAN 启用。	默认启用。用于配置 802.1Q 和 802.1P。
接收缓冲区（0=自动）	500 的增量。	默认值为“自动”。
接收端伸缩	启用或禁用。	默认启用
接收分段结合 (IPV4)	启用或禁用。	默认启用
接收分段结合 (IPV6)	启用或禁用。	默认启用
RSS 负载平衡配置文件	NUMA 伸缩静态、Closest 处理器、Closest 处理器静态、保守的伸缩、NUMA 伸缩。	默认 NUMA 伸缩静态。
速度和双工	1 Gbps 或 10 Gbps 或 25 Gbps 或自动协商。	10 Gbps 铜质端口可以自动协商速度，而 25 Gbps 端口设置为强制速度。
SR-IOV	启用或禁用。	默认启用。此参数与硬件配置的 SR-IOV 和 BIOS 配置的 SR-IOV 设置共同起作用。
TCP/UDP checksum offload IPV4	启用 TX/RX、启用 TX 或启用 RX 或禁用 offload。	默认启用 RX 和 TX。

表 21：Windows 驱动程序高级属性（待续）

驱动程序键	*参数*	*描述*
TCP/UDP checksum offload IPV6	启用 TX/RX、启用 TX 或启用 RX 或禁用 offload。	默认启用 RX 和 TX。
发送缓冲区（0 = 自动）	50 的增量。	默认自动。
虚拟机队列	启用或禁用。	默认启用。
VLAN ID	用户可配置的数字。	默认 0。

事件日志消息

表 22 提供了由 Windows NDIS 驱动程序记录到事件日志中的事件日志消息。

表 22：Windows 事件日志消息

消息 ID	*注释*
0x0001	内存分配失败。
0x0002	检测到链路断开。
0x0003	检测到链路连通。
0x0009	链路 1000 满载。
0x000A	链路 2500 满载。
0x000b	初始化成功。
0x000c	微型端口复位。
0x000d	初始化失败。
0x000E	链路 10Gb 成功。
0x000F	驱动程序层绑定失败。
0x0011	设置属性失败。
0x0012	扩散聚集 DMA 失败。
0x0013	默认队列初始化失败。
0x0014	固件版本不兼容。
0x0015	单一中断。

表 23：事件日志消息

0x0016	固件未能在分配的时间内响应。
0x0017	固件返回失败状态。
0x0018	固件状态未知。
0x0019	不支持光学模块。
0x001A	端口 1 和端口 2 之间选择的速度不兼容。报告的链路速度正确，但可能与“速度和双工”设置不匹配。
0x001B	端口 1 和端口 2 之间选择的速度不兼容。链路配置不合法。
0x001C	为 25Gb 全双工链路配置网络控制器。
0x0020	RDMA 支持初始化失败。

表 23：事件日志消息（待续）

0x0021	设备的 RDMA 固件与该驱动程序不兼容。
0x0022	门铃 BAR 对于 RDMA 来说太小。
0x0023	设备重置后 RDMA 重启失败。
0x0024	系统启动后 RDMA 重启失败
0x0025	RDMA 启动失败。资源不足。
0x0026	固件中未启用 RDMA。
0x0027	启动失败，未设置 MAC 地址。
0x0028	检测到发送失速。现在开始禁用发送流控制。

组合

Windows

安装在 Dell 平台上的 Broadcom NetXtreme-E 设备可以使用 Microsoft 组合解决方案参与 NIC 组合功能。参阅以下链接中描述的 Microsoft 公共文档：
https://www.microsoft.com/en-us/download/details.aspx?id=40319

Microsoft LBFO 是可以在 Windows 操作系统中使用的原生组合驱动程序。该组合驱动程序还提供 VLAN 标记功能。

Linux

Linux bonding 用于 Linux 下的组合。该概念是加载 bonding 驱动程序并向绑定中添加组员，该绑定将对通讯流量进行负载平衡。

使用下列步骤设置 Linux bonding：

1. 执行以下命令：

   `modprobe bonding mode="balance-alb"`。这会创建一个绑定接口。
2. 向绑定接口中添加绑定客户端。下面显示一个示例：

   ```
   ifenslave bond0 ethX; ifenslave bond0 ethY
   ```
3. 使用 `ifconfig bond0 IPV4Address netmask NetMask up` 分配一个 IPV4 地址以绑定接口。IPV4Address 和 NetMask 是 IPV4 地址和关联的网络掩码。

   注： `IPV4 address` 应替换为实际网络 IPV4 地址。`NetMask` 应替换为实际 IPV4 网络掩码。
4. 使用 `ifconfig bond0 IPV6Address netmask NetMask up` 分配一个 IPV6 地址以绑定接口。IPV6Address 和 NetMask 是 IPV6 地址和关联的网络掩码。

   注： `IPV6 address` 应替换为实际网络 IPV6 地址。`NetMask` 应替换为实际 IPV6 网络掩码。

参阅 Linux Bonding 文档，了解高级配置。

系统级配置

参阅以下各节，了解有关系统级 NIC 配置的信息。

UEFI HII 菜单

可以使用 HII（人机界面）菜单配置 Broadcom NetXtreme-E 系列控制器，进行预启动、iscsi 和 SR-IOV 等高级配置。

要配置这些设置，请在系统引导过程中选择 F2 -> System Setup（系统设置） -> Device Settings（设备设置）。选择所需的网络适配器，查看和更改配置。

主配置页面

此页面显示适配器和以太网设备的当前网络链路状态、PCI-E Bus:Device:Function、MAC 地址。

10GBaseT 卡允许用户启用或禁用节能以太网 (EEE)。

固件图像属性

主配置页面 -> 固件映像属性显示系列版本，包括控制器 BIOS、Multi Boot Agent (MBA)、UEFI、iSCSI 的版本号和 Comprehensive Configuration Management (CCM) 版本号。

设备级别配置

主配置页面 -> 设备级配置允许用户启用 SR-IOV 模式、每个物理功能的虚拟功能数、每个虚拟功能的 MSI-X 矢量和物理功能 MSI-X 矢量的最大数量。

NIC 配置

NIC 配置 -> 传统引导协议用来选择和配置 PXE、iSCSI 或禁用传统引导模式。引导程序类型可以是自动、int18h（中断 18h）、int19h（中断 19h）或 BBS。

MBA 和 iSCSI 也可以使用 CCM 进行配置。传统 BIOS 模式使用 CCM 进行配置。隐藏设置提示可用于禁用或启用标志显示。

可以启用或禁用 PXE 的 VLAN，VLAN ID 可由用户配置。请参阅 第 40 页上的“自动协商配置” 以了解链路速度设置选项的详细信息。

iSCSI 配置

iSCSI 引导配置可通过主配置页面 -> iSCSI 配置进行设置。IPV4 或 IPV6、iSCSI 启动程序或 iSCSI 目标等参数可通过此页面进行设置。

参阅第 49 页上的“iSCSI 引导”，了解详细的配置信息。

Comprehensive Configuration Management

预引导配置可使用 Comprehensive Configuration Management (CCM) 菜单选项进行配置。在系统 BIOS POST 过程中，将会显示 Broadcom 标志消息，并提供通过 Control-S 菜单更改参数的选项。当按下 Control-S 时，设备列表将会包含系统中找到的所有 Broadcom 网络适配器。选择所需的 NIC 进行配置。

设备硬件配置

使用此节可以配置的参数与 HII 菜单“设备级配置”相同。

MBA 配置菜单

使用此节可以配置的参数与 HII 菜单“NIC 配置”相同。

iSCSI 引导主菜单

使用此节可以配置的参数与 HII 菜单“iSCSI 配置”相同。

自动协商配置

注：在一个端口由多个 PCI 功能共享的 NPAR（NIC 分区）设备中，端口速度是预先配置的，不能由驱动程序更改。

Broadcom NetXtreme-E 控制器支持以下自动协商功能：

- 链路速度自动协商
- 暂停/流控制自动协商
- FEC – 前向纠错自动协商

注：对于链路速度自动协商，在使用 SFP+、SFP28 连接器时，可以使用支持自动协商的 DAC 或多模光纤收发器。确保链路伙伴端口已设置为匹配自动协商协议。例如，如果本地 Broadcom 端口已设置为 IEEE 802.3by 自动协商协议，则链路伙伴必须支持自动协商且必须设置为 IEEE 802.3by 自动协商协议。

注：对于双端口 NetXtreme-E 网络控制器，链路速度不支持 10 Gbps 和 25 Gbps 的组合。

双端口 NetXtreme-E 网络控制器支持的链路速度设置组合显示在第 41页上的 表 24。

表 24：支持的 Link Speed 设置组合

Port1 Link Speed Setting	Port 2 Link Setting									
	Forced 1G	**Forced 10G**	**Forced 25G**	**AN Enabled {1G}**	**AN Enabled {10G}**	**AN Enabled {25G}**	**AN Enabled {1/10G}**	**AN Enabled {1/25G}**	**AN Enabled {10/25G}**	**AN Enabled {1/10/25G}**
Forced 1G	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN
	P2: no AN	P2: no AN	P2: no AN	P2: {1G}	P2: AN {10G}	P2: AN {25G}	P2: AN {1/10G}	P2: AN {1/25G}	P2: AN {10/25G}	P2: AN {1/10/25G}
Forced 10G	P1: no AN	P1: no AN	Not supported	P1: no AN	P1: no AN	Not supported	P1: no AN	P1: no AN	P1: no AN	P1: no AN
	P2: no AN	P2: no AN		P2: {1G}	P2: {10G}		P2: AN {1/10G}	P2: AN {1G}	P2: AN {10G}	P2: AN {1/10G}
Forced 25G	P1: no AN	Not supported	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN	P1: no AN
	P2: no AN		P2: no AN	P2: no AN	P2: no AN	P2: no AN	P2: AN {1G}	P2: AN {1/25G}	P2: AN {25G}	P2: AN {1/25G}
AN Enabled {1G}	P1: {1G}	P1: {1G}	P1: {1G}	P1: AN {1G}	P1: AN {1G}	P1: AN {1G}	P1: AN {1G}	P1: AN {1G}	P1: AN {1G}	P1: AN {1G}
	P2: no AN	P2: no AN	P2: no AN	P2: AN {1G}	P2: AN {10G}	P2: AN {25G}	P2: AN {1/10G}	P2: AN {1/25G}	P2: AN {10/25G}	P2: AN {1/10/25G}
AN Enabled {10G}	P1: AN {10G}	P1: AN {25G}	Not supported	P1: AN {10G}	P1: AN {10G}	Not supported	P1: AN {25G}	P1: AN {10G}	P1: AN {10G}	P1: AN {10G}
	P2: no AN	P2: no AN		P2: AN {1G}	P2: AN {10G}		P2: AN {1G}	P2: AN {1G}	P2: AN {10G}	P2: AN {1/10G}
AN Enabled {25G}	P1: AN {25G}	Not supported	P1: AN {25G}	P1: AN {25G}	Not supported	P1: AN {25G}	P1: AN {1/10G}	P1: AN {25G}	P1: AN {25G}	P1: AN {25G}
	P2: no AN		P2: no AN	P2: AN {1G}		P2: AN {25G}	P2: AN {1/10G}	P2: AN {1/25G}	P2: AN {25G}	P2: AN {1/25G}
AN Enabled {1/10G}	P1: AN {1/10G}	P1: AN {1/10G}	P1: AN {1G}	P1: AN {1/10G}	P1: AN {1/10G}	P1: AN {1/10G}	P1: AN {1/25G}	P1: AN {1G}	P1: AN {1/10G}	P1: AN {1/10G}
	P2: no AN	P2: no AN	P2: no AN	P2: AN {1G}	P2: AN {10G}	P2: AN {25G}	P2: AN {1/10G}	P2: AN {1G}	P2: AN {10G}	P2: AN {1/10G}
AN Enabled {1/25G}	P1: AN {1/25G}	P1: {1G}	P1: AN {1/25G}	P1: AN {1/25G}	P1: AN {1G}	P1: AN {1/25G}	P1: AN {10/25G}	P1: AN {1/25G}	P1: AN {1/25G}	P1: AN {1/25G}
	P2: no AN	P2: no AN	P2: no AN	P2: AN {1G}	P2: AN {10G}	P2: AN {25G}	P2: AN {1/10G}	P2: AN {1/25G}	P2: AN {25G}	P2: AN {1/25G}
AN Enabled {10/25G}	P1: AN {10/25G}	P1: {10G}	P1: AN {25G}	P1: AN {10/25G}	P1: AN {10G}	P1: AN {25G}	P1: AN {1/10/25G}	P1: AN {25G}	P1: AN {10/25G}	P1: AN {10/25G}
	P2: no AN	P2: no AN	P2: no AN	P2: AN {1G}	P2: AN {10G}	P2: AN {25G}	P2: AN {1/10G}	P2: AN {25G}	P2: AN {10/25G}	P2: AN {1/10/25G}
AN Enabled {1/10/25G}	P1: AN {1/10/25G}	P1: {1/10G}	P1: AN {1/25G}	P1: AN {1/10/25G}	P1: AN {1/10G}	P1: AN {1/25G}	P1: AN {1/10/25G}	P1: AN {1/25G}	P1: AN {1/10/25G}	P1: AN {1/10/25G}
	P2: no AN	P2: no AN	P2: no AN	P2: AN {1G}	P2: AN {10G}	P2: AN {25G}	P2: AN {1/10G}	P2: AN {1/25G}	P2: AN {10/25G}	P2: AN {1/10/25G}

注：使用支持的光纤收发器或直连铜缆实现 1 Gbps linkspeed。

- P1 - 端口 1 设置
- P2 - 端口 2 设置
- AN - 自动协商
- 不可 AN - 强制速度

- {链路速度} - 预期链路速度
- 自动协商 {链路速度} - 需要受支持的自动协商链路速度。

预期的链路速度基于表 25 中显示的本地和链路伙伴设置。

表 25：预期的链路速度

Local Speed Settings	Link Partner Speed Settings									
	Forced 1G	Forced 10G	Forced 25G	AN Enabled {1G}	AN Enabled {10G}	AN Enabled {25G}	AN Enabled {1/10G}	AN Enabled {1/25G}	AN Enabled {10/25G}	AN Enabled {1/10/25G}
Forced 1G	1G	No link	No link	No link	No link	No link	No link	No link	No link	No link
Forced 10G	No link	10G	No link	No link	No link	No link	No link	No link	No link	No link
Forced 25G	No link	No link	25G	No link	No link	No link	No link	No link	No link	No link
AN {1G}	No link	No link	No link	1G	No link	No link	1G	1G	No link	1G
AN {10G}	No link	No link	No link	No link	10G	No link	10G	No link	10G	10G
AN {25G}	No link	No link	No link	No link	No link	25G	No link	25G	25G	25G
AN {1/10G}	No link	No link	No link	1G	10G	No link	10G	1G	10G	10G
AN {1/25G}	No link	No link	No link	1G	No link	25G	1G	25G	25G	25G
AN {10/25G}	No link	No link	No link	No link	10G	25G	10G	25G	25G	25G
AN {1/10/25G}	No link	No link	No link	1G	10G	25G	10G	25G	25G	25G

注：此版本当前不支持 SFP+/SFP28 的 1 Gbps 链路速度。

要启用链路速度自动协商，可在系统 BIOS HII 菜单或 CCM 中启用以下选项：

注：如果通过光纤收发器模块/电缆连接，在启用链路速度自动协商时，则必须启用介质自动检测。

系统 BIOS -> 设备设置 -> NetXtreme-E NIC -> 设备级配置

操作链路速度

此选项配置预启动（MBA 和 UEFI）驱动程序（Linux、ESX）、操作系统驱动程序和固件使用的链路速度。此设置将被操作系统当前状态下的驱动程序设置覆盖。Windows 驱动程序 (bnxtnd_.sys) 使用驱动程序 .inf 文件中的 linkspeed 设置。

固件链路速度

此选项配置的链路速度在设备处于 D3 中时供固件使用。

自动协商协议

这是受支持的自动协商协议，在与链路伙伴协商链路速度时使用。此选项必须与链路伙伴端口中的自动协商协议设置相匹配。Broadcom NetXtreme-E NIC 支持以下自动协商协议：IEEE 802.3by、25G/50G 联盟和 25G/50G BAM。默认情况下，此选项设置为 IEEE 802.3by 并回归 25G/50G 联盟。

链路速度和流控制/暂停必须在主机操作系统的驱动程序中进行配置。

Windows 驱动程序设置

要访问 Windows 驱动程序设置：

打开 **Windows 设备管理器 -> Broadcom NetXtreme E 系列适配器 -> 高级属性 ->“高级”选项卡**

流控制 = 自动协商

即启用流控制/暂停帧自动协商。

速度和双工 = 自动协商

即启用链路速度自动协商。

Linux 驱动程序设置

注：对于 10GBase-T NetXtreme-E 网络适配器，必须启用自动协商。

25G 广播是首次在 4.7 核的 ethtool 接口中定义的更新标准。要使自动协商完全支持这些新广播速度，需要 4.7（或更新版本）核和更新版本的 ethtool 实用程序（版本 4.8）。

ethtool -s eth0 speed 25000 autoneg off

此命令将关闭自动协商，并强制将链路速度设置为 25 Gbps。

ethtool -s eth0 autoneg on advertise 0x0

此命令将启用自动协商，且需要设备支持以下所有速度：1G、10G、25G。

以下是受支持的所需速度。

- 0x020 – 1000baseT Full
- 0x1000 – 10000baseT Full
- 0x80000000 – 25000baseCR Full

ethtool -A eth0 autoneg on|off

使用此命令启用/禁用“暂停帧”自动协商。

ethtool -a eth0

使用此命令显示当前流控制自动协商设置。

ESXi 驱动程序设置

注：对于 10GBase-T NetXtreme-E 网络适配器，必须启用自动协商。在 10GBase-T 适配器上使用强制速度将导致 esxcli 命令失败。

注：VMWare ESX6.0 不支持 25G 速度。这此情况下，使用第二实用程序 (BNXTNETCLI) 设置 25G 速度。ESX6.0U2 支持 25G 速度。

$ esxcli network nic get -n <iface>

此命令显示当前速度、双工、驱动程序版本、固件版本和链路状态。

$ esxcli network nic set -S 10000 -D full -n <iface>

此命令将速度强制设置为 10 Gbps。

$ esxcli network nic set -a -n <iface>

这将在接口 <iface> 上启用 linkspeed 自动协商。

$ esxcli network nic pauseParams list

使用此命令获取“暂停参数”列表。

$ esxcli network nic pauseParams set --auto <1/0> --rx <1/0> --tx <1/0> -n <iface>

使用此命令设置“暂停参数”。

注：只有在链路速度自动协商模式下配置接口时，才能设置流控制/暂停自动协商。

FEC 自动协商

要启用/禁用链路 FEC 自动协商，可在系统 BIOS HII 菜单或 CCM 中启用以下选项：

- **系统 BIOS -> 设备设置 -> NetXtreme-E NIC -> 设备级配置**

在协商交换过程中，FEC 自动协商使用两个参数：FEC capable 和 FEC request。

如果 NIC 将其作为 FEC 自动协商功能播发，则 FEC 设置由 Switch 驱动。随后，即可与启用 FEC 的交换机或禁用 FEC 的交换机连通。

示例：

– switch – capable=1, request = 1 then, the link is a FEC link.

– switch – capable= N/A, request= 0 = then the FEC is disabled.

对于 NetXtreme-E Ethernet 控制器，仅支持 Base-R FEC (CL74)。表 26 显示了带有链路伙伴的所有支持配置。

表 26：适用于 BCM5730X/BCM5740X 的 FEC 支持配置

本地 SFEC 设置	链路伙伴 FEC 设置			
	Force Speed No FEC	**Force Speed Base-R FEC CL74**	**AN（无）**	**AN（无，Base-R）**
Force Speed No FEC	链路无 FEC	无链路	无链路	无链路
Force Speed Base-R FEC CL74	无链路	Base-R FEC CL74	无链路	无链路
AN（无）	无链路	无链路	链路无 FEC	Base-R FEC CL74
AN（无，Base-R）	无链路	无链路	Base-R FEC CL74	Base-R FEC CL74

注：对于强制速度而言，两侧的速度设置必须相同。

注：AN {无} 表示 AN 播发 Base-R 功能位。在 IEEE802.3by 上设置 F0 位，在联盟上设置 F2 位。

注：AN {无，Base-R} 表示 AN 播发 Base-R 功能位和请求位。在 IEEE802.3by 上设置 F0 和 F1 位，在联盟上设置 F2 和 F4 位。

对于 NetXtreme-E Ethernet 控制器而言，FEC 支持 Base-R FEC (CL74) 和 RS-FEC (CL91/CL108)。表 27 显示了带有链路伙伴的所有支持配置。

表 27：适用于 BCM5741X 的 FEC 支持配置

本地 FEC 设置	链路伙伴 FEC 设置					
	Force Speed No FEC	**Force Speed Base-R FEC CL74**	**Force Speed RS-FEC CL91/ CL108**	**AN（无）**	**An （无，Base-R）**	**AN（无，Base-R，RS）**
Force No FEC	链路无 FEC	无链路	无链路	无链路	无链路	无链路
Force Speed Base-R FEC CL74	无链路	Base-R FEC CL74	无链路	无链路	无链路	无链路
强制 RS-FEC CL91/CL108	无链路	无链路	RS-FEC CL91/ CL108	无链路	无链路	无链路
AN（无）	无链路	无链路	无链路	链路无 FEC	Base-R FEC CL74	RS-FEC CL91/CL108
AN（无，Base-R）	无链路	无链路	无链路	Base-R FEC CL74	Base-R FEC CL74	RS-FEC CL91/CL108
AN（无，Base-R，RS）	无链路	无链路	无链路	RS-FEC CL91/CL108	RS-FEC CL91/ CL108	RS-FEC CL91/CL108

注：对于强制速度而言，两侧的速度设置必须相同。

注：AN {无} 表示 AN 播发 Base-R 功能位。在 IEEE802.3by 上设置 F0 位，在联盟上设置 F2 位。

注：AN {无，Base-R} 表示 AN 播发 Base-R 功能位和请求位。在 IEEE802.3by 上设置 F0 和 F1 位，在联盟上设置 F2 和 F4 位。

链路训练

链路训练允许两个端点、Broadcom 适配器和另一侧调整功率设置以及其他调节参数，充分发挥两个设备间通信通道的可靠性和效率。目标是消除对于在不同电缆长度和类型间进行通道特定调节的需求。链路训练先于自动协商针对 CR/KR 速度执行。启用自动协商时，可操作链路训练。如果链路训练不能实现与链路伙伴连通，链路策略将自动禁用链路训练。此链路策略可确保与不支持链路训练的链路伙伴兼容。

表 28 显示了介质类型与 BCM5730X、BCM5740X 和 BCM5741X Ethernet 控制器之间的关系。

表 28：介质类型与速度之间的链路训练关系

本地和介质电缆类型	链路伙伴链路训练设置		
	Force Speed Link Training Disabled	**Force Speed Link Training Enabled**	**AN（自动链路训练）**
Force Speed DAC (SFP+/SFP28/ QSFP28)	链路无链路训练	链路有链路训练	链路有链路训练
Force Speed optical (Transceiver/AOC)	链路无链路训练	链路无链路训练	无
AN DAC (SFP+SFP28/QSFP28)	无链路	无链路	链路有链路训练

介质自动检测

由于 SerDes 不支持并行检测，所以，固件实施了一种用于增强链路检测的方法，称为**介质自动检测**。此功能受 CCM/HII 控制，如图 13 所示。

启用**介质自动检测**时，链路策略在状态机之后（参见图 13）与链路伙伴建立链路。此行为取决于介质类型。对于 DAC 电缆而言，该方法回归其他强制模式，并且在链路连通时停止。

注： 以 25G 或 10GbE 速率执行自动协商时，不支持并行检测。这表示，如果一侧执行自动协商，而另一侧不执行，则无法连通链路。

图 13：适用于介质自动检测功能的状态机

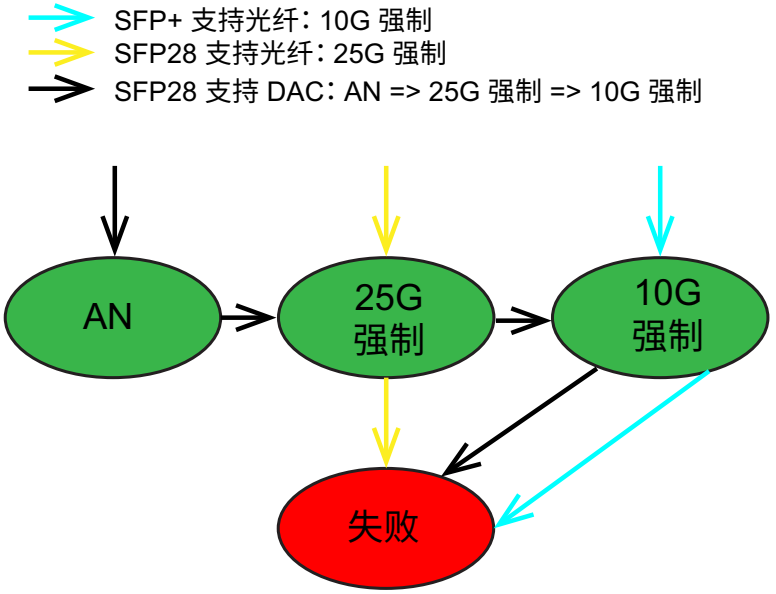


表 29 和表 30 显示了启用**介质自动检测**的链路结果。

表 29：适用于 BCM5730X 和 BCM5740X 的介质自动检测

		链路伙伴链路训练设置	
链路伙伴设置		介质自动检测	
速度	FEC	无 FEC	Base-R FEC
10G	无 FEC	链路	链路支持 Base-R
	Base-R	无链路	链路支持 Base-R
25G	无 FEC	链路	链路
	Base-R	无链路	链路支持 Base-R
AN	禁用	链路	链路
	自动 FEC	链路	链路支持 Base-R
	Base-R	链路支持 Base-R	链路支持 Base-R

表 30：适用于 BCM5741X 的介质自动检测

		链路伙伴链路训练设置		
链路伙伴设置		介质自动检测		
速度	FEC	无 FEC	Base-R FEC	RS-FEC
10G	无 FEC	链路	链路支持 Base-R	无链路
	Base-R	无链路	链路支持 Base-R	无链路
	RS	无链路	无链路	链路支持 RS-FEC
25G	无 FEC	链路	无链路	无链路
	Base-R	无链路	链路支持 Base-R	无链路
	RS	无链路	无链路	链路支持 RS-FEC
AN	禁用	链路	链路	链路
	自动 FEC	链路	链路支持 Base-R	链路支持 RS-FEC
	Base-R	链路支持 Base-R	链路支持 Base-R	链路支持 RS-FEC
	RS	链路支持 RS	链路支持 RS-FEC	链路支持 RS-FEC

ISCSI 引导

Broadcom NetXtreme-E 以太网适配器支持 iSCSI 引导，从而实现无盘系统的操作系统网络引导。iSCSI 引导允许 Windows、Linux 或 VMware 操作系统通过标准 IP 网络从位于远程的 iSCSI 目标计算机引导。

适用于 iSCSI 引导的支持的操作系统

Broadcom NetXtreme-E Gigabit 以太网适配器支持以下操作系统上的 iSCSI 引导：

- Windows Server 2012 及更高版本 64 位
- Linux RHEL 7.1 及更高版本，SLES11 SP4 或更高版本
- VMware 6.0 U2

设置 iSCSI 引导

参阅以下各节，了解有关设置 iSCSI 引导的信息。

配置 iSCSI 目标

注：Windows 2016（或早期版本）自带的驱动程序可能已经过期，仅支持 Forced Linkspeed。为避免无法连通不匹配链路伙伴等安装问题，强烈建议用户使用最新版本的 NIC 驱动程序（支持 Linkspeed 自动协商）自定义 Windows 安装映像/介质。

配置 iSCSI 目标视目标供应商而异。有关配置 iSCSI 目标的信息，请参阅供应商提供的文档。一般步骤包括：

1. 创建一个 iSCSI 目标。
2. 创建一个虚拟盘。
3. 将虚拟盘映射到在第 49页上的步骤 1 中创建的 iSCSI 目标。
4. 将 iSCSI 启动程序与 iSCSI 目标关联。
5. 记下 iSCSI 目标名称、TCP 端口号、iSCSI 逻辑单元号 (LUN)、启动程序 Internet 限定名称 (IQN) 和 CHAP 身份验证详细信息。
6. 配置 iSCSI 目标之后，获取以下信息：
   - 目标 IQN
   - 目标 IP 地址
   - 目标 TCP 端口号
   - 目标 LUN
   - 启动程序 IQN
   - CHAP ID 和密钥

配置 iSCSI 引导参数

配置 Broadcom iSCSI 引导软件以获得静态或动态配置。有关“常规参数”菜单上提供的配置选项的信息，请参阅表 31。表 31 列出了适用于 IPv4 和 IPv6 的参数。指出了特定于 IPv4 或 IPv6 的参数。

表 31：配置选项

选项	*描述*
通过 DHCP 配置 TCP/IP 参数	该选项特定于 IPv4。控制 iSCSI 引导主机软件是使用 DHCP 获得 IP 地址信息（启用）还是使用静态 IP 配置（禁用）。
IP 自动配置	该选项特定于 IPv6。控制显示并使用了 DHCPv6（启用）时，iSCSI 引导主机软件是否配置无状态链接本地地址和/或有状态地址。Router Solicit 数据包在每次重试间，每隔 4 秒最多发送 3 次。或使用静态 IP 配置（禁用）。
通过 DHCP 配置 iSCSI 参数	控制 iSCSI 引导主机软件是使用 DHCP 获得其 iSCSI 目标参数（启用）还是通过静态配置（禁用）。通过 iSCSI 启动程序参数配置屏幕输入静态信息。
CHAP 身份验证	控制 iSCSI 引导主机软件在连接到 iSCSI 目标时是否使用 CHAP 身份验证。如果启用了 CHAP 身份验证，可通过 iSCSI 启动程序参数配置屏幕输入 CHAP ID 和 CHAP 密钥。
DHCP 供应商 ID	控制 iSCSI 引导主机软件如何解释在 DHCP 期间使用的 Vendor Class ID 字段。如果 DHCP Offer 数据包中的 Vendor Class ID 字段与该字段的值匹配，iSCSI 引导主机软件将进一步查看 DHCP 选项 43 字段以获得所需的 iSCSI 引导扩展。如果禁用 DHCP，不必设置此值。
链路连通延迟时间	控制 iSCSI 引导主机软件从建立 Ethernet 链路之后到通过网络发送数据所等待的时间（以秒为单位）。有效值为 0 至 255。例如，如果客户端系统的交换机接口上启用了某个网络协议（如生成树），用户可能需要为该选项设置值。
使用 TCP 时间戳	控制 TCP 时间戳选项是否启用。
目标为第一个 HDD	允许指定 iSCSI 目标驱动器作为系统中的第一个硬盘驱动器。
LUN 忙时重试次数	控制 iSCSI 引导启动程序在 iSCSI 目标 LUN 忙时将会尝试的连接重试次数。
IP 版本	该选项特定于 IPv6。在 IPv4 或 IPv6 协议间切换。从一个协议版本切换到另一个协议版本时，所有 IP 设置都将丢失。

MBA 引导协议配置

配置引导协议：

1. 重新启动系统。
2. 从 PXE 标志中，选择 **CTRL+S**。此时会显示 **MBA 配置菜单**。
3. 从 **MBA 配置菜单**中，使用**上箭头**或**下箭头**移至引导协议选项。使用**左箭头**或**右箭头**更改 iSCSI 的引导协议选项。
4. 从**主菜单**中选择 **iSCSI** 引导配置。

iSCSI 引导配置

有两种配置 iSCSI 引导的方法：

- 静态 iSCSI 引导配置。
- 动态 iSCSI 引导配置。

静态 iSCSI 引导配置

在静态配置中，必须为系统的 IP 地址、系统的启动程序 IQN 和在第 49 页上的“配置 iSCSI 目标”中获得的目标参数输入数据。要了解关于配置选项的信息，请参见第 50页上的 表 31。

使用静态配置来配置 iSCSI 引导参数：

1. 从**常规参数**菜单，设置以下各项：
   - 通过 DHCP 配置 TCP/IP 参数 – 禁用。（对于 IPv4。）
   - IP 自动配置 – 禁用。（对于 IPv6、non-offload。）
   - 通过 DHCP 配置 iSCSI 参数 – 禁用
   - CHAP 身份验证 – 禁用
   - DHCP 供应商 ID – BRCM ISAN
   - 链路连通延迟时间 – 0
   - 使用 TCP 时间戳 – 启用（对于 Dell/EMC AX100i 等目标，有必要启用“使用 TCP 时间戳”）
   - 目标为第一个 HDD – 禁用
   - LUN 忙时重试次数 – 0
   - IP 版本 – IPv6。（对于 IPv6、non-offload。）
2. 选择 **Esc** 返回到**主**菜单。
3. 从**主**菜单中，选择**启动程序参数**。
4. 从启动程序参数屏幕中，输入以下各项值：
   - IP 地址（未指定的 IPv4 和 IPv6 地址应分别为“0.0.0.0”和“::”）
   - 子网掩码前缀
   - 默认网关
   - 主 DNS
   - 从属 DNS
   - iSCSI 名称（与客户端系统要使用的 iSCSI 启动程序名称相对应）

注：输入 IP 地址。不对 IP 地址执行错误检查以查看重复或错误的段/网络分配。

5. 选择 **Esc** 返回到**主**菜单。

6. 从**主**菜单中，选择**第一个目标参数**。

注：对于初始设置，不支持配置第二个目标。

7. 从**第一个目标参数**屏幕中，启用**连接，**以连接到 iSCSI 目标。使用配置 iSCSI 目标时使用的值为以下各项输入值：
   - IP 地址
   - TCP 端口
   - 引导 LUN
   - iSCSI 名称
8. 选择 **Esc** 返回到**主**菜单。
9. 选择 **Esc** 并选择退出并保存配置。
10. 选择 **F4** 以保存 MBA 配置。

动态 iSCSI 引导配置

在动态配置中，指定系统的 IP 地址和目标/启动程序信息由 DHCP 服务器提供（请参见第 54 页上的“配置 DHCP 服务器以支持 iSCSI 引导”中的 IPv4 和 IPv6 配置）。对于 IPv4，除了启动程序 iSCSI 名称之外，启动程序参数、第一个目标参数或第二个目标参数屏幕上的任何设置均被忽略，不需要清除。对于 IPv6，除了 CHAP ID 和密钥之外，启动程序参数、第一个目标参数或第二个目标参数屏幕上的任何设置均被忽略，不需要清除。要了解关于配置选项的信息，请参见第 50页上的 表 31。

注：使用 DHCP 服务器时，DNS 服务器条目将被 DHCP 服务器提供的值覆盖。即使本地提供的值有效并且 DHCP 服务器不提供 DNS 服务器信息，也可能出现这种情况。当 DHCP 服务器不提供 DNS 服务器信息时，主 DNS 和辅助 DNS 服务器值均设置为 0.0.0.0。当 Windows 操作系统获得控制权时，Microsoft iSCSI 启动程序将检索 iSCSI 启动程序参数并静态配置相应的注册表。这将覆盖配置的内容。由于 DHCP 守护进程在 Windows 环境中作为一个用户进程运行，在堆栈在 iSCSI 引导环境中产生之前必须静态配置所有 TCP/IP 参数。

- 如果使用了 DHCP 选项 17，目标信息将由 DHCP 服务器提供，并且从启动程序参数屏幕编入的值中检索启动程序 iSCSI 名称。如果未选择任何值，控制器默认为名称：

  ```
  iqn.1995-05.com.broadcom.<11.22.33.44.55.66>.iscsiboot
  ```

  其中的字符串 **11.22.33.44.55.66** 对应于控制器的 MAC 地址。
- 如果使用了 DHCP 选项 43（仅 IPv4），启动程序参数、第一个目标参数或第二个目标参数屏幕上的任何设置均被忽略，不需要清除。

使用动态配置来配置 iSCSI 引导参数：

1. 从**常规参数**菜单屏幕中，设置以下参数：
   - 通过 DHCP 配置 TCP/IP 参数 – 启用。（对于 IPv4。）
   - IP 自动配置 – 启用。（对于 IPv6、non-offload。）
   - 通过 DHCP 配置 iSCSI 参数 – 启用
   - CHAP 身份验证 – 禁用
   - DHCP 供应商 ID – BRCM ISAN
   - 链路连通延迟时间 – 0
   - 使用 TCP 时间戳 – 启用（对于 Dell/EMC AX100i 等目标，有必要启用“使用 TCP 时间戳”）
   - 目标为第一个 HDD – 禁用
   - LUN 忙时重试次数 – 0
   - IP 版本 – IPv6。（对于 IPv6、non-offload。）
2. 选择 **Esc** 返回到**主**菜单。

注： 启动程序参数和第一个目标参数屏幕上的信息均将忽略，不需要清除。

3. 选择退出并保存配置。

启用 CHAP 身份验证

确保在目标上启用了 CHAP 身份验证。

启用 CHAP 身份验证：

1. 从**常规参数**屏幕中，将 **CHAP 身份验证**设置为**启用**。
2. 从**启动程序参数**屏幕中，输入以下各项参数：
   - CHAP ID（最多 128 字节）
   - CHAP 密钥（如果需要身份验证，长度必须为 12 个字符或更长）
3. 选择 **Esc** 返回到**主**菜单。
4. 从**主**菜单中，选择第一个目标参数。
5. 从第一个目标参数屏幕中，使用配置 iSCSI 目标时使用的值为以下各项输入值：
   - CHAP ID（双向 CHAP 时可选）
   - CHAP 密钥（双向 CHAP 时可选，长度必须为 12 个字符或更长）
6. 选择 **Esc** 返回到**主**菜单。
7. 选择 **Esc** 并选择退出并保存配置。

配置 DHCP 服务器以支持 iSCSI 引导

DHCP 服务器是可选组件，只有动态 iSCSI 引导配置设置才需要它（请参见第 52 页上的“动态 iSCSI 引导配置”）。

配置 DHCP 服务器以支持 iSCSI 引导对于 IPv4 和 IPv6 是不同的。参阅以下各节：

IPv4 的 DHCP iSCSI 引导配置

DHCP 协议包括许多为 DHCP 客户端提供配置信息的选项。对于 iSCSI 引导，Broadcom 适配器支持以下 DHCP 配置：

DHCP 选项 17，根路径

选项 17 用于将 iSCSI 目标信息传递给 iSCSI 客户端。IETC RFC 4173 中定义的根路径的格式为：

```
iscsi:"<servername>":"<protocol>":"<port>":"<LUN>":"<targetname>
```

参数在表 32 中定义。

表 32：DHCP 选项 17 参数定义

参数	*定义*
"iscsi:"	字符串。
<servername>	iSCSI 目标的 IP 地址或 FQDN
":"	分隔符。
<protocol>	用于访问 iSCSI 目标的 IP 协议。目前，仅支持 TCP，因此协议为 6。
<port>	与协议关联的端口号。iSCSI 的标准端口号为 3260。
<LUN>	要在 iSCSI 目标上使用的逻辑单元号。LUN 值必须以十六进制形式表示。ID 为 64 的 LUN 必须在 DHCP 服务器上配置为选项 17 参数内的 40。
<targetname>	目标名称为 IQN 或 EUI 格式（有关 IQN 和 EUI 格式的详细信息，请参见 RFC 3720）。示例 IQN 名称为“iqn.1995-05.com.broadcom:iscsi-target”。

DHCP 选项 43，供应商特定信息

与 DHCP 选项 17 相比，DHCP 选项 43（供应商特定信息）提供给 iSCSI 客户端的配置选项更多。在此配置中，提供了三个附加子选项，用于为 iSCSI 引导客户端分配启动程序 IQN，以及分配另外两个可用于引导的 iSCSI 目标 IQN。iSCSI 目标 IQN 的格式与 DHCP 选项 17 的格式相同，而 iSCSI 启动程序 IQN 就是该启动程序的 IQN。

注：DHCP 选项 43 仅在 IPv4 上受支持。

这些子选项在下面列出。

表 33：DHCP 选项 43 子选项定义

子选项	*定义*
201	采用标准根路径格式的第一条 iSCSI 目标信息 iscsi:"<servername>":"<protocol>":"<port>":"<LUN>":"<targetname>
203	iSCSI 启动程序 IQN

与 DHCP 选项 17 相比，使用 DHCP 选项 43 需要更多配置，但它提供更丰富的环境和更多配置选项。Broadcom 建议客户在执行动态 iSCSI 引导配置时使用 DHCP 选项 43。

配置 DHCP 服务器

配置 DHCP 服务器以支持选项 17 或选项 43。

注：如果使用选项 43，则配置选项 60。选项 60 的值应该与 DHCP 供应商 ID 值匹配。DHCP 供应商 ID 值为 BRCM ISAN，如 iSCSI 引导配置菜单的常规参数中所示。

IPv6 的 DHCP iSCSI 引导配置

DHCPv6 服务器可提供多个选项，包括无状态或有状态 IP 配置以及向 DHCPv6 客户端发送信息。对于 iSCSI 引导，Broadcom 适配器支持以下 DHCP 配置：

注：DHCPv6 标准根路径选项尚不可用。Broadcom 建议使用选项 16 或选项 17 以获得动态 iSCSI 引导 IPv6 支持。

DHCPv6 选项 16，Vendor Class 选项

DHCPv6 选项 16（Vendor Class 选项）必须出现，且必须包含与所配置的 DHCP 供应商 ID 参数相匹配的字符串。DHCP 供应商 ID 值为 BRCM ISAN，如 iSCSI 引导配置菜单的常规参数中所示。

选项 16 的内容应为 <2 字节长度> <DHCP 供应商 ID>。

DHCPv6 选项 17，供应商特定信息

DHCPv6 选项 17（供应商特定信息）向 iSCSI 客户端提供更多的配置选项。在此配置中，提供了三个附加子选项，用于为 iSCSI 引导客户端分配启动程序 IQN，以及分配另外两个可用于引导的 iSCSI 目标 IQN。

子选项在第 56页上的 表 34 中列出。

表 34：DHCP 选项 17 子选项定义

子选项	*定义*
201	采用标准根路径格式的第一条 iSCSI 目标信息 "iscsi:"[<servername>]":"<protocol>":"<port>":"<LUN>":"<targetname>"
203	iSCSI 启动程序 IQN

注：在表 34 中，方括号 [ ] 是 IPv6 地址所必需的。

选项 17 的内容应为 <2-byte Option Number 201|202|203> <2-byte length> <data>。

配置 DHCP 服务器

配置 DHCP 服务器以支持选项 16 和选项 17。

注：DHCPv6 选项 16 和选项 17 的格式在 RFC 3315 中进行了完整的定义。

VXLAN：配置和使用案例示例

VXLAN 封装允许驻留在一台服务器上的多个 3 层主机发送和接收帧，封装到与安装在同一服务器上的 NIC 卡关联的单个 IP 地址上。

此示例讨论两台 RHEL 服务器之间的基本 VxLan 连通性。每台服务器都启用了一个物理 NIC，外部 IP 地址设置为 1.1.1.4 和 1.1.1.2。

使用多播组 239.0.0.10 创建 VXLAN ID 为 10 的 VXLAN10 接口，并与每台服务器上的物理网络端口 pxp1 关联。

在每台服务器上创建主机的 IP 地址，并将其与 VXLAN 接口关联。VXLAN 接口产生后，系统 1 中存在的主机便可与系统 2 中存在的主机进行通信。VLXAN 格式如表 35 中所示。

表 35：VXLAN 帧格式

MAC 标头	协议为 UDP 的外部 IP 标头	目的地端口为 VXLAN 的 UDP 标头	VXLAN 标头（标志、VNI）	原始 L2 帧	FCS

表 36 提供 VXLAN 命令和配置示例。

表 36：VXLAN 命令和配置示例

系统 1	*系统 2*
PxPy: ifconfig PxPy 1.1.1.4/24	PxPy: ifconfig PxPy 1.1.1.2/24
ip link add vxlan10 type vxlan id 10 group 239.0.0.10 dev PxPy dstport 4789	ip link add vxlan10 type vxlan id 10 group 239.0.0.10 dev PxPy dstport 4789
ip addr add 192.168.1.5/24 broadcast 192.168.1.255 dev vxlan10	ip addr add 192.168.1.10/24 broadcast 192.168.1.255 dev vxlan10
ip link set vxlan10 up	ip link set vxlan10 up
ip –d link show vxlan10	
Ping 192.168.1.10	ifconfig vxlan10 (MTU 1450) (SUSE and RHEL)

注：x 代表系统中物理适配器的 PCIe 总线号。y 代表物理适配器上的端口号。

SR-IOV：配置和使用案例示例

SR-IOV 可以在 10Gb 和 25Gb Broadcom NetExtreme-E NIC 上配置、启用和使用。

Linux 使用案例

1. 启用 NIC 卡中的 SR-IOV：
   a. 可以使用 **HII** 菜单启用 NIC 卡中的 SR-IOV。在系统引导过程中，访问系统 **BIOS -> 设备设置 -> NetXtreme-E NIC ->** 设备级配置。
   b. 设置 SR-IOV 虚拟化模式。
   c. 设置每个物理功能的虚拟功能数。
   d. 设置每个虚拟功能的 MSI-X 矢量数和物理功能 MSI-X 矢量的最大数量。如果虚拟功能耗尽资源，则使用 CCM 平衡每个虚拟机的 MSI-X 矢量数。
2. 在 BIOS 中启用虚拟化：
   a. 在系统引导过程中，进入系统 **BIOS -> 处理器设置 -> 虚拟化技术**，将其设置为**已启用**。
   b. 在系统引导过程中，进入系统 **BIOS -> 集成设备 -> SR-IOV Global**，将其设置为**已启用**。
3. 在启用虚拟化（libvirt 和 Qemu）的情况下安装所需的 Linux 版本。
4. 启用 iommu 核心参数。
   a. 通过编辑 `/etc/default/grub.cfg`，然后运行 `grub2-mkconfig -o /boot/grub2/grub.cfg` 传统模式，启用 IOMMU 核心参数。对于 UEFI 模式，编辑 `/etc/default/grub.cfg`，然后运行 `grub2-mkconfig -o /etc/grub2-efi.cfg`。参阅以下示例：

```
Linuxefi /vmlinuz-3.10.0-229.el7.x86_64 root=/dev/mapper/rhel-root ro rd.lvm.lv=rhel/swap
crashkernel=auto rd.lvm.lv=rhel/root rhgb intel_iommu=on quiet LANG=en_US.UTF.8
```

5. 安装 bnxt_en 驱动程序：
   a. 将 `bnxt_en` 驱动程序复制到操作系统上，然后运行 `make; make install; modprobe bnxt_en`。

注：使用 netxtreme-bnxt_en<version>.tar.gz，在 SRIOV VF 上安装 bnxt_re 和 bnxt_en 以获取 RDMA 功能。

6. 通过核心参数启用虚拟功能：
   a. 安装驱动程序后，lspci 便会显示系统中存在 NetXtreme-E NIC。激活虚拟功能需要提供总线、设备和功能。
   b. 要激活虚拟功能，请输入如下所示的命令：

```
echo X >/sys/bus/pci/device/0000\:Bus\:Dev.Function/sriov_numvfs
```

注：确保 PF 接口已设置。只有设置 PF 后，才能创建 VF。X 是将要导出到操作系统的虚拟功能数量。

典型的示例为：

```
echo 4 > /sys/bus/pci/devices/0000\:04\:00.0/sriov_numvfs
```

7. 检查 PCI-E 虚拟功能：
   a. lspci 命令将显示虚拟功能，并将 BCM57402/BCM57404/BCM57406 的 DID 设置为 16D3、将非 RDMA BCM57412/BCM57414/BCM57416 的该参数设置为 16DC、将启用 RDMA 的 BCM57412/BCM57414/BCM57416 的该参数设置为 16C1。
8. 使用 Virtual Manager 安装虚拟化客户端系统 (VM)。
   关于 Virtual Manager 安装，请参阅 Linux 文档。确保该管理程序的内置驱动程序已经移除。示例为 NIC:d7:73:a7 rtl8139。移除此驱动程序。
9. 为来宾 VM 分配一个虚拟功能。
   a. 将此适配器作为物理 PCI 设备分配给来宾 VM。参阅 Linux 文档，了解有关向 VM 来宾分配虚拟功能的信息。
10. 在 VM 上安装 bnxt_en 驱动程序：
    a. 在来宾 VM 上复制 netxtreme-bnxt_en-<version>.tar.gz 源文件，然后提取 tar.gz 文件。更改每个驱动程序的目录，然后运行 make; make install; modprobe bnxt_en（及 bnxt_re，如启用 RDMA）。通过使用 modinfo 命令检查接口，确保驱动程序正确加载。在加载最新构建模块之前，用户可能需要运行 modprobe -r bnxt_en 卸载现有驱动程序或随附的 bnxt_en 模块。
11. 测试来宾 VM 与外部世界的连通性：
    a. 为适配器分配适当的 IP 地址并测试网络连通性。

Windows 情况下

1. 启用 NIC 卡中的 SR-IOV：
   a. 可以使用 HII 菜单启用 NIC 卡中的 SR-IOV。在系统引导过程中，访问系统 **BIOS -> 设备设置-> NetXtreme-E NIC -> 设备级配置**。
   b. 设置 SR-IOV 虚拟化模式。

c. 设置每个物理功能的虚拟功能数。

d. 设置每个虚拟功能的 MSI-X 矢量数和物理功能 MSI-X 矢量的最大数量。如果虚拟功能耗尽资源，则使用 CCM 平衡每个虚拟机的 MSI-X 矢量数。

2. 在 BIOS 中启用虚拟化：

a. 在系统引导过程中，进入系统 **BIOS -> 处理器设置 -> 虚拟化技术**，将其设置为**已启用**。

b. 在系统引导过程中，进入系统 **BIOS -> 集成设备 -> SR-IOV Global**，将其设置为**已启用**。

3. 为您的 Windows 2012 R2 或 Windows 2016 操作系统安装最新的 KB 更新。

4. 安装合适的虚拟化 (Hyper-V) 选项。了解关于设置 Hyper-V、虚拟交换机和虚拟机的更多详细要求和步骤，请访问 Microsoft.com：
https://technet.microsoft.com/en-us/windows-server-docs/compute/hyper-v/system-requirements-for-hyper-v-on-windows
https://technet.microsoft.com/en-us/windows-server-docs/compute/hyper-v/get-started/install-the-hyper-v-role-on-windows-server

5. 在 Hyper-V 上安装最新的 NetXtreme-E 驱动程序。

6. 在 NDIS 微型端口高级属性中启用 SR-IOV。

7. 在 Hyper-V 管理器中，使用选定的 NetXtreme-E 接口创建自己的虚拟交换机。

8. 在创建 Hyper-V 虚拟适配器时，勾选**启用单域根 I/O 虚拟化 (SR-IOV)** 框。

9. 创建虚拟机 (VM) 然后添加所需数量的虚拟适配器。

10. 在“虚拟机的网络适配器”设置中，在**硬件加速**部分为每个虚拟适配器勾选**启用 SR-IOV**。

11. 启动您的 VM，然后安装所需的客户机操作系统。

12. 为每个客户机操作系统安装相应的 NetXtreme-E 驱动程序。

注：NetXtreme-E 的虚拟功能 (VF) 驱动程序与基本驱动程序相同。例如，如果客户机操作系统是 Windows 2012 R2，则客户需要在 VM 中安装 Bnxtnd64.sys。用户可以通过运行 NetXtreme-E 驱动程序安装可执行程序达到此目的。在客户机操作系统中安装驱动程序后，用户可以在 VM 客户机操作系统的“设备管理器”中看到 VF 驱动程序接口。

VMWare SRIOV 情况下

1. 启用 NIC 卡中的 SR-IOV：

a. 可以使用 HII 菜单启用 NIC 卡中的 SR-IOV。在系统引导过程中，访问系统 **BIOS -> 设备设置-> NetXtreme-E NIC -> 设备级配置**。

b. 设置 SR-IOV 虚拟化模式。

c. 设置每个物理功能的虚拟功能数。

d. 设置每个虚拟功能的 MSI-X 矢量数和物理功能 MSI-X 矢量的最大数量。如果虚拟功能耗尽资源，则使用 CCM 平衡每个虚拟机的 MSI-X 矢量数。

2. 在 BIOS 中启用虚拟化：

a. 在系统引导过程中，进入系统 **BIOS -> 处理器设置 -> 虚拟化技术**，将其设置为**已启用**。

b. 在系统引导过程中，进入系统 **BIOS -> 集成设备 -> SR-IOV Global**，将其设置为**已启用**。

3. 在 ESXi 上按以下步骤安装 Bnxtnet 驱动程序：

a. 在 /var/log/vmware 中复制 <bnxtnet>-<driver version>.vib 文件。

```
$ cd /var/log/vmware。
$ esxcli software vib install --no-sig-check -v <bnxtnet>-<driver version>.vib。
```

b. 重启计算机。

c. 验证驱动程序是否正确安装：

```
$ esxcli software vib list | grep bnxtnet
```

4. 安装 Broadcom 提供的 BNXTNETCLI (esxcli bnxtnet) 实用程序，以设置/查看 esxcli 中原本不支持的杂项驱动程序参数，例如将链路速度设置为 25G，显示驱动程序/固件芯片信息，显示 NIC 配置（NPAR、SRIOV）。有关详细信息，请参阅 bnxtnet 驱动程序的 README.txt 文件。

要安装此实用程序：

a. 在 /var/log/vmwareCopy 中复制 BCM-ESX-bnxtnetcli-<version>.vib 文件。

```
$ cd /var/log/vmware
$ esxcli software vib install --no-sig-check -v /BCM-ESX-bnxtnetcli-<version>.vib
```

b. 重启系统。

c. 验证 vib 是否正确安装：

```
$ esxcli software vib list | grep bcm-esx-bnxtnetcli
```

d. 将速度设置为 10/20/25

```
$ esxcli bnxtnet link set -S <speed> -D <full> -n <iface>
```

如果速度设置正确，将返回“确定”消息。

示例：

```
$ esxcli bnxtnet link set -S 25000 -D full -n vmnic5
```

e. 显示链路统计数据

```
$ esxcli bnxtnet link get -n vmnic6
```

f. 显示驱动程序/固件/芯片信息

```
$ esxcli bnxtnet drvinfo get -n vmnic4
```

g. 显示 NIC 信息（例如 BDF，NPAR、SRIOV 配置）

```
$ esxcli bnxtnet nic get -n vmnic4
```

5. 启用 SRIOV VF：

只有 PF 能自动启用。如果 PF 支持 SR-IOV，则 PF (vmknicX) 是以下命令的输出结果的一部分。

```
esxcli network sriovnic list
```

要启用一个或多个 VF，驱动程序使用模块参数“max_vfs”为 PF 启用所需数量的 VF。例如，要在 PF1 上启用 4 个 VF：

```
esxcfg-module -s 'max_vfs=4' bnxtnet（需要重启）
```

要在一组 PF 上启用 VF，请使用以下所示格式的命令。例如，要在 PF 0 上启用 4 个 VF，在 PF 2 上启用 2 个 VF：

```
esxcfg-module -s 'max_vfs=4,2' bnxtnet（需要重启）
```

在 PF 启用过程中，每个受支持的 PF 上所需的 VF 将按顺序启用。请参阅 VMware 文档，了解如何将 VF 映射到 VM 的相关信息：

https://pubs.vmware.com/vsphere-60/index.jsp#com.vmware.vsphere.networking.doc/GUID-EE03DC6F-32CA-42EF-98FC-12FDE06C0BE0.html

https://pubs.vmware.com/vsphere-60/index.jsp#com.vmware.vsphere.networking.doc/GUID-CC021803-30EA-444D-BCBE-618E0D836B9F.html

NPAR – 配置和使用案例示例

功能和要求

- 与 OS/BIOS 无关 - 分区在操作系统中如同“真实的”网络接口，所以无须特殊的 BIOS 或 OS 支持，如 SR-IOV。
- 额外的 NIC，无须其他交换机端口、布线、PCIe 扩展槽。
- 通信量调整功能 - 可以控制每个分区的带宽分配，以便根据需要进行限制或保留。
- 可用于交换机不相关方式 - 交换机不需要任何特殊配置，也不需要知道 NPAR 的启用状态。
- 可与 RoCE 和 SR-IOV 配合使用。
- 支持无状态卸载，如 LSO、TPA、RSS/TSS 和 RoCE（每端口仅 2 个 PF）。
- 替代路由 ID 支持每个物理设备超过 8 个功能。

注：在 **Dell UEFI HII 菜单 -> 主配置 -> 设备级配置**页面中，用户可以启用 **NParEP**，以允许 NXE 适配器在具有 ARI 功能的系统中支持每设备最多 16 个 PF。对于 2 端口设备，表示每个端口最多 8 个 PF。

限制

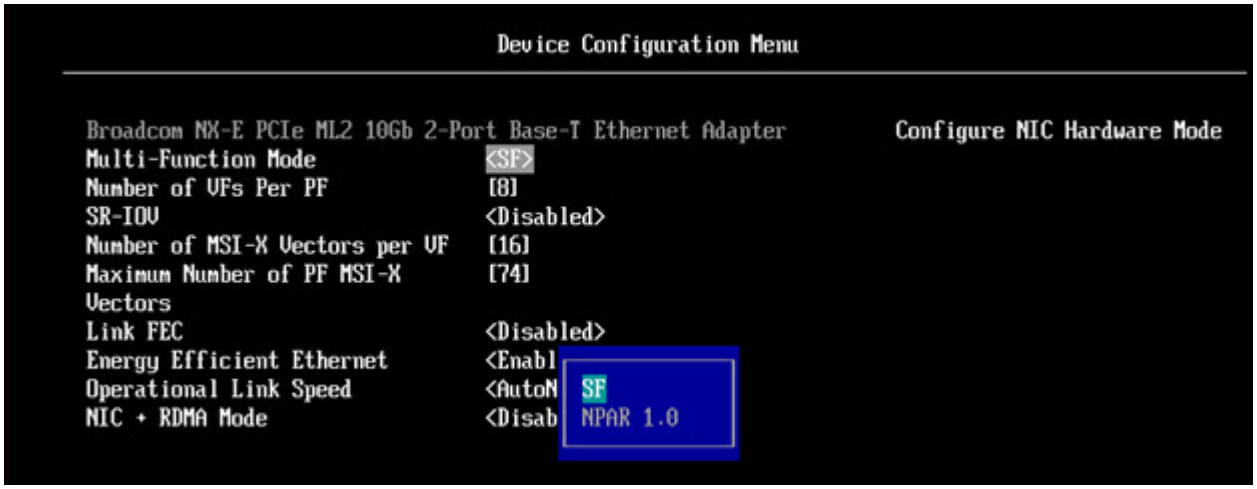
- 必须中止共享设置以避免争用。例如：速度、双工、流控制和类似物理设置都由设备驱动程序隐藏，以避免争用。
- 在不支持 ARI 的系统中，每个物理设备仅启用 8 个分区。
- 只有每个物理端口的前两个分区或每个物理设备的总计 4 个分区中支持 RoCE。在 NPAR + SRIOV 模式下，每个父级物理端口只有 2 个 VF 可以启用 RDMA 支持，或每个物理设备只有总计 4 个 VF + RDMA。

配置

可以使用 BIOS 配置 HII 菜单或在传统引导系统中使用 Broadcom CCM 实用程序对 NPAR 进行配置。一些供应商还通过其他专有接口公开配置。

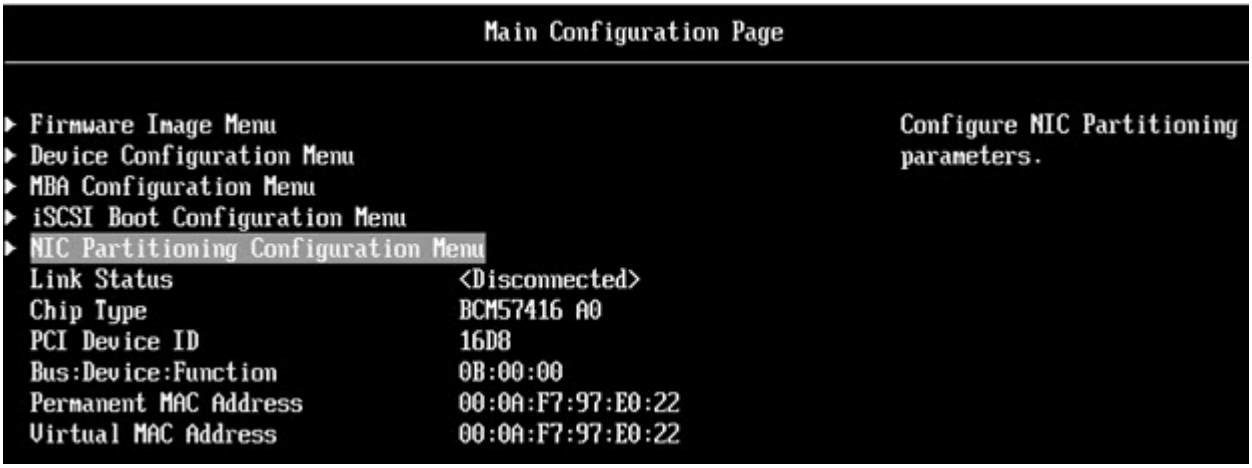
启用 NPAR：

1. 从 BIOS HII 菜单或 CCM 接口选择目标 NIC，然后设置“多功能模式”或“虚拟化模式”选项。

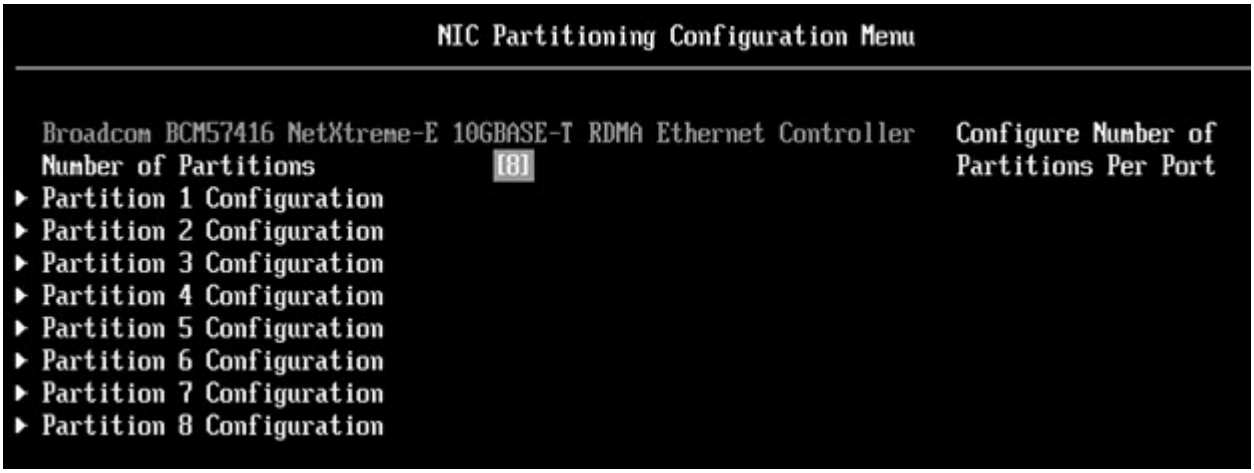


NPAR 已启用，与 SR-IOV 组合使用。对于某些具有 ARI 功能的 OEM 系统，**NParEP** 按钮可用于明确允许 BCM5741X 支持最多 16 个分区。从“单功能”模式切换到“多功能”模式需要重新枚举设备，所以在系统重启之前，修改不会生效。

2. 启用 NPAR 后，在每个物理端口相关的主 NIC 配置菜单中，“NIC 分区主配置菜单”选项可用。



3. “NIC 分区配置菜单”（如下所示）允许用户选择应从选定物理端口进行分配的分区数量。在具有 ARI 功能的服务器上，每个 BCM5741X NIC 支持最多 16 个分区。默认情况下，为每个物理端口的 8 个分区配置双端口适配器。还可通过此菜单访问每个分区的配置选项。对于某些 OEM 系统，HII 菜单还包括“全球带宽分配”页面，可同时为所有分区配置最小（保留）和最大（限制）TX 带宽。



4. 设置 NIC 分区配置参数（请参阅 第 64页上的 表 37）。

表 37：NPAR 参数

参数	*描述*	*有效选项*
BW 预留	应为此分区保留的总可用带宽百分比。0 表示所有分区的带宽相等。	值 0-100
BW 限制	此分区允许的最大可用带宽百分比。	值 0-100
BW 保留有效	用作 BW 保留设置的 on/off 开关功能。	True/False
BW 限制有效	用作 BW 限制设置的 on/off 开关功能。	True/False
支持 RDMA	用作此分区 RDMA 支持的 on/off 开关功能。请注意：只有每物理端口两个分区可以支持 RDMA。对于双端口设备，最多 4 个 NPAR 分区可以支持 RDMA。	已启用/已禁用
MAC 地址	此分区的 MAC 地址。	–

降低 NIC 内存消耗注意事项

由于此 NIC 支持较快的链路速度，因此接收缓冲区的默认数量更大。当链路速度较快时，在给定的时间间隔内到达的数据包更多，如果主机系统在处理接收中断情况时有延迟，那么只要所有可用接收缓冲区都在使用，NIC 将必须丢失数据包。

选择接收缓冲区默认值时要使其在典型配置下工作良好。但是，如果系统中有多个 NIC、多个 NIC 上已启用 NPAR 或只有少量 RAM，则您可参阅设备管理器代码 12 的“黄色叹号”了解 NIC 相关信息。代码 12 表示：由于没有足够资源，驱动程序加载失败。在此情况下，资源是特定类型的内核内存，被称为非分页内存池 (NPP) 内存。

如果您得到代码 12 或因为其他原因想降低 NIC 消耗的 NPP 内存量：

- 减少 RSS 队列的数量，从默认的 8 降至 4 或 2。每个 RSS 队列都被分配一组自己的接收缓冲区，所以减少 RSS 队列的数量将减少所分配的 NPP 内存。减少 RSS 队列的数量可能会影响性能，因为参与处理该 NIC 的接收数据包的内核数量变少。应该监控每个处理器的 CPU 利用率，以确保在此更改后没有变“热”的处理器。
- 通过减少所分配接收缓冲区的数量降低分配的内存。默认值 0 表示驱动程序应自动确定接收缓冲区数量。对于典型配置，设置为 0（自动）将为每个队列映射到 XXXX 接收缓冲区。您可以选择一个更小的值，如 1500、1000 或 500。（该值必须在 500 到 15000 之间，且为 500 的倍数。）如上所述，接收缓冲区数量变少将提高丢包的风险，因而影响数据包发送并提高吞吐量。

参数“RSS 队列的最大数量”和“接收缓冲区（0 = 自动）”，可通过**设备管理器**中每个 NIC 的**高级**属性选项卡进行修改。如果您想同时修改多个 NIC，使用 *Set-NetAdapterAdvancedProperty PowerShell cmdlet* 操作更快。例如，要为系统中所有名称以“SI”开头的 NIC 分配 2 个 RSS 队列，请运行以下命令：

```
Set-NetAdapterAdvancedProperty Sl* -RegistryKeyword *NumRSSQueues -RegistryValue 2
```

同样，要将接收缓冲区的数量设置为 1500，请运行以下命令：

```
Set-NetAdapterAdvancedProperty Sl* -RegistryKeyword *ReceiveBuffers -RegistryValue 1500
```

请参阅 https://blogs.technet.microsoft.com/wincat/2012/08/27/using-powershell-for-nic-configuration-tasks/ 以了解使用 PowerShell 修改 NIC 属性的概述。

RoCE - 配置和使用案例示例

本节将提供 RoCE 的配置和使用案例示例。

要为 PF 或 VF 启用 RoCE，用户必须在 BIOS 的 HII 菜单中启用 RDMA 选择，然后 RDMA 选项才可在主机或客户机操作系统中生效。

要在单功能模式下启用 RDMA（如果**虚拟模式**为**无**或 **SR-IOV**）：

1. 在系统引导期间，访问**系统设置 -> 设备设置 -> NetXtreme-E NIC -> 主配置页面**，然后将 **NIC+ RMDA 模式**设置为**已启用**。

在虚拟模式为 NPAR 或 NPAR+SR-IOV 的情况下启用 RDMA：

1. 在系统引导期间，访问**系统设置 -> 设备设置 -> NetXtreme-E NIC -> NIC 分区配置 -> 分区 1（或 2）配置**，然后将 **NIC+ RMDA 模式**设置为**已启用**。

注： 如果使用 NPAR+SRIOV 模式，则每个父级物理端口只有两个 VF 可以启用 RDMA 支持，或者每台物理设备总共有四个 VF+RDMA。

Linux 配置

要求

要在 Linux 中配置 RoCE，需要以下项目：

- Bnxt_en-roce（支持 RoCE 的 bnxt_en 驱动程序，是已推出的 gzip 压缩 tar 存档的一部分）
- bnxt_re（RoCE 驱动程序）
- libbnxtre（用户模式 RoCE 库模块）

BNXT_RE 驱动程序依存关系

Bnxt_re 驱动程序需要已启用 RoCE 的特殊版本的 bnxt_en，它包含在 netxtreme-bnxt_en-1.7.9.tar.gz（或更新的）程序包中。Bnxt_re 驱动程序编译取决于是否需要 IB 堆栈与操作系统分布或外部 OFED 同时可用。

注：有必要加载正确的 bnxt_en 版本，它包含在相同的 netxtreme-bnxt_en-1.7.x.tar.gz 程序包中。Bnxt_re 和 Bnxt_en 作为一对，用于启用 RoCE 通信。使用这两种驱动程序的不匹配版本会产生不可信或不可预测的结果。

- 拥有操作系统分布与 IB 堆栈同时可用的发行版有：

  RH7.1/7.2/7.3/6.7/6.8、SLES12SP2 和 Ubuntu 16.04

  如果尚未安装，可以通过以下命令在 Redhat 中安装 IB 堆栈和有用的工具包，然后再编译 bnxt_re：

  ```
  yum -y install libibverbs* inifiniband-diag perftest qperf librdmacm utils
  ```

要编译 bnxt_re：

```
$make
```

- 需要安装外部 OFED 的发行版有：
  SLES11SP4
  请访问以下链接，参阅 OFED 发行说明并安装 OFED，然后再编译 bnxt_re 驱动程序。
  http://downloads.openfabrics.org/downloads/OFED/release_notes/OFED_3.18-2_release_notes
  要编译 bnxt_re：

```
$export OFED_VERSION=OFED-3.18-2
$make
```

安装

要在 Linux 中安装 RoCE：

1. 使用支持 RoCE 的固件程序包，将 NIC NVRAM 从软件发行版 20.06.04.01 或更新版本进行升级。
2. 在操作系统中，解压缩、构建并安装 BCM5741X Linux L2 和 RoCE 驱动程序。
   a. # tar -xzf netxtreme-bnxt_en-1.7.9.tar.gz
   b. # cd netxtreme-bnxt_en-bnxt_re
   c. # make build && make install
3. 解压缩、构建和安装 NetXtreme-E Linux RoCE 用户库。
   a. # tar xzf libbnxtre-0.0.18.tar.gz
   b. #cd libbnxtre-0.0.18
   c. # configure && make && make install。
   d. # cp bnxtre.driver /etc/libibverbs.d/
   e. # echo "/usr/local/lib" >> /etc/ld.so.conf
   f. # ldconfig -v

请参阅 bnxt_re README.txt 了解有关可配置选项和建议的更多详细信息。

限制

在双端口 NIC 上，如果两个端口位于同一子网上，则 rdma perftest 命令可能会失败。可能的原因是 Linux 操作系统中出现 arp 通量问题。要避开此限制，请使用多个子网进行测试或将第二个端口/接口降低。

已知问题

Bnxt_en 和 Bnxt_re 设计为成对发挥作用。版本 1.7.x 之前的较旧 Bnxt_en 驱动程序不支持 RDMA 且无法与 Bnxt_re (RDMA) 驱动程序同时加载。如果 Bnxt_re 与较旧 Bnxt_en 驱动程序一同加载，则用户可能会经历系统崩溃及重启。建议用户从同一 netxtreme-bnxt_en-<1.7.x>.tar.gz 包中加载 Bnxt_en 和 Bnxt_re 模块。

为防止加载不匹配的 bnxt_en 和 bnxt_re 组合，需要如下操作：

- 如果已使用带 bnxt_en DUD 或内核模块 RPM 的 PXEboot 将 RedHat/CentOS 7.2 操作系统安装到目标系统中，请删除 /lib/modules/$(uname -r)/extra/bnxt_en/bnxt_en.ko 中的 bnxt_en.ko 文件或编辑 /etc/depmod.d/.
- 用 Bnxt_en.conf 进行覆写，以使用更新的版本。用户还可以使用 rpm -e kmod-bnxt_en 命令清除当前的 BCM5741X Linux 内核驱动程序。RHEL 7.3/SLES 12 Sp2 内置 bnxt_en 驱动程序（v1.7.x 之前的版本）。必须移除此驱动程序并添加最新的 bnxt_en，才可应用 bnxt_re（RoCE 驱动程序）。

Windows

内核模式

如果为 RDMA 启用了两端，则 Windows Server 2012 及更新版本将为 SMB 文件通信量调用 NIC 中的 RDMA 功能。Broadcom NDIS 微型端口 bnxtnd.sys v20.6.2 和更新版本通过 NDKPI 接口支持 RoCEv1 和 RoCEv2。默认设置为 RoCEv1。

要启用 RDMA：

1. 使用适当的板程序包升级 NIC NVRAM。在 CCM 或 UEFI HII 中，启用对 RDMA 的支持。
2. 转至适配器**高级属性**页面，然后将每个 BCM5741X 微型端口的 **NetworkDirect 功能**设置为**已启用**，或者使用 PowerShell 窗口，运行以下命令：

   ```
   Set-NetAdapterAdvancedProperty -RegistryKeyword *NetworkDirect -RegistryValue 1
   ```

3. 如果 **NetworkDriect** 已启用，则以下 PowerShell 命令将返回 true。
   a. Get-NetOffLoadGlobalSetting
   b. Get-NetAdapterRDMA

验证 RDMA

要验证 RDMA：

1. 在远程系统上创建文件共享，然后使用 Windows 资源管理器或“net use ….”打开该共享。为避免硬盘读取/写入速度出现瓶颈，建议将 RAM 盘作为正在进行测试的网络共享。
2. 在 PowerShell 中，运行以下命令：

   ```
   Get-SmbMultichannelConnection | fl *RDMA*
   ClientRdmaCapable : True
   ServerRdmaCapable : True
   ```

   如果客户机和服务器都显示 True，则任何文件都通过使用 SMB 的此 SMB 连接进行传输。

3. 可以使用以下命令启用/禁用 SMB Multichannel：

   服务器端：

   - 启用：Set-SmbServerConfiguration -EnableMultiChannel $true
   - 禁用：Set-SmbServerConfiguration -EnableMultiChannel $false

客户机端：

- 启用：Set-SmbClientConfiguration -EnableMultiChannel $true
- 禁用：Set-SmbClientConfiguration -EnableMultiChannel $false

注：默认情况下，该驱动程序为每个 IP 地址的每个网络共享最多设置两个 RDMA 连接（在唯一子网上）。用户可以通过为正在进行测试的同一物理端口添加多个 IP 地址（每个都有不同的子网），以此增加 RDMA 连接的数量。可以使用创建的唯一 IP 地址创建多个网络共享并将其映射到每个链路伙伴。

例如：

```
在服务器 1 上，为 网络端口 1 创建以下 IP 地址。
172.1.10.1
172.2.10.2
172.3.10.3

在同一服务器 1 上，创建 3 个共享。
Share1
Share2
Share3

在网络链路伙伴上，
连接到 \\172.1.10.1\share1
连接到 \\172.2.10.2\share2
连接到 \\172.3.10.3\share3
等等
```

用户模式

可以运行写入到 NDSPI 的用户模式应用程序之前，复制并安装 `bxndspi.dll` 用户模式驱动程序。要复制和安装用户模式驱动程序：

1. 复制 `bxndspi.dll` 到 `C:\Windows\System32.`
2. 通过运行以下命令安装驱动程序：

```
rundll32.exe .\bxndspi.dll,Config install|more
```

VMware ESX

限制

支持 RoCE 的当前版本驱动程序需要 ESXi-6.5.0 GA 版本 4564106 或更高版本。

BNXT RoCE 驱动程序要求

必须将 BNXTNET L2 驱动程序与 `disable_roce=0` 模块参数一同安装，然后才能安装驱动程序。

要设置模块参数，请运行以下命令：

```
esxcfg-module -s "disable_roce=0" bnxtnet
```

请使用 ESX6.5 L2 驱动程序版本 20.6.9.0（支持 RoCE 的 L2 驱动器）或更高版本。

安装

要安装 RoCE 驱动程序：

1. 使用以下命令在 `/var/log/vmware` 中复制 `<bnxtroce>-<driver version>.vib` 文件：

   ```
   $ cd /var/log/vmware
   $ esxcli software vib install --no-sig-check -v <bnxtroce>-<driver version>.vib
   ```

2. 重启计算机。
3. 使用以下命令验证驱动程序是否正确安装：

   ```
   esxcli software vib list | grep bnxtroce
   ```

4. 要为 RoCE 通信量禁用 ECN（默认启用），请为 bnxtroce 使用“tos_ecn=0”模块参数。

配置 Paravirtualized RDMA 网络适配器

请参阅下方 vmware 连接，了解有关设置和使用 Paravirtualized RDMA (PVRDMA) 网络适配器的更多信息。

https://pubs.vmware.com/vsphere-65/index.jsp#com.vmware.vsphere.networking.doc/GUID-4A5EBD44-FB1E-4A83-BB47-BBC65181E1C2.html

为 PVRDMA 配置虚拟中心

要为 PVRDMA 配置虚拟中心：

1. 创建 DVS（PVRDMA 需要分布式虚拟交换机）
2. 将主机添加到 DVS。

为 ESX 主机上的 PVRDMA 标记 vmknic

要为 PVRDMA 标记 vmknic 以在 ESX 主机上使用：

1. 选择主机，然后右键单击**设置**切换到**管理**选项卡的设置页面。
2. 在**设置**页面中，展开**系统**，然后单击**高级系统设置**，以显示“高级系统设置”密钥对及其摘要。
3. 单击**编辑**出现**编辑高级系统设置**。
   过滤 **PVRDMA** 以将所有设置的范围缩小至只有 `Net.PVRDMAVmknic`。
4. 将 `Net.PVRDMAVmknic` 值设置为 vmknic，如示例中的 vmk0

为 PVRDMA 设置防火墙规则

要为 PVRDMA 设置防火墙规则：

1. 选择主机，然后右键单击**设置**切换到**管理**选项卡的设置页面。
2. 在**设置**页面中，展开**系统**，然后单击**安全配置文件**以显示防火墙摘要。
3. 单击**编辑**出现**编辑安全配置文件**。
4. 向下滚动找到 pvrdma，然后勾选相应的框设置防火墙。

将 PVRDMA 设备添加到 VM

要将 PVRDMA 设备添加到 VM：

1. 选择 VM，然后右键单击**编辑设置**。
2. 添加新的网络适配器。
3. 选择网络作为**分布式虚拟交换机**和**端口组**。
4. **适配器类型**选择 **PVRDMA**，然后单击**确定**。

在 Linux 客户机操作系统上配置 VM

注：用户必须安装适当的开发工具（包括 git），然后才能继续执行下方配置步骤。

1. 使用以下命令下载 PVRDMA 驱动程序和库：

   ```
   git clone git://git.openfabrics.org/~aditr/pvrdma_driver.git
   git clone git://git.openfabrics.org/~aditr/libpvrdma.git
   ```

2. 编译并安装 PVRDMA 客户机驱动程序和库。
3. 要安装该驱动程序，请在驱动程序目录中执行 `make && sudo insmod pvrdma.ko`。
   必须先加载已配对的 vmxnet3 驱动程序，然后再加载该驱动程序。

已安装的 RDMA 内核模块可能与 PVRDMA 驱动程序不兼容。如果不兼容，请移除当前安装，然后重启。然后按安装说明进行操作。请阅读驱动程序目录中的 README 了解有关不同 RDMA 堆栈的更多信息。

4. 要安装库，请在库目录中执行 `./autogen.sh && ./configure --sysconfdir=/etc && make && sudo make install`。

注：库的安装路径需要位于共享的库缓存中。按库目录中 INSTALL 文件中的说明操作。

注：可能需要将防火墙设置修改为允许 RDMA 通信量。请确保进行了适当的防火墙设置

5. 将 `/usr/lib` 添加到 `/etc/ld.so.conf` 文件中，然后通过运行 `ldconfig` 重新加载 `ldconf`
6. 使用 `modprobe rdma_ucm` 加载 ib 模块。
7. 使用 `insmod pvrdma.ko` 加载 PVRDMA 内核模块。
8. 为 PVRDMA 接口分配 IP 地址。
9. 通过运行 `ibv_devinfo -v` 命令验证是否已创建 IB 设备。

DCBX - 数据中心桥接

Broadcom NetXtreme-E 控制器支持 IEEE802.1Qaz DCBX 及更旧的 CEE DCBX 规格。通过将本地配置的设置与链路对等端交换获取 DCB 配置。由于链路的两端可能配置不同，因此 DCBX 使用“愿意”的概念指示链路的哪一端准备好接受另一端的参数。这在 DCBX 协议的 ETS 配置和 PFC TLV 中使用一位指示，这一位不用于 ETS 建议和应用程序优先级 TLV。默认情况下，NetXtreme-E NIC 处于“愿意”模式，而链路伙伴网络交换机处于“不愿意”模式。这将确保交换机上的相同 DCBX 设置延伸到整个网络。

用户可以将 NetXtreme-E NIC 手动设置为“不愿意”模式，并从主机端执行各种 PFC、严格优先级、ETS、APP 配置。请参阅驱动程序的 readme.txt 了解可用配置的更多详细信息。此文档将提供示例，使您了解如何在具有 Windows PowerShell 的 Windows 中进行此类设置。在单独的白皮书中详细描述了关于 DCBX、QoS 和相关案例的更多信息，其内容超过本用户手册的范围。

UEFI HII 菜单中以下设置需要启用 DCBX 支持：

系统设置 -> 设备设置 -> NetXtreme-E NIC -> 设备级配置

QoS 配置文件 – 默认 QoS 列队配置文件

服务器质量 (QoS) 资源配置需要支持各种 PFC 和 ETS 要求，需要为带宽分配进行更精细的调节。NetXtreme-E 允许管理员选择将 NIC 硬件资源用于支持 Jumbo 帧和/或有损和无损服务等级队列（CoS 队列）组合。有多种可能的配置组合，所以计算很复杂。此选项允许用户从预先计算的 QoS 队列配置文件列表中选择。这些预先计算的配置文件在典型客户部署中用于优化对 PFC 和 ETS 要求的支持。

以下是每个 QoS 配置文件的摘要描述。

表 38：QoS 配置文件

配置文件编号	*Jumbo 帧支持*	*有损 CoS 队列/端口数量*	*无损 CoS 队列/端口数量*	*支持 2 端口 SKU*
配置文件 #1	是	0	1（支持 PFC）	是 (25 Gbps)
配置文件 #2	是	4	2（支持 PFC）	否
配置文件 #3	否。 (MTU <= 2 KB)	6	2（支持 PFC）	是 (25 Gbps)
配置文件 #4	是	1	2（支持 PFC）	是 (25 Gbps)
配置文件 #5	是	1	0（不支持 PFC）	是 (25 Gbps)
配置文件 #6	是	8	0（不支持 PFC）	是 (25 Gbps)
配置文件 #7	此配置的数据包缓冲区最多分配到两个无损 CoS 应用队列在权衡灵活性的同时最大限度地提高 RoCE 性能。			
	是	0	2	是 (25 Gbps)
默认	是	与配置文件 #4 相同		是

DCBX 模式 = 启用（仅 IEEE）

此选项允许用户启用/禁用指定规格的 DCBX。IEEE 只表示已选中 IEEE802.1Qaz DCBX。

Windows 驱动程序设置：

在 UEFI HII 菜单中启用指定选项以设置固件级别设置后，在 Windows 驱动程序高级属性中执行以下选择。

打开 **Windows 设备管理器 -> Broadcom NetXtreme E 系列适配器 -> 高级属性 ->“高级”选项卡**

服务质量 = 已启用

优先级和 VLAN = 优先级和 VLAN 已启用

VLAN = <ID>

设置所需 VLAN ID

要在 Windows PowerShell 中验证 DCB 相关命令，请安装相应的 DCB Windows 功能。

1. 在**任务栏**中，右键单击 Windows PowerShell 图标，然后单击**以管理员身份运行**。Windows PowerShell 以提升模式打开。
2. 在 Windows PowerShell 控制台输入：

```
Install-WindowsFeature "data-center-bridging"
```

DCBX Willing 位

DCBX willing 位在 DCB 规范中指定。如果设备中的 Willing 位为 true，则设备愿意接受通过 DCBX 连接的远程设备的配置。如果设备中的 Willing 位为 false，则设备拒绝任何远程设备的配置，仅强制使用本地配置。

使用以下命令将 willing 位设置为“True”或“False”。1 为启用，0 为禁用。

例如 `set-netQoSdcbxSetting -Willing 1`

使用以下命令创建通信量类别。

```
C:\> New-NetQosTrafficClass -name "SMB class" -priority 4 -bandwidthPercentage 30 -Algorithm ETS
```

注：默认情况下，所有 802.1p 的值映射到默认通信量类别，该类别拥有物理链路 100% 的带宽。如上所示的命令将创建新的通信量类别，任何标记有 8 个 IEEE 802.1p 值 4 的数据包映射到该类别，传输选择算法 (TSA) 为 ETS 且占用 30% 的带宽。

最多可创建 7 个新的通信量类别。除默认通信量类别外，系统中最多有 8 个通信量类别。

使用以下命令显示创建的通信量类别：

```
C:\> Get-NetQoSTrafficClass
名称            算法   带宽（百分比）    优先级
---------------------------
[默认]          ETS                 70         0-3，5-7
SMB 类别         ETS               30             4
```

使用以下命令修改通信量类别：

```
PS C:\> Set-NetQoSTrafficClass -Name "SMB class" -BandwidthPercentage 40
PS C:\> get-NetQosTrafficClass
名称 算法 带宽（百分比） 优先级
-----------------------------------------------------------------
[默认] ETS            60 0-3，5-7
SMB 类别 ETS          40       4
```

使用以下命令移除通信量类别：

```
PS C:\> Remove-NetQosTrafficClass -Name "SMB class"
PS C:\> Get-NetQosTrafficClass
名称 算法 带宽（百分比） 优先级
-----------------------------------------------------------------
[默认] ETS          100              0-7
```

使用以下命令创建通信量类别（严格优先级）：

```
C:\> New-NetQosTrafficClass -name "SMB class" -priority 4 -bandwidthPercentage 30-Algorithm
Strict
```

启用 PFC：

```
PS C:\> Enable-NetQosFlowControl -priority 4
PS C:\> Get-NetQosFlowControl -priority 4
启用优先级
-----------------------------------------------------------------
4 True

PS C:\> Get-NetQosFlowControl
```

禁用 PFC：

```
PS C:\> disable-NetQosflowControl -priority 4
PS C:\> get-NetQosFlowControl -priority 4
启用优先级
-----------------------------------------------------------------
4 False
```

使用以下命令创建 QoS 策略：

```
PS C:\> New-NetQosPolicy -Name "SMB policy" -SMB -PriorityValue8021Action 4

名称：SMB 策略
所有者：组策略（计算机）
网络配置文件：全部
优先级：127
```

注：上述命令为 SMB 创建新策略。-SMB 是随附的过滤器，与 TCP 端口 445 匹配（为 SMB 保留）。如果数据包被发送到 TCP 端口 445，将被操作系统标记为 802.1p 值 4，然后才能传递到网络微型端口驱动程序。

除 -SMB 外的其他默认过滤器包括：-iSCSI（匹配 TCP 端口 3260）、-NFS （匹配 TCP 端口 2049）、-LiveMigration（匹配 TCP 端口 6600）、-FCOE（匹配 EtherType 0x8906）和 -NetworkDirect。

NetworkDirect 是我们在网络适配器的任何 RDMA 实施的顶层创建的抽象层。-NetworkDirect 必须后跟 Network Direct 端口。

除默认过滤器外，用户可以按应用程序的可执行文件名称对通信进行分类（如方第一个示例），或者按 IP 地址、端口或协议分类。

使用以下命令根据源/目的地址创建 QoS 策略：

```
PS C:\> New-NetQosPolicy "Network Management" -IPDstPrefixMatchCondition 10.240.1.0/24 -
IPProtocolMatchCondition both -NetworkProfile all -PriorityValue8021Action 7
名称：网络管理
所有者：组策略（计算机）
网络配置文件：全部
优先级：127
IP 协议：两者
目的 IP 地址前缀：10.240.1.0/24
优先级值：7
```

使用以下命令显示 QoS 策略：

```
PS C:\> Get-NetQosPolicy
名称：网络管理
所有者：(382ACFAD-1E73-46BD-A0A-6-4EE0E587B95)
网络配置文件：全部
优先级：127
IP 协议：两者
目的 IP 地址前缀：10.240.1.0/24
优先级值：7
名称：SMB 策略
所有者：(382AFAD-1E73-46BD-A0A-6-4EE0E587B95)
网络配置文件：全部
优先级：127
模板：SMB
优先级值：4
```

使用以下命令修改 QoS 策略：

```
PS C:\> Set-NetqosPolicy -Name "Network Management" -IPSrcPrefixMatchCondition 10.235.2.0/24 -
IPProtocolMatchCondition both -PriorityValue802.1Action 7
PS C:\> Get-NetQosPolicy -name "network management"
名称：网络管理
所有者：{382ACFD-1E73-46BD-A0A0-4EE0E587B95}
网络配置文件：全部
优先级：127
IP 协议：两者
源 IP 地址前缀：10.235.2.0/24
目的 IP 地址前缀：10.240.1.0/24
优先级值：7
```

使用以下命令移除 QoS 策略：

```
PS C:\> Remove-NetQosPolicy -Name "Network Management"
```

常见问题

- 是否支持 25G 速度下的 AutoNeg？

  是。请参阅 第 40 页上的“自动协商配置” 了解更多详细信息。

- 如何将 SFP28 电缆连接到 QSFP 端口？

  有从 QSFP 到 4 个 SFP28 端口的分支电缆。

- 什么是兼容端口速度？

  对于 BCM57404AXXXX/BCM57414 双端口设备，每个端口的端口速度必须与另一端口的端口速度兼容。10 Gbps 和 25 Gbps 链路速度不兼容。如果一个端口设置为 10 Gbps，则另一个端口不能设置为 25 Gbps。如果用户尝试设置不兼容的端口速度，则第二个要启用的端口将无法链接。请参阅 第 40 页上的“自动协商配置” 了解更多详细信息。

- 我能对 25 Gbps 端口上的 PXE 连接使用 10 Gbps 吗？

  目前仅支持 25 Gbps PXE 速度。不建议在 25 Gbps 适配器上使用 10 Gbps PXE 连接。这是由于现有的 10 Gbps 交换机不支持自动协商，而且会因为不兼容的端口链路速度设置而引起复杂问题。