□↑从英语翻译成中文(简体) - www.onlinedoctranslator.com



NVIDIA ConnectX-7卡(适用于 OCP Spec 3.0)用户手册

目录

1	关于本手册	5
	1.1 订购部件编号	5
	1.2 目标读者	6
	1.3 技术支持	6
	1.4 相关文档	6
	1.5 文档约定	7
2	简介	8
	2.1 产品概述	8
	2.2 系统要求	9
	2.3 包装内容	9
	2.4 特点和优点	10
3	支持的接口	15
	3.1 ConnectX-7 布局和接口信息	15
	3.2 接口详细说明	17
	3.2.1 ConnectX-7 芯片	17
	3.2.1.1 NVIDIA Multi-HostTM 支持	17
	3.2.2 PCI Express 接口	18
	3.2.3 网络接口	19
	3.2.4 网络端口 LED 规格	20
	3.2.5 FRU EEPROM	21
	3.2.6 SMBus 接口	21
	3.2.7 电压调节器	21
4	硬件安装	22
	4.1 安全警告	22
	4.2 安装步骤概述	22
	4.2.1 系统要求	23

	4.2.1.1	硬件要求	23
	4.2.1.2	气流要求	23
	4.2.1.3	软件要求	24
	4.2.2	安全预防措施	24
	4.2.3	安装前检查清单	25
	4.2.4	电缆和模块	26
	4.2.5	识别系统中的卡	27
	4.2.6	提取说明	28
5	驱动程序罗	₹装	29
	5.1 Windows	驱动程序安装	29
	5.1.1 软件	要求	29
	5.1.2 下载	WinOF-2 驱动程序	30
	5.1.3 安装 9	WinOF-2 驱动程序	31
	5.1.3.1 7	有人值守的安装	31
	5.1.3.2	无人值守安装	41
	5.1.4 固件	升级	43
	5.2 Linux 驱动	b安装	43
	5.2.1 先决	条件	43
	5.2.2 下载	MLNX_OFED	44
	5.3 VMware 3	区动安装	69
	5.3.1 硬件	和软件要求	69
	5.3.2 安装运	5用于 VMware vSphere 的 NATIVE ESXi 驱动程序	70
	5.3.3 删除与	早期版本的 NVIDIA 驱动程序	70
	5.3.4 固件	编程	71
6	更新适配	器固件	72
7	设置高速	端口的链路类型	73
	7.1 mlnxconf	ig	73
	7.2 UEFI		73
8	故障排除		74
_	-XI T 2 11 1 W ***		,

8.1 常规故障排除	74
8.3 Windows 故障排除	76
规格	78
9.1 MCX75343AMC-NEAC / MCX75343AMS-NEAC 规格79	
9.2 MCX753436MS-HEAB / MCX753436MC-HEAB 规格8	1
9.3.1 电路板图纸和尺寸	83
9.3.2 支架图纸及尺寸	84
O 监控	85
10.1 热传感器	85
10.2 散热器	85
1 查找适配卡上的 MAC 地址和序列号	. 86
2 文档修订历史	. 88
	8.2 Linux 故障排除

1 关于本手册

这是基于 NVIDIA® ConnectX®-7 集成电路设备(适用于开放计算项目规范 3.0)的适配卡的用户指南。这些适配器的连接性为支持企业数据中心和高性能计算环境中使用的 OCP 规范 3.0 适配卡的服务器提供了性能最高、延迟最低且最灵活的互连解决方案。

1.1 订购部件编号

下表提供了适用于 OCP Spec 3.0 的 ConnectX-7 卡的订购部件编号 (OPN)。

NVIDIA SKU	旧版 OPN	端口类型	支持速度	OCP3.0 形狀因數	PCIe 表达	多主机 和 插座 直接的	加密	安全的 引导	支架类型	生命周期
900-9X760- 0018-MB2	MCX753436MC- Z炔蒸氨基苯甲酸	双端口 QSFP112	InfiniBand: NDR200/HDR 以太网: 200GbE (默认模式)	外形小巧	PCIe 5.0 x16	✓	✓	√	翼形螺钉(拉环) 括号	大量的 生产
900-9X760- 0078-MB0	MCX753436MS- 乙炔基氨基苯甲酸	双端口 QSFP112	InfiniBand: NDR200/HDR 以太网: 200GbE (默认模式)	外形小巧	PCIe 5.0 x16	✓	-	√	翼形螺钉(拉环) 括号	大量的 生产
900-9X7AX- 004NMC0	MCX75343AMC- 东北亜航空中心	单端口	InfiniBand: NDR (默认模式) 以太网: 400GbE	高小体型 因子 (TSFF)	PCIe 5.0 x16	√	√	✓	翼形螺钉(拉环) 特许金融期货括号	大量的 生产
900-9X7AX- 003NMC0	MCX75343AMS- 东北亚航空中心	单端口	InfiniBand: NDR (默认模式) 以太网: 400GbE	高小体型 因子 (TSFF)	PCIe 5.0 x16	√	-	√	翼形螺钉(拉环) 特许金融期货括号	大量的 生产

1.2 目标受众

本手册适用于这些卡的安装人员和用户。

本手册假设您基本熟悉 InfiniBand 和以太网网络和架构规范。

1.3 技术支持

直接从 NVIDIA 购买 NVIDIA 产品的客户可以通过以下方式与我们联系:

• 网址: <u>www.nvidia.com</u> →支持

• 电子邮件: enterprisesupport@nvidia.com

购买了 NVIDIA M-1 全球支持服务的客户,请参阅合同以了解有关技术支持的详细信息。通过 NVIDIA 认可的经销商购买 NVIDIA 产品的客户应首先通过其经销商寻求帮助。

1.4 相关文档

MLNX_OFED for Linux 用 户手册和发行说明	用户手册描述了 OFED 的功能、性能、频带诊断、工具内容和配置。请参阅 MLNX_OFED Linux 文档。
WinOF-2 for Windows 用 户手册和发行说明	用户手册描述了 WinOF-2 的功能、性能、以太网诊断、工具内容和配置。请参阅 <u>WinOF-2 for</u> <u>Windows 文档</u> 。
NVIDIA VMware 以太网用户 手册	用户手册和发行说明描述了 NVIDIA ConnectX [®] NATIVE ESXi 堆栈的各个组件。请参阅 <u>VMware[®] ESXi</u> <u>驱动程序文档</u> 。
NVIDIA 固件实用程序 (mlxup) 用户手册和发行说明	NVIDIA 固件更新和查询实用程序,用于更新固件。请参阅 <u>固件实用程序 (mlxup) 文档</u> 。
NVIDIA 固件工具 (MFT) 用户 手册	描述单个节点的 MFT 固件管理工具集的用户手册。请参阅 <u>MFT 用户手册</u> 。

InfiniBand 架构规范版本 1.2.1,第 2 卷 - 版本 1.4 和第 2 卷 - 版本 1.5	InfiniBand 规格
IEEE 标准 802.3 规范	IEEE 以太网规范
开放计算项目 3.0 规格	https://www.opencompute.org/
PCI Express 规格	行业标准 PCI Express 基础和卡机电规格。请参阅 <u>PCI-SIG 规范</u> 。
LinkX 互连解决方案	LinkX InfiniBand 电缆和收发器旨在最大限度地提高高性能计算网络的性能,需要在计算节点和交换机节点之间建立高带宽、低延迟的连接。NVIDIA 提供业界最广泛的 QDR/FDR10 (40Gb/s)、FDR (56Gb/s)、EDR/HDR100 (100Gb/s)、HDR (200Gb/s) 和 NDR (400Gb/s) 电缆产品组合之一,包括直连铜缆 (DAC)、铜分路器电缆、有源光缆 (AOC) 和长度范围从 0.5 米到 10 公里的收发器。除了满足 IBTA 标准外,NVIDIA 还在端到端环境中测试每款产品,确保误码率低于 1E-15。了解更多信息,请访问LinkX 电缆和收发器。

1.5 文档约定

在讨论内存大小时,本文档使用 MB 和 MBytes 来表示以兆字节为单位的大小。使用 Mb 或 Mbits(小写 b)表示以兆位为单位的大小。本文档使用 PCIe 来表示 PCI Express。

2 简介

2.1 产品概述

这是*用户指南*基于 ConnectX-7 集成电路设备的 InfiniBand/VPI 适配卡,符合 OCP Spec 3.0 标准。这些适配器的连接性为企业数据中心和高性能计算环境中使用的支持 OCP 3.0 的服务器提供了性能最高、最灵活的互连解决方案。ConnectX-7 网络适配器有两种外形和各种类型:直立式 PCIe 和开放计算项目 (OCP) Spec 3.0 卡。本用户手册 涵盖 OCP 3.0 卡,有关低调 PCIe 直立式卡硬件用户手册,请参阅ConnectX-7 PCIe 直立卡用户手册。

ConnectX-7 提供最多两个 NDR200 和 200GbE 端口或一个 NDR 和 400GbE 连接端口以及 PCIe Gen 5.0 x16 主机连接,是 NVIDIA 世界一流、屡获殊荣的 ConnectX 系列网络适配器的成员。ConnectX-7 延续了 NVIDIA 在网络领域的持续创新,在各个规模上都提供灵活性和效率。ConnectX-7 为不折不扣的数据中心提供尖端的性能和安全性。

除了小型 (SFF) 规格外,ConnectX-7 for OCP 3.0 卡还提供新添加的高 SFF (TSFF) 规格规格,考虑到卡的增加高度,以实现更好的散热性能。请联系 NVIDIA 了解更多详细信息。

2.2 系统要求

物品	描述
PCI Express 插槽	在 PCle x16 配置中 PCle Gen 5.0/4.0(16 GT/s / 32GT/s)通过 x16 边缘连接器。
系统电源 供应	參閱 <u>规格</u>
操作系统	 适用于主要操作系统的内置驱动程序: Linux: RHEL、Ubuntu 视窗 虚拟化和容器 VMware ESXi (SR-IOV) Kubernetes OpenFabrics 企业发行版 (OFED) OpenFabrics Windows 发行版 (WinOF-2)
连接	 可与 1/10/25/40/50/100/200/400 Gb/s 以太网交换机以及 SDR/DDR/EDR/HDR100/HDR/NDR200/NDR InfiniBand 交换机互操作 具有 ESD 保护功能的无源铜缆 用于光学和有源电缆支持的供电连接器

2.3 包装内容

类别	数量	物品
牌	1	适用于 OCP 3.0 的 ConnectX-7 适配卡
配件	1	翼形螺钉(拉环)支架

有关详细信息,请参阅<u>规格</u> 。

2.4 特点和优点

本节介绍硬件特性和功能。请参阅相关驱动程序和/或固件版本说明以了解功能可用性。

	1							
特征	描述	描述						
无限带宽 建筑学 规范 v1.5 合规	ConnectX-7 为高性能计算 (HPC)、人工智能 (AI) 和超大规模云数据中心应用提供低延迟、高带宽和计算效率。 ConnectX-7 符合 InfiniBand 架构规范 v1.5。 InfiniBand 网络协议和速率:							
	协议	标准	标准 速率	(Gb/s)	评论			
			4x 端口 (4 车道)	2x 端口 (2 车道)				
	无交易记录/无交易记录200	IBTA 第 2 卷 1.5	425	212.5	PAM4 256b/257b 编码和 RS-FEC			
	HDR/HDR100	IBTA 第 2 卷 1.4	212.5	106.25	PAM4 256b/257b 编码和 RS-FEC			
	增强数据链路层	IBTA Vol2 1.3.1	103.125	51.5625	NRZ 64b/66b 编码			
	罗斯福	IBTA 第 2 卷 1.2	56.25	不适用	NRZ 64b/66b 编码			
高达 400Gb/s 以太网		ConnectX-7 适配卡符合以下 IEEE 802.3 标准: 400GbE / 200GbE / 100GbE / 50GbE / 40GbE / 25GbE / 10GbE						
	协议			平均可动用利率				
	IEEE802.3ck	IEEE802.3ck			100/200/400Gb/s 千兆以太网 (包括 ETC 增强功能)			
	IEEE802.3cd	IEEE802.3cd			50/100/200/400Gb/s 千兆以太网			
	IEEE802.3bs	IEEE802.3bs			C增强功能)			
	IEEE802.3厘米							
	IEEE802.3cn							
	IEEE802.3cu							

特征	描述			
	IEEE 802.3bj IEEE 802.3bm	100 千兆以太网		
	IEEE 802.3by 以太网联盟25	25/50千兆以太网		
	IEEE 802.3ba	40千兆以太网		
	IEEE 802.3ae	10 千兆以太网		
	IEEE 802.3cb	2.5/5 千兆以太网 (对于 2.5:仅支持 2.5 x1000BASE-X)		
	IEEE 802.3ap	基于自动协商和KR启动		
	IEEE 802.3ad IEEE 802.1AX	链路聚合		
	IEEE 802.1Q IEEE 802.1P VLAN 标签和优先级			
	IEEE 802.1Qau(QCN) 拥塞通知 IEEE 802.1Qaz (ETS) EEE 802.1Qbb(PFC) IEEE 802.1Qbg IEEE 1588v2 IEEE 802.1AE(MAC安全) 巨型帧支持 (9.6KB)			
记忆 成分	 SPI——包括256Mbit SPI 四闪存设备。 FRU EEPROM - 存储卡的参数和个性。EEPROM 容量为 128Kbit(或 16KB)。FRU I2C 地址为 (0x50) 通过 PCIe SMBus 访问。(注:地址 0x58 被保留。) 			
覆盖网络	为了更好地扩展网络,数据中心运营商通常会创建覆盖网络,通过逻辑隧道以 NVGRE 和 VXLAN 等封装格式传输来自各个虚拟机的流量。虽然这解决了网络可扩展性问题,但它会向硬件卸载引擎隐藏 TCP 数据包,从而给主机带来更高的负载			

特征	描述
	CPU。ConnectX-7 通过提供封装和解封装覆盖协议的先进 NVGRE 和 VXLAN 硬件卸载引擎,有效地解决了这个问题。
服务质量	支持基于端口的服务质量,满足各种应用对延迟和 SLA 的要求。
基于硬件的 I/O 虚拟化	ConnectX-7 为服务器内的虚拟机提供专用适配器资源和保证的隔离和保护。
存储加速	与多结构网络相比,整合的计算和存储网络具有显著的性价比优势。标准块和文件访问协议可以利用: · 用于高性能存储访问的 RDMA · NVMe over Fabric 为目标机器卸载 · NVMe over TCP 加速
SRIOV	ConnectX-7 SR-IOV 技术为服务器内的虚拟机 (VM) 提供专用适配器资源并保证隔离和保护。
高性能 加速度	 集体操作卸载 向量集合操作卸载 MPI 标签匹配 MPI_Alltoall 卸载 会合协议卸载
RDMA 消息速率	每秒3.3至3.7亿条消息。
安全启动	安全启动过程可确保启动要在 ConnectX-7 上运行的真实固件/软件。这是通过使用非对称加密的加密原语实现的。 ConnectX-7 在其 HW 信任根 (RoT) 中支持多种加密功能,其密钥存储在片上 FUSES 中。
安全固件 更新	安全固件更新功能使设备能够验证新固件二进制文件的数字签名,以确保只有经过官方批准的版本才能从主机、网络或板卡管理控制器 (BMC) 安装。具有"安全固件更新"功能(安全 FW)的设备固件会限制对特定命令和寄存器的访问,这些命令和寄存器可用于修改闪存上的固件二进制映像,以及可能危害一般安全性的命令。如需了解更多信息,请参阅MFT 用户手册。

特征	描述
高级存储 能力	块级加密和校验和卸载。
主机管理	ConnectX-7 技术通过 BMC 保持对主机可管理性的支持。ConnectX-7 PCIe 直立式适配器可以使用 MCTP over SMBus 或 MCTP over PCIe 协议连接到 BMC,就像它是标准的 NVIDIA PCIe 直立式适配器卡一样。如需为服务器使用的特定可管理性解决方案配置适配器,请联系 NVIDIA 支持。 · 协议: PLDM、NCSI · 传输层 – RBT、SMBus 上的 MCTP 和 PCIe 上的 MCTP · 物理层: 用于设备控制和配置的SMBus 2.0 / I2C接口,PCIe · 用于监控和控制 DSP0248 的 PLDM · 用于固件更新的 PLDM DSP026 · IEEE 1149.6 · 安全固件更新 · FW 恢复 · 网卡重置 · 监测和控制 · 网络端口设置 · 启动设置
RDMA 和 RDMA 过度收敛 以太网 (RoCE)	ConnectX-7 利用 IBTA RDMA(远程数据内存访问)和 RoCE(融合以太网上的 RDMA)技术,在 InfiniBand 和以太 网网络上提供低延迟和高性能。利用数据中心桥接 (DCB) 功能以及 ConnectX-7 高级拥塞控制硬件机制,RoCE 在第 2 层和第 3 层网络上提供高效的低延迟 RDMA 服务。
NVIDIA PeerDirect™	PeerDirect™ 通信通过消除 PCIe 总线上组件之间不必要的内部数据复制(例如从 GPU 到 CPU),提供高效的 RDMA 访问,从而显著缩短应用程序运行时间。ConnectX-7 先进加速技术可实现更高的集群效率和数万个节点的可扩展性。
CPU 卸载	适配器功能可以减少 CPU 开销,从而提供更多的 CPU 可用于计算任务。 ・ 灵活的匹配动作流表 ・ 使用 ASAP 进行 Open VSwitch (OVS) 卸载2®

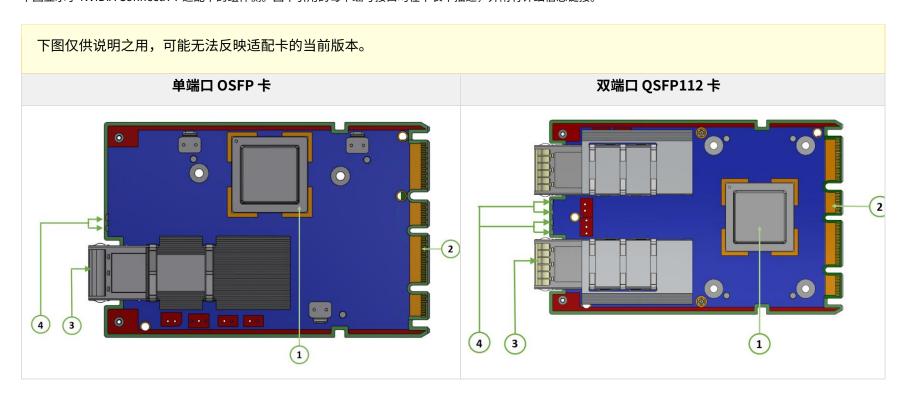
特征	描述
	・ 隧道封装/解封装
NVIDIA 多主机	NVIDIA® Multi-Host 技术使下一代云、Web 2.0 和高性能数据中心能够设计和构建新的横向扩展异构计算和存储机架,并在多个主机和集中式网络控制器之间实现直接连接。这可以实现最低延迟的直接数据访问,从而显著提高密度并最大限度地提高数据传输速率。如需了解更多信息,请访问NVIDIA 多主机解决方案。

3 支持的接口

本节介绍 ConnectX-7 支持的接口。图中引用的每个编号接口均在下表中描述,并附有详细信息链接。

3.1 ConnectX-7 布局和接口信息

下图显示了 NVIDIA ConnectX-7 适配卡的组件侧。图中引用的每个编号接口均在下表中描述,并附有详细信息链接。



物品	界面	描述
1	ConnectX-7 IC	ConnectX-7 集成电路
2	PCI Express 界面	PCle Gen 5.0 通过 x16 边缘连接器
	<u>联网</u> 接口	网络流量通过适配卡 QSFP112/OSFP 连接器传输。网络连接器允许使用模块、光纤和无源电缆互连解决方案。
4	网络端口	每个端口有两个 I/O LED,用于指示速度和链路状态
	可擦除只读存储器	FRU EEPROM容量16KB
	SMBus 接口	允许使用 MCTP over SMBus 或 MCTP over PCIe 协议进行 BMC 连接。
	电压 监管机构	为板载调节器供电的电压供应引脚。
	CPLD 接口	控制网络端口逻辑 LED(LED0)并实现 OCP 3.0 主机扫描链。
	<u>散热器</u>	散发热量。

3.2 接口详细说明

3.2.1 ConnectX-7 芯片

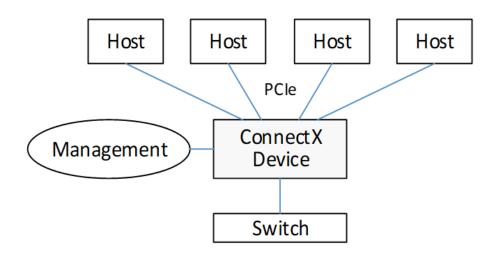
ConnectX-7 系列适配器 IC 设备提供两个 NDR200/200GbE 端口或一个 NDR/400GbE 连接端口,并配备一流的硬件功能,可加速和保护云和数据中心工作负载。

3.2.1.1 NVIDIA Multi-HostTM 支持

除了为数据中心构建异常高的带宽之外,ConnectX-7设备还可以利用其 NVIDIA Multi-Host 功能在整个数据中心内发挥这种速度。

使用其 16 通道 PCI Express 接口,单个 ConnectX-7 设备可以为最多四个独立主机提供 400GbE 互连,而不会降低性能。

下图显示了带有 NVIDIA Multi-Host 的 ConnectX-7 设备连接到四个独立主机,每个主机都有一个 PCIe x4 接口,一侧连接到另一侧的交换机。



对于适配器的 x16 PCIe 接口,以下分叉是可选的:

x1 PCle x16、x1 PCle x8、x1 PCle x4

- x2 PCle x8、x2 PCle x4、x2 PCle x2、x2 PCle x1
- x4 PCle x4、x4 PCle x2、x4 PCle x1

支持多主机的卡还支持 Socket-Direct 应用程序,并可作为常规的单主机卡工作,具体取决于它们插入的服务器类型,假设该服务器符合 OCP 3.0 规范。

根据 OCP 3.0 规范,适配卡通过 PRSNTB[3:0]# 引脚来宣传其功能。服务器通过 BIF[2:0]# 引脚确定配置,并将其驱动到适配卡。

这**NVIDIA OCP3.0卡**具有使用 BIF[2:0]#数据并确定启动的正确操作模式的内部逻辑。PRSNTB[3:0]#和 BIF[2:0]#引脚的组合确定性地设置了给定 OCP 3.0 卡和底板组合的 PCIe 通道宽度。逻辑和解码表可在 OCP 3.0 规范(第 1 章)中找到**3.5 PCIe分叉机制)**。

例如:

然后VIDIA OCP 3.0 多主机适配器驱动器将 PRSNTB[3:0]# 上的 0100 发送至服务器。

如果服务器驱动程序	适配器 PCIe 模式
000	单主机模式: x1 PCle x16
001	套接字直接模式: x2 PCIe x8
010	套接字直接模式: x4 PCle x4
101	多主机模式: x2 PCle x8
110	多主机模式: x4 PCle x4

3.2.2 PCI Express 接口

下表描述了 ConnectX-7 OCP 3.0 适配卡中支持的 PCIe 接口。

- 符合 PCIe Gen 5.0 标准,兼容 4.0、3.0、2.0 和 1.1
- 2.5、5.0、8.0、16.0 和 32GT/s 链路速率 x16
- 支持 PCIe 分叉: 自动协商为 x16、x8、x4、x2 或 x1
- · NVIDIA Multi-Host™ 支持最多 4 台主机的连接
- 事务层数据包 (TLP) 处理提示 (TPH)

- PCIe 交换机下游端口遏制 (DPC)
- 高级错误报告 (AER)
- 用于点对点安全通信的访问控制服务 (ACS)
- 进程地址空间 ID (PASID)
- 地址翻译服务 (ATS)
- 支持 MSI/MSI-X 机制
- ・ 支持 SR-IOV

3.2.3 网络接口

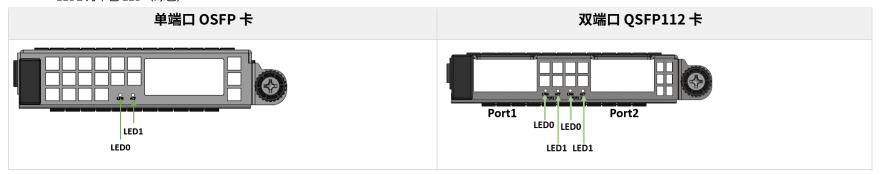
适配卡包含特殊电路,用于在插入铜缆时保护卡/服务器免受 ESD 冲击。

CHO! CHISM CHI / NOS EMA (I) MASSING (I) MASSING CHISM		
协议	规格	
以太网	网络端口符合 IEEE 802.3 以太网标准 <u>特点和优点</u> . 以太网流量通过适配卡上的网络连接器传输。	
无限带宽	网络端口符合 <i>InfiniBand 架构规范,版本 1.5。</i> InfiniBand 流量通过卡的网络连接器传输。	

3.2.4 网络端口 LED 规格

每个端口有两个 I/O LED,用于指示端口速度和链路状态。

- · LED0 是双色 LED(黄色和绿色)
- LED2 为单色 LED (绿色)



状态	双色 LED (i	黄色/绿色)		单色 LED
用于定位适配卡的 Beacon 命令	命令 1Hz 闪烁黄色		离开	
错误	4Hz 闪烁黄色 表示链接错误。错误可能是以下之一:			在
	错误类型	描述	LED 行为	
	我2碳	我 ₂ C 访问网络端口失败	闪烁直至错误修复	
	超过- 当前的	网络端口的过流情况	闪烁直至错误修复	
衔接	全端口速度:绿色 LED 常亮端口速度低于全速时:黄色 LED 常亮			在
体力活動	绿色 LED 将闪烁。		闪烁	

3.2.5 FRU 电可擦除只读存储器

FRU EEPROM 允许底板识别不同类型的 OCP 3.0 卡。FRU EEPROM 可通过 SMCLK 和 SMDATA 访问。FRU EEPROM 地址根据 SLOT_ID0 和 SLOT_ID1 定义,其容量为 16KB。

3.2.6 SMBus 接口

ConnectX-7 技术通过 BMC 保持对可管理性的支持。ConnectX-7 OCP 3.0 适配器可以使用 MCTP over SMBus 或 MCTP over PCIe 协议连接到 BMC,就像它是标准 NVIDIA OCP 3.0 适配器一样。如需为服务器使用的特定可管理性解决方案配置适配器,请联系 NVIDIA 支持。

3.2.7 电压调节器

该适配卡集成了实现OCP 3.0主机扫描链的CPLD器件和c控制网络端口逻辑 LED (LED0). 它从 3.3V_EDGE 和 12V_EDGE 导轨获取电源。

4 硬件安装

安装和初始化ConnectX-7适配卡OCP Spec 3.0需要注意机械属性、电源规格以及电子设备的注意事项。

4.1 安全警告

此处提供英语版安全警告。如需其他语言版安全警告,请参阅适配器安装安全说明。

4.2 安装步骤概述

安装过程包括以下步骤:

步	程序	直接链接
1	检查系统的硬件和软件要求。	參閱 <u>系统要求</u>
2	注意主机系统内的气流考虑	參閱 <u>气流要求</u>
3	遵守安全预防措施	參閱安全预防措施
4	遵循安装前检查表	參閱安装前检查清单
5	(可选)将组装好的 OCP 3.0 支架更换为所需尺寸的支架	參閱OCP 3.0 支架更换说明
6	在系统中安装适配卡	參閱 <u>安装说明</u>
7	将电缆或模块连接到卡上	參閱 <u>电缆和模块</u>
8	识别系统中的适配卡	參閱 <u>识别系统中的卡</u>

4.2.1 系统要求

除非另有说明,NVIDIA产品设计用于环境受控、气体和灰尘(颗粒)污染程度较低的数据中心。

操作环境应满足 ISA 71.04 对气体污染的严重程度 G1 等级和 ISO 14644-1 对清洁度的 8 级等级要求。

4.2.1.1 硬件要求

为了确保正常运行和性能,请确保使用具有相应总线宽度且能为您的卡提供足够电力的 PCIe 插槽。有关更多电力要求,请参阅手册的"规格"部分。

安装该卡需要具有符合 OCP 规范 3.0 的 PCI Express x16 插槽的系统。

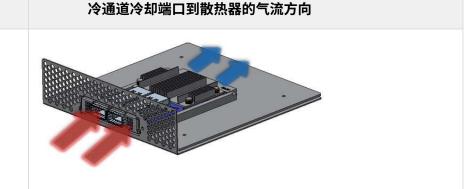
4.2.1.2 气流要求

ConnectX-7 适配卡提供两种气流模式:从散热器到网络端口,反之亦然,如下所示。

请参阅"规格"章节以了解每种特定卡型号的气流数字。

系统中的所有卡都应规划相同的气流方向。

热通道冷却散热器到端口的气流方向



4.2.1.3 软件要求

- 请参阅简介部分下的操作系统/发行版部分。
- 软件堆栈 NVIDIA OpenFabric 软件包 MLNX_OFED(适用于 Linux)、WinOF-2(适用于 Windows)和 VMware。请参阅驱动程序安装 部分。

4.2.2 安全注意事项

适配器安装在使用可能致命的电压的系统中。打开系统机壳之前,请遵守以下预防措施,以避免受伤并防止损坏系统组件。

- 1. 取下手上和手腕上的所有金属物体。
- 2. 确保仅使用绝缘工具。
- 3. 确认系统已关闭并且已拔下电源插头。
- 4. 强烈建议使用 ESD 腕带或其他防静电设备。

4.2.3 安装前检查清单

1.**打开 ConnectX-7 适配卡的包装。**打开包装并取出 ConnectX-7 卡。检查零件是否有在运输过程中发生的明显损坏。请注意,必须将卡放在防静电表面上。

请注意,如果匆忙将卡从防静电袋中取出,塑料密封袋可能会损坏网络连接器上的 EMI 指。小心地将卡从防静电袋中取出,以免损坏 EMI 指。

2.如果系统处于活动状态,请关闭系统:

关闭系统电源,并断开电源线。请参阅系统文档以获取说明。在安装 ConnectX-7 卡之前,请确保系统已断开电源。

4.2.4 电缆和模块

要获取您的适配器支持的 NVIDIA 电缆列表,请参阅电缆参考表网络配置工具。

电缆安装

- 1. 在设备通电的情况下可以插入或拔出所有电缆。
- 2. 要插入电缆,请将连接器按入端口插座,直至连接器牢固就位。
 - a. 在将电缆连接到适配卡之前,支撑电缆的重量。可以使用电缆固定器或将电缆绑在机架上。
 - b. 插入连接器前,确定连接器与卡的正确方向。不要尝试将连接器倒置插入。这可能会损坏适配卡。
 - c. 将连接器插入适配卡。小心地将连接器笔直插入笼子。不要对适配卡中的连接器笼子施加任何上下扭矩。
 - d. 确保连接器锁定到位。

安装电缆时,确保闩锁已啮合。

安装和移除电缆时,务必通过与卡沿直线推或拉电缆和连接器来安装和移除电缆。

- 3. 将电缆插入端口后,黄色或绿色 LED0 指示灯将在物理连接建立时亮起(即,当设备通电并将电缆插入端口,而连接器的另一端插入正常工作的端口时)。请参阅<u>适配</u> 卡 LED 操作 。
- 4. 插入电缆后,使用电缆供应商特有的锁定机制锁定连接器。当建立逻辑连接时,绿色 LED1 将亮起。当正在传输数据时,绿色 LED1 将闪烁。
- 5. 应注意不要阻碍通过通风孔排出的空气。使用长度合适的电缆,使其能够水平绕过底盘的一侧,然后再在机架中向上或向下弯曲。
- 6. 要移除电缆,请松开锁,然后慢慢将连接器从端口插座中拉出。当电缆拔出时,LED 指示灯将熄灭。

要卸载适配卡,请参阅卸载卡。

4.2.5 识别系统中的卡

在Linux上

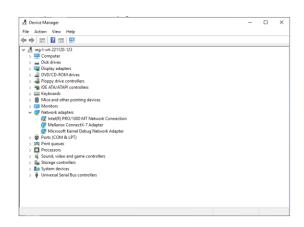
通过运行 lspci 并定位包含字符串 "Mellanox Technologies"的行来获取 PCI 总线上的设备位置:

Ispci | grep -i Mellanox

网络控制器: Mellanox Technologies MT2910系列[ConnectX-7]

在Windows上

- 1. 在服务器上打开设备管理器。单击**开始=>跑步**,然后输入**设备管理控制台**。
- 2. 扩展**系统设备**并找到您的 NVIDIA ConnectX-7 适配卡。
- 3. 右键单击 适配器行并选择特性显示适配卡属性窗口。
- 4. 点击细节选项卡并选择硬件 ID(Windows 2022/2019/2016/2012 R2)从财产下拉菜單。 PCI 设备(示例)



- 1.在**价值**显示框中,选中字段 VEN 和 DEV(字段以"&"分隔)。在上面的显示示例中,请注意子字符串"PCI\VEN_15B3&DEV_1021": VEN 等于 0x15B3 这是 Mellanox Technologies 的供应商 ID; DEV 等于 1021(对于 ConnectX-7)
 - 这是一个有效的 NVIDIA PCI 设备 ID。

如果 PCI 设备没有 NVIDIA 适配器 ID,请返回步骤 2 检查另一个设备。

NVIDIA PCI 设备 ID 列表可在以下位置找到: PCI ID 存储库。

4.2.6 提取说明

无法呈现包含或摘录包含。无法检索页面。

5 驱动程序安装

请使用相关的驱动程序安装部分。

- · Windows 驱动程序安装
- · Linux 驱动程序安装
- · VMware 驱动程序安装

5.1 Windows 驱动程序安装

对于 Windows,请通过 NVIDIA 网站下载并安装最新的适用于 Windows 的 WinOF-2 软件包,网址为: <u>WinOF-2 网页</u> . 按照下载包中包含的安装说明进行操作(也可从下载页面获取)。

以下部分中的快照仅用于说明目的。安装界面可能会略有不同,具体取决于所使用的操作系统。

5.1.1 软件要求

描述	包裹
Windows Server 2022	MLNX_WinOF2-<版本>_All_x64.exe
Windows 服务器 2019	
Windows 服务器 2016	
Windows Server 2012 R2	
Windows 11 客户端(仅 64 位)	
Windows 10 客户端(仅 64 位)	

描述	包裹
Windows 8.1 客户端(仅 64 位)	

注意:上面列出的操作系统必须以管理员权限运行。

5.1.2 下载 WinOF-2 驱动程序

- 》 要根据您的操作系统下载.exe 文件,请按照以下步骤操作:
 - 1.获取机器架构。
 - a. 要转到"开始"菜单,请将鼠标放在屏幕远程桌面的右下角。
 - b. 打开 CMD 控制台(单击任务管理器-->文件 --> 运行新任务并输入 CMD)。
 - c.输入以下命令。

选择%处理器架构%

在 x64 (64 位) 机器上,输出将是"AMD64"。

- 2. 转到 WinOF-2 网页: https://www.nvidia.com/en-us/networking/ > 产品 > 软件 > InfiniBand 驱动程序(了解更多) > Nvidia WinOF-2。
- 3. 根据你机器的架构下载 .exe 镜像(见 $\underline{步骤1}$)。.exe 的名称格式如下:MLNX_WinOF2-
<version>_<arch>.exe。

禁止安装错误的.exe文件。如果这样做,将显示错误消息。

例如,如果您在 32 位计算机上安装 64 位 .exe,向导将显示以下(或类似)错误消息:"此处理器类型不支持安装程序包。请联系您的供应商"

5.1.3 安装 WinOF-2 驱动程序

以下章节中的快照仅用于说明目的。根据所使用的操作系统,安装界面可能会略有不同。

本节提供两种安装过程的说明,都需要管理员权限:

- <u>有人值守安装</u> 需要用户频繁干预的安装过程。
- <u>无人值守安装</u> 无需用户干预的自动化安装过程。

5.1.3.1 有人看管的安装

以下是安装会话的示例。

- 1. 双击 .exe 并按照 GUI 说明安装 MLNX_WinOF2。
- 2.[选修的]手动配置您的设置以包含日志选项(用相关目录替换"LogFile")。

LNX_WinOF2_<修订版本>_All_Arch.exe /v"/l*vx [日志文件]"

3.**[选修的]**如果您不想升级您的固件版本(即MT_SKIPFWUPGRD默认值为False)。

LNX_WinOF2_<修订版本>_All_Arch.exe /v "MT_SKIPFWUPGRD = 1"

4.[选修的]如果您不想安装 Rshim 驱动程序,请运行。

LNX_WinOF2_<修订版本>_All_Arch.exe /v "MT_DISABLE_RSHIM_INSTALL = 1"

如果已安装先前的 Rshim 驱动程序,则 Rshim 驱动程序安装将失败。日志中将显示以下失败消息:

"错误!! 由于以下错误导致安装失败: MlxRshim 驱动程序安装已禁用且 MlxRshim 驱动程序已安装,请从驱动程序存储区中删除以下 oem inf 文件: <oem inf 列表>"

5.[选修的]如果您想跳过对不受支持的设备的检查,请运行。

LNX_WinOF2_<修订版本>_All_Arch.exe /v "SKIPUNSUPPORTEDDEVCHECK=1"

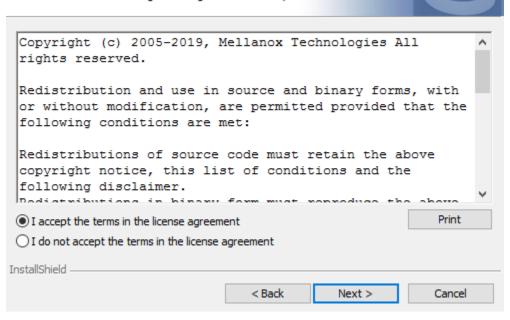
6. 在欢迎屏幕中单击"下一步"。



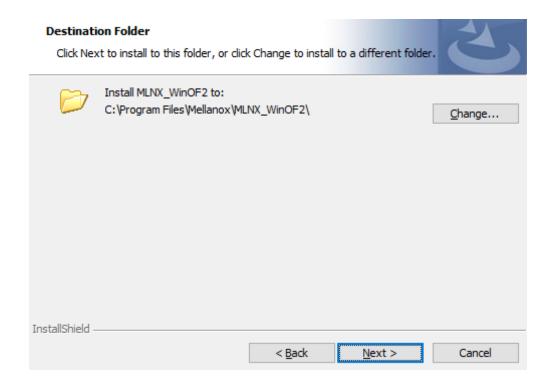
7.阅读并接受许可协议,然后单击下一步。

License Agreement

Please read the following license agreement carefully.



8.选择安装的目标文件夹。



9. 以下情况将显示固件升级界面:

- o 如果用户有OEM卡。在这种情况下,固件将不会显示。
- o 如果用户拥有的是标准 NVIDIA® 卡,且固件版本较旧,则固件将相应更新。但是,如果用户同时拥有 OEM 卡和 NVIDIA® 卡,则只有 NVIDIA® 卡会更新。

Firmware Upgrade



Cancel

Upgrade the HCA's firmware version (Recommended).

Upgrading the firmware version will reboot your machine.

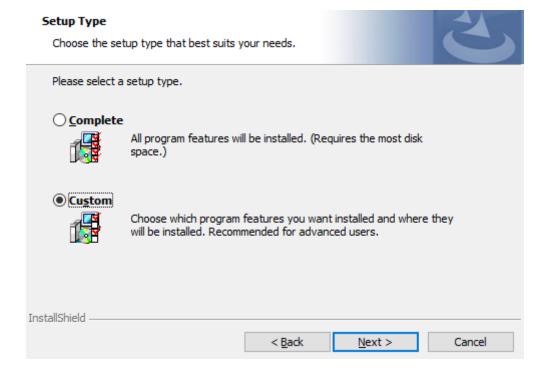
Note: One or more of your HCA adapters has an old firmware version. We recommend upgrading to a newer firmware version to enable improved functionality and support driver's capabilities.

< Back

Next >

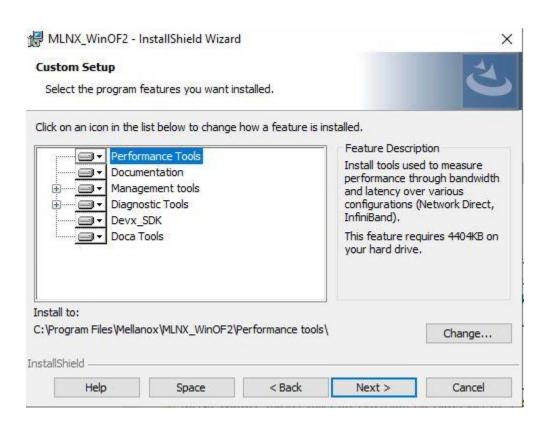
10. 选择"完整"或"自定义"安装,然后按照步骤 向前。

InstallShield

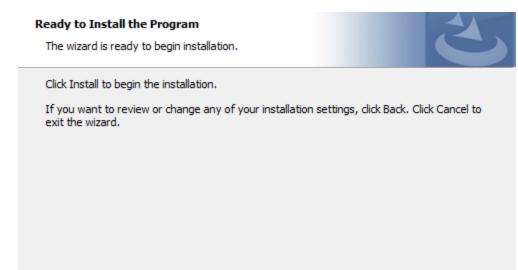


a. 选择要安装的功能:

- 性能工具-安装用于测量用户环境中性能的性能工具
- 文档 包含用户手册和发行说明
- 管理工具——用于管理的安装工具,例如 mlxstat
- 诊断工具 用于诊断的安装工具,例如 mlx5cmd
- b. 单击"下一步"安装所需的工具。



11.单击"安装"开始安装。



< Back

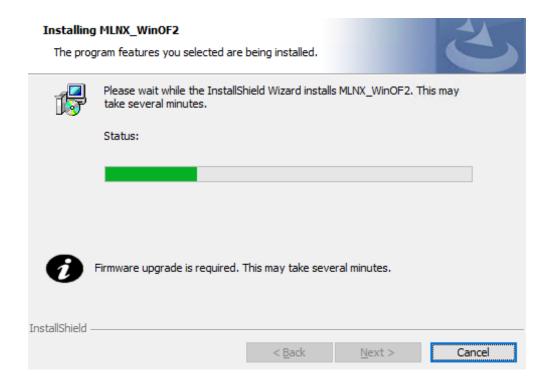
<u>I</u>nstall

Cancel

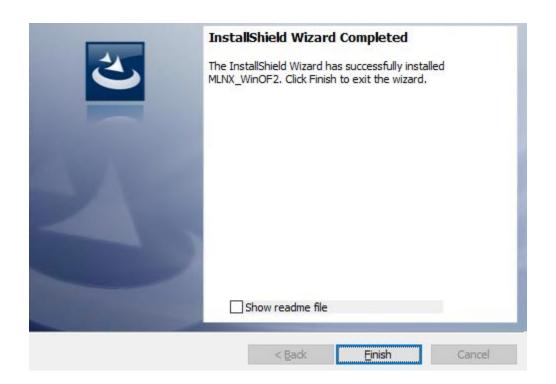
12. 如果选中了固件升级选项步骤7,如果需要固件升级,您将收到通知(请参阅)。

InstallShield

0



13.单击"完成"以完成安装。



5.1.3.2 无人值守安装

如果未指定重新启动选项,安装程序将在必要时重新启动计算机,而不会向用户显示任何提示或警告。

要控制重启,请使用/启动或者/强制重启标准命令行选项。

以下是无人值守安装会话的示例。

- 1.打开CMD控制台->单击"开始"->任务管理器文件->运行新任务->并输入 CMD。
- 2.安装驱动程序。运行:

LNX_WinOF2-[驱动程序/版本]_<修订版本>_All_-Arch.exe /S /v/qn

3.[选修的]手动配置您的设置以包含日志选项:

LNX_WinOF2-[驱动程序/版本]_<修订版本>_All_-Arch.exe /S /v/qn /v"/l*vx [日志档案]"

4.[选修的]如果你想控制是否安装 ND 提供程序(即MT_NDPROPERTY 默认值为 True)。

LNX_WinOF2-[驱动程序/版本]_<修订版本>_All_Arch.exe /vMT_NDPROPERTY=1

5.[选修的]如果您不想升级固件版本(即MT_SKIPFWUPGRD默认值为 False)。

LNX_WinOF2-[驱动程序/版本]_<修订版本>_All_Arch.exe /vMT_SKIPFWUPGRD=1

6.[选修的]如果您不想安装 Rshim 驱动程序,请运行。

LNX_WinOF2_<修订版本>_All_Arch.exe /v "MT_DISABLE_RSHIM_INSTALL = 1"

如果已安装先前的 Rshim 驱动程序,则 Rshim 驱动程序安装将失败。日志中将显示以下失败消息:

"错误!! 由于以下错误导致安装失败: MlxRshim 驱动程序安装已禁用且 MlxRshim 驱动程序已安装,请从驱动程序存储区中删除以下 oem inf 文件: <oem inf 列表>"

7.[选修的]如果要启用 Rivermax 的默认配置,请运行。

LNX_WinOF2_<修订版本>_All_Arch.exe /v "MT_RIVERMAX=1 /l*vx C:\Users\<用户>\log.txt"

8.[选修的]如果要跳过对不受支持设备的检查,请运行/

LNX_WinOF2_<修订版本>_All_Arch.exe /v "SKIPUNSUPPORTEDDEVCHECK=1"

5.1.4 固件升级

如果机器配有标准 NVIDIA® 卡且固件版本较旧,则固件将作为 NVIDIA® WinOF-2 软件包安装的一部分自动更新。有关如何手动升级固件的信息,请参阅MFT 用户手册。

如果机器具有 DDA(直通)功能,则仅在主机中支持固件更新。因此,要更新固件,必须执行以下操作:

- 1. 将网络适配器返回给主机。
- 2. 按照以下步骤更新固件MFT 用户手册。
- 3. 使用 DDA 工具将适配器重新连接到 VM。

5.2 Linux 驱动安装

本节介绍如何在安装了 ConnectX-7 适配卡的单台服务器上安装和测试 MLNX_OFED for Linux 包。

5.2.1 先决条件

要求	描述
平台	安装了 ConnectX-7 InfiniBand/以太网适配卡的服务器平台。
安装所需的磁盘空间	1GB

要求	描述
操作系统	Linux 操作系统。 有关受支持的操作系统发行版和内核的列表,请参阅 <i>MLNX_OFED 发行说明</i> 。
安装人员权限	安装需要目标机器上的管理员(root)权限。

5.2.2 下载 MLNX_OFED

1. 通过运行 lspci 命令验证系统是否安装了网络适配器。下表提供了每个 ConnectX-7 卡配置的输出示例。

ConnectX-7 卡 配置	输出示例
单端口 PCIe x16 卡	[root@mftqa-009~]# lspci grep mellanox -ia 3:00.0 Infiniband 控制器: Mellanox Technologies MT2910 系列 [ConnectX-7]
双端口 PCIe x16 卡片	[root@mftqa-009~]# lspci grep mellanox -ia 86: 00.0 网络控制器: Mellanox Technologies MT2910 系列 [ConnectX-7] 86: 00.1 网络控制器: Mellanox Technologies MT2910 系列 [ConnectX-7]

2. 将 ISO 映像下载到您的主机。

图片名称的格式为MLNX_OFED_LINUX-<ver>--OS 标签><CPU arch>.iso。

您可以通过 NVIDIA 网站下载并安装最新的 OpenFabrics Enterprise Distribution (OFED) 软件包,网址为: <u>nvidia.com/en-us/networking→</u> 产品→软件 →InfiniBand 驱动程序→<u>NVIDIA MLNX_OFED</u>

一个。

- i. 向下滚动到下载向导,然后单击下载选项卡。
- ii. 根据您的主机操作系统选择相关的软件包。
- iii. 单击所需的 ISO/tgz 包。

iv. 要获取下载链接,请接受最终用户许可协议 (EULA)。

3. 使用 Hash 实用程序确认 ISO 映像的文件完整性。运行以下命令并将结果与 下载页面上提供的值进行比较。

HA256 MLNX_OFED_LINUX-<版本>-<操作系统标签>.iso

安装 MLNX_OFED

安装脚本

安装脚本 mlnxofedinstall 执行以下操作:

- o 发现当前安装的内核
- o 卸载标准操作系统发行版或其他供应商的商业堆栈中的任何软件堆栈
- o 安装 MLNX_OFED_LINUX 二进制 RPM(如果它们适用于当前内核)
- o 识别当前安装的 InfiniBand 和以太网网络适配器并自动升级固件 **笔记**:要使用自定义固件二进制文件执行固件升级,可以通过运行 --fw-image-dir 提供包含固件二进制文件的文件夹的路径。使用此选项,MLNX_OFED 包中嵌入的固件版本将被忽略。

例子:

/mlnxofedinstall --fw-image-dir /tmp/my_fw_bin_files

如果驱动程序检测到系统上不支持的卡,它将中止安装过程。为避免这种情况, 请确保添加--跳过不支持的设备检查 安装期间标记。

用法

/mnt/mlnxofedinstall [选项]

安装脚本将删除所有之前安装的 OFED 软件包并从头开始重新安装。系统将提示您确认删除 旧软件包。

预先存在的配置文件将以".conf.rpmsave"扩展名保存。

- o 如果您需要在整个(同构)集群上安装 OFED,一种常见的策略是将 ISO 映像挂载到其中一个集群节点上,然后将其复制到共享文件系统(如 NFS)。要在所有集群节点上安装,请使用集群感知工具(如 aspdsh)。
- o 如果您的内核版本与任何提供的预构建 RPM 不匹配,您可以使用 MLNX_OFED 包内的 "mlnx_add_kernel_support.sh"脚本添加您的 内核版本。

在安装了勘误版内核的 Redhat 和 SLES 发行版中,无需使用 mlnx_add_kernel_support.sh 脚本。可以执行常规安装,弱更新机制将创 建指向 MLNX_OFED 内核模块的符号链接。

如果为自定义内核重新生成内核模块(使用--添加内核支持)软件包安装不会自动重新生成 initramfs。在某些情况下,例如通过 ConnectX 卡安装根文件系统的系统,不重新生成 initramfs 甚至可能导致系统无法重启。

在这种情况下,安装程序将建议运行以下命令来更新 initramfs:

dracut-f

在某些操作系统上,dracut-f可能会导致以下错误消息,可以安全忽略。

libkmod: kmod_module_new_from_path: kmod_module 'mdev'已存在于不同的路径中

"mlnx_add_kernel_support.sh"脚本可以直接从 mlnxofedinstall 脚本执行。有关更多信息,请参阅下面的"--add-kernel-support"选项。

在 Ubuntu 和 Debian 发行版中,驱动程序安装使用动态内核模块支持 (DKMS) 框架。因此,驱动程序的编译将在 MLNX_OFED 安装期间在主机上进行。

因此,使用"mlnx_add_kernel_support.sh"与 Ubuntu 和 Debian 发行版无关。

例子: 以下命令将在 /tmp 目录下为 RedHat 7.3 创建 MLNX_OFED_LINUX ISO 映像。

./MLNX_OFED_LINUX-xx-x-rhel7。3-x86_64/mlnx_add_kernel_support.sh -m /tmp/ MLNX_OFED_LINUX-xx-x-rhel7。3-x86_64/——make-tgz

注意:该程序将创建 MLNX_OFED_LINUX TGZ**为了**rhel7。3/tmp 下 目录。

ll Mellanox、OEM、OFED 或 Distribution IB 软件包将被删除。o 您想要继续?[是/否]:是

ee 日志文件 /tmp/mlnx_ofed_iso.ee21642。日志

正在构建 OFED RPM。请稍候… 正在删除 OFED

RPM..

已创建 /tmp/MLNX_OFED_LINUX-xx-x-rhel7。3-x86_64-ext.tgz

- o 该脚本为 MPI 等用户空间组件将以下行添加到 /etc/security/limits.conf 中:
 - * 软 memlock 无限制
 - * 硬内存锁不受限制

• 这些设置将用户空间应用程序可以固定的内存量设置为无限制。如果需要,可以将值无限制调整为特定的 RAM 量。

为了使您的机器成为 InfiniBand/VPI 结构的一部分,子网管理器必须在其中一个结构节点上运行。此时,OFED for Linux 您的机器上已安装了 OpenSM 子网管理器。

要获取安装选项列表,请运行:

/mlnxofedinstall--h

安装程序

本节介绍 NVIDIA 适配卡上 MLNX_OFED 的安装过程。

- a. 以 root 身份登录安装机器。
- b. 在您的机器上安装 ISO 映像。

ost1# mount -o ro,loop MLNX_OFED_LINUX-<版本>-<操作系统标签>-<CPU 架构>.iso /mnt

c. 运行安装脚本。

mnt/mlnxofedinstall

ogs 目录: /tmp/MLNX_OFED_LINUX-xx-x.logs

该程序将安装 MLNX_OFED_LINUX包裹在您的机器上。请注意,所有其他 Mellanox、OEM、OFED、RDMA 或 Distribution IB 包将

被移除。

由于与 MLNX_OFED_LINUX 冲突,软管包被删除,<mark>做</mark>不是 重新安装它们。

开始 MLNX_OFED_LINUX-xxx 安装....

•••••

安装成功完成。尝试执行固件更新...查询 Mellanox 设备

固件...

对于无人值守安装,请在运行 MLNX_OFED 安装脚本时使用 --force 安装选项:

/mnt/mlnxofedinstall——force

Ubuntu 的 MLNX_OFED 应在 chroot 环境中使用以下标志进行安装:

./mlnxofedinstall --without-dkms --add-kernel-support --kernel <chroot 中的内核版本> --without-fw-update --force

例如:

./mlnxofedinstall--不带-dkms--添加内核支持--内核 3.13.0-85-通用--不 带-fw-update--force

请注意,如果源不在其默认位置,则应添加内核源 (--kernel-sources) 的路 径。 如果您的机器具有最新的固件,则不会发生固件更新,并且安装脚本将在安装结束时打印类似以下内容的消息:

设备#1:

设备类型: ConnectX-X 部件编号: MCXXXX-XXX PSID: MT <version>

PCI 设备名称: 0b:00.0 基本 MAC: 0000e41d2d5cf810 版本: 当前可用 FW XX.XX.XXXX

状态: 最新

如果您的机器有不受支持的网络适配器设备,则不会进行固件更新,并且会 打印以下错误消息之一。请联系您的硬件供应商以获取固件更新帮助。错误 消息 #1:

设备#1:

.

设备类型: ConnectX-X 部件编号: MCXXXX-XXX PSID: MT_<version>

PCI 设备名称: 0b:00.0 基本 MAC: 0000e41d2d5cf810 版

本: 当前可用

交货日期 XX.XX.XXXX

状态:未找到匹配的图像错误消

息 #2:

该设备的固件未在 NVIDIA 驱动程序内分发: 0000: 01: 00.0

(PSID: IBM2150110033)

要获取此设备的固件,请联系您的硬件供应商。

d.**案例 A**:如果安装脚本已对您的网络适配器执行了固件更新,则您需要重新启动驱动程序或重新启动系统,然后固件更新才能生效。请参阅下表以找到适合 您的特定卡的操作。

动作∖适配器	驱动程序重启	标准重启 (软复位)	冷重启(硬 重置)
标准 ConnectX- 4/ConnectX-4 Lx 或 更高	-	+	-
适配器与多 主办方支持	-	-	+
插座直接卡	-	-	+

案例 B:如果安装脚本尚未对您的网络适配器执行固件升级,请运行以下命令重新启动驱动程序:"/etc/init.d/openibd 重新启动"。

- e. (仅限 InfiniBand)运行 hca_self_test.ofed 实用程序以验证 InfiniBand 链路是否已启动。该实用程序还会检查并显示其他信息,例如:
- o HCA 固件版本
- o 内核架构
- o 驱动程序版本
- o 活动 HCA 端口数量及其状态

o 节点 GUID 有关 hca_self_test.ofed 的更多详细信息,请参阅文件 docs/readme_and_user_manual/hca_self_test.readme。

安装完成后,可以通过以下方式检索有关 OFED 安装的信息,例如前缀、内核版本和安装参数 运行命令 /etc/infiniband/info。大多数 OFED 组件可以在安装后通过修改相关配置文件进行配置或重新配置。有关详细信息,请参阅本手册中的相关章节。

可以在 /etc/infiniband/openib.conf 文件中找到启动时自动加载的模块列表。

安装 OFED 将替换 RDMA 堆栈并删除现有的第三方 RDMA 连接器。

安装结果

软件	• 大多数 MLNX_OFED 软件包都安装在"/usr"目录下,除了以下软件包安装在"/opt"目录下:		
	o fca和 ibutils		
	o iproute2(rdma 工具)- 安装在 /opt/ Mellanox/iproute2/sbin/rdma 下		
	• 内核模块安装在		
	o SLES 和 Fedora 发行版上的 /lib/modules/ uname -r /ùpdates		
	o RHEL 和其他类似 RedHat 发行版上的 /lib/ modules/`uname -r /extra/mlnx-ofa_kernel		
	o Ubuntu 上的 /lib/modules/`uname -r /ùpdates/ dkms/		

固件

- 如果满足以下两个条件,现有网络适配器设备的固件将被更新:
 - o 安装脚本以默认模式运行;即不带选项 "--without-fw-update"
 - o 适配器设备的固件版本比 OFED ISO 映像中包含的固件版本旧

笔记: 如果适配器的闪存最初是用扩展 ROM 映像进行编程的,则自动固件更新也将刻录扩展 ROM 映像。

· 如果您的机器具有不受支持的网络适配器设备,则不 会进行固件更新,并且会打印以下错误消息。

"此设备的固件未在 NVIDIA 驱动程序中分发: 0000: 01: 00.0 (PSID: IBM2150110033) 要获取此设备的固件,请联系您的硬件供应商。"

安装日志

在安装 MLNX_OFED 时,每个选定包的安装日志将保存在单独的日志文件中。运行安装脚本后将显示包含日志文件的目录路径,格式如下:

例子:

ogs 目录: /tmp/MLNX_OFED_LINUX-4.4-1.0。0.0.IBMM2150110033.日志

系统启动时加载驱动程序

系统启动时,NVIDIA 驱动程序将自动加载。

> 为了防止系统启动时自动加载 NVIDIA 驱动程序:

a. 将以下行添加到 "/etc/modprobe.d/mlnx.conf" 文件。

缺乏列表 mlx5_core 缺乏列表 mlx5_ib

- b. 在 "/etc/infiniband/openib.conf" 文件中设置"ONBOOT=no"。
- c. 如果模块存在于 initramfs 文件中,则内核可以自动加载它们。要防止此行为,请使用操作系统的标准工具更新 initramfs。**笔记**:更新 initramfs 的过程 会添加步骤 1 中的黑名单,并阻止内核自动加载模块。

mlnxofedinstall 返回代码

下表列出了 mlnxofedinstall 脚本返回代码及其含义。

返回 代码	意义
0	安装成功结束
1	安装失败
2	未找到适配器设备的固件
22	参数无效
二十八	可用空间不足
171	不适用于此系统配置。当系统上没有所需硬件时,可能会发生这种情况

返回 代码	意义
172	未满足先决条件。例如,缺少安装所需的软件或硬件配置不正确
173	无法启动 mst 驱动程序

软件	• 大多数 MLNX_OFED 软件包都安装在"/usr"目录下,除了以下软件包安装在"/opt"目录下:
	o fca和ibutils
	o iproute2(rdma 工具)- 安装在 /opt/ Mellanox/iproute2/sbin/rdma 下
	• 内核模块安装在
	o SLES 和 Fedora 发行版上的 /lib/modules/ uname -r /ùpdates
	o RHEL 和其他类似 RedHat 发行版上的 /lib/ modules/`uname -r /extra/mlnx-ofa_kernel
	o Ubuntu 上的 /lib/modules/`uname -r /ùpdates/ dkms/
o 固件	如果满足以下两个条件,现有网络适配器设备的固件将被更新:
	o 安装脚本以默认模式运行;即不带选项" without-fw-update"

o 适配器设备的固件版本比 OFED ISO 映像中包含的固件版本旧

笔记: 如果适配器的闪存最初是用扩展 ROM 映像进行编程的,则自动固件更新也将刻录扩展 ROM 映像。

· 如果您的机器具有不受支持的网络适配器设备,则不 会进行固件更新,并且会打印以下错误消息。

"此设备的固件未在 NVIDIA 驱动程序中分发: 0000: 01: 00.0 (PSID: IBM2150110033) 要获取 此设备的固件,请联系您的硬件供应商。"

安装日志

在安装 MLNX_OFED 时,每个选定包的安装日志将保存在单独的日志文件中。 运行安装脚本后将显示包含日志文件的目录路径,格式如下:

例子:

ogs 目录: /tmp/MLNX_OFED_LINUX-4.4-1.0。0.0.IBMM2150110033.日志

卸载 MLNX_OFED

使用脚本 /usr/sbin/ofed_uninstall.sh 卸载 MLNX_OFED 包。该脚本是 ofed-scripts RPM 的一部分。

附加安装程序

使用 YUM 安装 MLNX_OFED

该安装类型适用于RedHat/OL和Fedora操作系统。

设置 MLNX_OFED YUM 存储库

- a. 以 root 身份登录安装计算机。
- b. 在您的机器上安装 ISO 映像并将其内容复制到网络中的共享位置。

mount -o ro,loop MLNX_OFED_LINUX-<版本>-<操作系统标签>-<CPU 架构>.iso /mnt

c.下载并安装NVIDIA的GPG-KEY:

可以通过以下链接下载密钥: http://www.mellanox.com/downloads/ ofed/RPM-GPG-KEY-Mellanox

wget http://www.mellanox.com/downloads/ofed/RPM-GPG-KEY-Mellanox - 2018-01-25 13:52: 三十-- http://www.mellanox.com/downloads/ofed/RPM-GPG-KEY-Mellanox

解决 www.mellanox.com...72.3。194.0

连接到www.mellanox.com|72.3。194.0|:80... 已连接。TTP 请求已发送,正在等待响应...200好的,结束:1354(1.3K)[text/plain] 保存至: ? RPM-GPG-KEY-Mellanox?

00%[=======]], 354 --.-千克/秒 在0秒内

018-01-25 13: 52: 三十(247MB/s)-? RPM-GPG-KEY-Mellanox? 已保存[1354/1354]

d. 安装密钥。

sudo rpm -- 进口RPM-GPG-KEY-Mellanox

警告:rpmts_HdrFromFdno: Header V3 DSA/SHA1 签名,密钥 ID 6224c050:

诺基

从文件中检索密钥: ///repos/MLNX_OFED/<MLNX_OFED 文件>/RPM-GPG-KEY-

Mellanox

导入 GPG 密钥0x6224C050:

用户身份: Mellanox 技术 (Mellanox 技术 - 签名密钥 v2)

< support@mellanox.com >"

来自:/repos/MLNX_OFED/<MLNX_OFED文件>/RPM-GPG-KEY-Mellanox s这确定[y/N]:

e.检查密钥是否已成功导入。

rpm -q gpg-pubkey --qf'%{名称}-%{版本}-%{发布}\t%{概要}\n'| grep

Mellanox

pg-公钥-a9e4b643-520791ba gpg(Mellanox 技术公司

<支持@mellanox.com>)

f. 创建一个名为 "/etc/yum.repos.d/mlnx_ofed.repo"的 yum 存储库配置文件,其内容如下:

mlnx_ofed]

ame=MLNX_OFED 存储库

aseurl=文件:///<提取的 MLNX_OFED 包的路径>/RPMS 已禁用=1

pgkey=文件: ///<下载密钥 RPM-GPG-KEY-Mellanox 的路径> 检查pg=1

g.检查存储库是否已成功添加。

yum 重新排序

加载插件: product-id、security、subscription-manager

该系统未注册到 Red Hat 订阅管理。您可以使用订阅管理器进行注册。epo id repo 名称

地位

lnx_ofed MLNX_OFED 存储库

108

普林福吉 RHEL 6Server - RPMforge.net - dag

4, 597

撰稿人: 8,351

使用 --add-kernel-support 设置 MLNX_OFED YUM 存储库

- a. 以 root 身份登录安装计算机。
- b. 在您的机器上安装 ISO 映像并将其内容复制到网络中的共享位置。

mount -o ro,loop MLNX_OFED_LINUX-<版本>-<操作系统标签>-<CPU 架构>.iso /mnt

c. 构建包含内核支持的软件包并创建 tarball。

注意:该程序将创建 MLNX_OFED_LINUX TGZ**为了**rhel7。6/tmp 下 目录。

你想继续?[是/否]:是

ee 日志文件 /tmp/mlnx_iso.4120_logs/mlnx_ofed_iso。4120。日志

赫金**如果**所有需要的软件包都已安装...正在构建 MLNX_OFED_LINUX RPMS。请

稍候...创建 metadata-rpms**为了**3.10。0-957.21。3.el7.x86_64 ...

警告:如果您要配置**这个包**作为存储库,然后

请注意

警告:它包含未签名的 RPM,因此您需要禁用

gpgcheck

警告:通过设置'gpgcheck=0'在存储库 conf 文件中。创建 /tmp/MLNX_OFED_LINUX-5.2-

0.5。5.0-rhel7。6-x86_64-ext.tgz

d. 打开 tarball。

cd /tmp/ tar -xvf /tmp/MLNX_OFED_LINUX-5.2-0.5。5.0-rhel7。6-x86_64-ext.tgz

e. 创建一个名为 "/etc/yum.repos.d/mlnx_ofed.repo"的 YUM 存储库配置文件,其内容如下:

mlnx_ofed]
ame=MLNX_OFED 存储库
aseurl=文件:///<提取的 MLNX_OFED 包的路径>/RPMS 已禁用=1
检查pg=0

f.检查存储库是否已成功添加。

yum 重新排序

加载插件: product-id、security、subscription-manager

该系统未注册到 Red Hat 订阅管理。您可以使用

订阅管理器进行注册。epoid repo名称

地位

lnx_ofed MLNX_OFED 存储库

108 4, 597

普林福吉 RHEL 6Server - RPMforge.net - dag

撰稿人: 8,351

使用 YUM 工具安装 MLNX_OFED

为 MLNX_OFED 包设置 YUM 存储库后,执行以下操作:

a. 通过调用以下命令查看可用的软件包组:

```
yum 搜索 mlnx-ofed-
lnx-ofed-all.noarch: MLNX OFED 所有安装程序包裹(带有 KMP 支持)lnx-ofed-all-user-only.noarch
: MLNX OFED 所有用户专用安装程序包裹
    (仅限用户空间包)
Inx-ofed-basic.noarch: MLNX_OFED 基本安装程序包裹(带有 KMP 支持)lnx-ofed-basic-user-only.noarch
: MLNX_OFED basic-user-only 安装程序包裹
    (仅限用户空间包)
Inx-ofed-bluefield.noarch: MLNX OFED bluefield 安装程序包裹(使用 KMP
  支持)
Inx-ofed-bluefield-user-only.noarch: MLNX OFED bluefield-user-only 安装程序
  包裹 (仅限用户空间包)
Inx-ofed-dpdk.noarch: MLNX OFED dpdk 安装程序包裹(带有 KMP 支持)lnx-ofed-dpdk-upstream-
libs.noarch: MLNX OFED dpdk-upstream-libs 安装程序
  包裹(带有 KMP 支持)lnx-ofed-dpdk-upstream-libs-user-only.noarch: MLNX_OFED dpdk-upstream-
  libs-
  仅限用户安装程序包裹(仅限用户空间包)lnx-ofed-dpdk-user-only.noarch: MLNX_OFED dpdk-user-
only 安装程序包裹
   (仅限用户空间包)
Inx-ofed-eth-only-user-only.noarch: MLNX_OFED eth-only-user-only 安装程序
   包裹 (仅限用户空间包)
Inx-ofed-guest.noarch: MLNX_OFED 客户机安装程序包裹(带有 KMP 支持)Inx-ofed-guest-user-
only.noarch: MLNX_OFED 仅限来宾用户的安装程序包裹
   (仅限用户空间包)
lnx-ofed-hpc.noarch: MLNX OFED hpc 安装程序包裹(带有 KMP 支持)lnx-ofed-hpc-user-
only.noarch: MLNX_OFED hpc-user-only 安装程序包裹
    (仅限用户空间包)
Inx-ofed-hypervisor.noarch: MLNX OFED 虚拟机管理程序安装程序包裹(和
  KMP 支持)
Inx-ofed-hypervisor-user-only.noarch: MLNX_OFED 虚拟机管理程序用户专用
  安装程序包裹(仅限用户空间包)lnx-ofed-kernel-only.noarch: MLNX_OFED 仅内核安装程序包裹(和
  KMP 支持)
lnx-ofed-vma.noarch: MLNX OFED vma 安装程序包裹(带有 KMP 支持)lnx-ofed-vma-eth.noarch:
MLNX OFED vma-eth 安装程序包裹(使用 KMP
  支持)
Inx-ofed-vma-eth-user-only.noarch: MLNX OFED vma-eth-user-only 安装程序
  包裹 (仅限用户空间包)
Inx-ofed-vma-user-only.noarch: MLNX_OFED vma-user-only 安装程序包裹
    (仅限用户空间包)
Inx-ofed-vma-vpi.noarch: MLNX_OFED vma-vpi 安装程序包裹(使用 KMP
Inx-ofed-vma-vpi-user-only.noarch: MLNX_OFED vma-vpi-user-only 安装程序
  包裹(仅限用户空间包
```

在哪里:

mlnx-ofed-全部	安装 MLNX_OFED 中所有可用的包
mlnx-ofed-基本版	安装运行 NVIDIA 卡所需的基本软件包
mlnx-ofed-客人	安装客户操作系统所需的软件包
mlnx-ofed-hpc	安装 HPC 所需的包
mlnx-ofed-虚拟机管理程序	安装虚拟机管理程序操作系统所需的软件包
mlnx-ofed-vma	安装 VMA 所需的软件包
mlnx-ofed-vma-eth	安装 VMA 通过以太网工作所需的软件包
mlnx-ofed-vma-vpi	安装 VMA 所需的软件包以支持 VPI
布卢菲尔德	安装 BlueField 所需的软件包
开发工具包	安装 DPDK 所需的软件包
dpdk-上游库	使用 RDMA-Core 安装 DPDK 所需的软件包
仅限内核	安装非默认内核所需的软件包

笔记: MLNX_OFED 为 RHEL 和 SLES 提供带有 KMP 支持的内核模块 RPM 包。对于其他操作系统,内核模块 RPM 软件包仅针对操作系统的默认内核提供。在这种情况下,RPM 软件包组的名称中包含支持的内核版本。

例子:

```
Inx-ofed-全部-3.17。4-301.fc21.x86_64.noarch: MLNX_OFED 所有安装程序包裹
   为了核心3.17。4-301.fc21.x86 64(不支持 KMP)lnx-ofed-basic-3.17。4-301.fc21.x86 64.noarch
: MLNX OFED 基本安装程序
   包裹核心3.17。4-301.fc21.x86 64(不支持 KMP)Inx-ofed-guest-3.17。4-301
.fc21.x86_64.noarch: MLNX_OFED 来宾安装程序
   包裹核心3.17。4-301.fc21.x86_64(不支持 KMP)lnx-ofed-hpc-3.17。4-301.fc21.x86_64.noarch:
MLNX OFED HPC 安装程序包裹
   为了核心3.17。4-301.fc21.x86_64(不支持 KMP)lnx-ofed-hypervisor-3.17。4-301
.fc21.x86_64.noarch: MLNX_OFED 虚拟机管理程序
   安装程序包裹核心3.17。4-301.fc21.x86 64(不支持 KMP)lnx-ofed-vma-3.17。4-301
.fc21.x86 64.noarch: MLNX OFED vma 安装程序包裹
   为了核心3.17。4-301.fc21.x86 64(不支持 KMP)lnx-ofed-vma-eth-3.17。4-301.fc21.x86 64.noarch:
MLNX OFED vma-eth 安装程序
   包裹核心3.17。4-301.fc21.x86_64(不支持 KMP)lnx-ofed-vma-vpi-3.17。4-301.fc21.x86_64.noarch:
MLNX_OFED vma-vpi 安装程序
   包裹核心3.17。4-301.fc21.x86_64(不支持 KMP)lnx-ofed-hypervisor-3.17。4-301
.fc21.x86_64.noarch: MLNX_OFED 虚拟机管理程序
   安装程序包裹核心3.17。4-301.fc21.x86 64(不支持 KMP) lnx-ofed-vma-3.17。4-301
.fc21.x86_64.noarch: MLNX_OFED vma 安装程序包裹
   为了核心3.17。4-301.fc21.x86_64(不支持 KMP)lnx-ofed-vma-eth-3.17。4-301.fc21.x86_64.noarch:
MLNX OFED vma-eth 安装程序
   包裹核心3.17。4-301.fc21.x86_64(不支持 KMP)lnx-ofed-vma-vpi-3.17。4-301.fc21.x86_64.noarch:
MLNX_OFED vma-vpi 安装程序
   包裹核心3.17。4-301.fc21.x86 64(不支持 KMP)
```

当使用不同于 RHEL 或 SLES 的操作系统时,或者您安装了 MLNX OFED 中默认不支持的内核时,

您可以使用 mlnx_add_kernel_support.sh 脚本为您的内核构建 MLNX_OFED。

该脚本将自动为您的内核构建匹配的组 RPM 包,以便您仍然可以通过 yum 安装 MLNX_OFED。请注意,生成的 MLNX_OFED 存储库将包含未签名的 RPM,因此,您应该在存储库配置文件中设置"gpgcheck=0"。

b. 安装所需的组。

使用 "YUM"工具安装 MLNX_OFED 不会自动更新固件。

要将固件更新为 MLNX_OFED 包中包含的版本,请运行: # yum 安装 mlnx-fw-updater

使用 apt-get 安装 MLNX_OFED

此类安装适用于Debian和Ubuntu操作系统。

设置 MLNX_OFED apt-get 存储库

a. 以 root 身份登录安装计算机。

b. 在网络中的共享位置提取 MLNX_OFED 包。

可以从以下网址下载https://www.nvidia.com/en-us/networking/ →产品→软件→InfiniBand 驱动程序。

c. 创建一个名为 "/etc/apt/sources.list.d/mlnx_ofed.list"的 apt-get 存储库配置文件,其内容如下:

eb 文件: /<提取的 MLNX_OFED 的路径**包裹**>/德布斯./

d. 下载并安装 NVIDIA's Technologies GPG-KEY。

wget -qO-http: //www.mellanox.com/downloads/ofed/RPM-GPG-KEY-Mellanox | sudo apt-key 添加-

e. 验证密钥是否已成功导入。

apt-key 列表

乌巴托1024D/A9E4B6432013-08-11

ID Mellanox Technologies <支持@mellanox.com> 1024g/

乌巴托09FCC2692013-08-11

f. 更新 apt-get 缓存。

sudo apt-get 更新

使用 --add-kernel-support 设置 MLNX_OFED apt-get 存储库

- a. 以 root 身份登录安装计算机。
- b. 在您的机器上安装 ISO 映像并将其内容复制到网络中的共享位置。

mount -o ro,loop MLNX_OFED_LINUX-<版本>-<操作系统标签>-<CPU 架构>.iso /mnt

c. 构建包含内核支持的软件包并创建 tarball。

| 65

/mnt/mlnx_add_kernel_support.sh --make-tgz <可选 --kmp> -k \$(uname -r) -米/mnt/

注意:该程序将创建 MLNX_OFED_LINUX TGZ**为了**rhel7。6/tmp 下 目录。

你想**继续**?[是/否]:是

ee 日志文件 /tmp/mlnx_iso.4120_logs/mlnx_ofed_iso。4120。日志

赫金如果所有需要的软件包都已安装…正在构建 MLNX_OFED_LINUX RPMS。请稍候…创建 metadata-rpms为了3.10。0-957.21。3.el7.x86_64 ...

警告:如果您要配置**这个包**作为存储库,然后 请注意

警告:它包含未签名的 RPM,因此您需要禁用

gpgcheck

警告:通过设置'gpgcheck=0'在存储库 conf 文件中。创建 /tmp/MLNX_OFED_LINUX-5.2-

0.5。5.0-rhel7。6-x86_64-ext.tgz

d. 打开 tarball。

cd /tmp/ tar -xvf /tmp/MLNX_OFED_LINUX-5.2-0.5。5.0-rhel7。6-x86_64-ext.tgz

e. 创建一个名为 "/etc/apt/sources.list.d/mlnx_ofed.list"的 apt-get 存储库配置文件,其内容如下:

eb [trusted=yes] file:/<提取的 MLNX_OFED 的路径包裹>/德布斯./

f. 更新 apt-get 缓存。

sudo apt-get 更新

使用 apt-get 工具安装 MLNX_OFED

为 MLNX_OFED 包设置 apt-get 存储库后,执行以下操作:

a. 通过调用以下命令查看可用的软件包组:

```
apt-cache 搜索 mlnx-ofedpt-cache 搜索 mlnx-
 ofed ......
nem-dkms——DKMS 支持为了mlnx-ofed 内核模块 lnx-ofed-kernel-dkms - DKMS 支持为了mlnx-ofed 内核模
块 lnx-ofed-kernel-utils - 用于重启和调整 mlnx-ofed 内核的用户空间工具
Inx-ofed-vma-vpi - MLNX_OFED vma-vpi 安装程序包裹(带有 DKMS 支持)lnx-ofed-kernel-only -
MLNX OFED 仅内核安装程序包裹(使用 DKMS
Inx-ofed-bluefield - MLNX OFED bluefield 安装程序包裹(使用 DKMS
Inx-ofed-hpc-user-only - MLNX OFED hpc-user-only 安装程序包裹(用户
   仅限太空套餐)
Inx-ofed-dpdk-user-only - MLNX_OFED dpdk-user-only 安装程序包裹(用户
   仅限太空套餐)
Inx-ofed-all-exact - MLNX_OFED 所有安装程序包裹(具有 DKMS 支持)
lnx-ofed-all - MLNX_OFED 所有安装程序包裹(带有 DKMS 支持)lnx-ofed-vma-vpi-user-only -
MLNX_OFED vma-vpi-user-only 安装程序包裹
    (仅限用户空间包)
Inx-ofed-eth-only-user-only - MLNX_OFED eth-only-user-only 安装程序包裹
    (仅限用户空间包)
Inx-ofed-vma-user-only - MLNX_OFED vma-user-only 安装程序包裹(用户
   仅限太空套餐)
lnx-ofed-hpc - MLNX_OFED hpc 安装程序包裹(带有 DKMS 支持)lnx-ofed-bluefield-user-only -
MLNX OFED bluefield-user-only 安装程序
   包裹 (仅限用户空间包)
Inx-ofed-dpdk - MLNX_OFED dpdk 安装程序包裹(带有 DKMS 支持)lnx-ofed-vma-eth-user-only -
MLNX_OFED vma-eth-user-only 安装程序包裹
    (仅限用户空间包)
Inx-ofed-all-user-only - MLNX OFED 所有用户专用安装程序包裹(用户
   仅限太空套餐)
lnx-ofed-vma-eth - MLNX OFED vma-eth 安装程序包裹(具有 DKMS 支持)lnx-ofed-vma - MLNX OFED
vma 安装程序包裹(带有 DKMS 支持)lnx-ofed-dpdk-upstream-libs-user-only - MLNX_OFED dpdk-
upstream-libs-user-
   仅安装程序包裹(仅限用户空间包)Inx-ofed-basic-user-only - MLNX OFED basic-user-only 安装程序包裹
   仅限太空套餐)
Inx-ofed-basic-exact - MLNX_OFED 基本安装程序包裹(具有 DKMS 支持)
lnx-ofed-basic - MLNX_OFED 基本安装程序包裹(带有 DKMS 支持)lnx-ofed-dpdk-upstream-libs -
MLNX_OFED dpdk-upstream-libs 安装程序包裹
    (有 DKMS 支持)
```

在哪里:

mlnx-ofed-全部	MLNX_OFED 所有安装程序包
mlnx-ofed-基本版	MLNX_OFED 基本安装程序包
mlnx-ofed-vma	MLNX_OFED vma 安装程序包
mlnx-ofed-hpc	MLNX_OFED HPC 安装程序包
mlnx-ofed-vma-eth	MLNX_OFED vma-eth 安装程序包
mlnx-ofed-vma-vpi	MLNX_OFED vma-vpi 安装程序包
克内姆	MLNX_OFED DKMS 支持 mlnx-ofed 内核模块
内核-dkms	MLNX_OFED kernel-dkms 安装程序包
仅限内核	MLNX_OFED 仅内核安装程序包
布卢菲尔德	MLNX_OFED bluefield 安装程序包
mlnx-ofed-全部精确	MLNX_OFED mlnx-ofed-all-exact 安装程序包
开发工具包	MLNX_OFED dpdk 安装程序包
mlnx-ofed-基本-精确	MLNX_OFED mlnx-ofed-basic-exact 安装程序包
dpdk-上游库	MLNX_OFED dpdk-upstream-libs 安装程序包

b. 安装所需的组。

pt-get 安装'<组名>'

例子:

pt-get 安装 mlnx-ofed-all

使用"apt-get"工具安装 MLNX_OFED 不会自动更新固件。

要将固件更新为 MLNX_OFED 包中包含的版本,请运行: 安装 mlnx-fw-updater

性能调优

根据用户系统的应用,可能需要根据以下情况修改网络适配器的默认配置: ConnectX® 适配器。如果需要进行调整,请参阅NVIDIA 网络适配器性能调优指南。

5.3 VMware 驱动安装

本节介绍 VMware 驱动程序安装。

5.3.1 硬件和软件要求

要求	描述
平台	基于NVIDIA设备的适配卡的服务器平台:
	ConnectX®-7(InfiniBand/以太网)(固件: fw-ConnectX7)
操作系统	ESXi 8.x
安装人员权限	安装需要目标机器的管理员权限。

5.3.2 安装适用于 VMware vSphere 的 NATIVE ESXi 驱动程序

在安装新版本之前,请卸载所有以前的驱动程序包。

要安装驱动程序:

- 1.使用 root 权限登录 ESXi 服务器。
- 2.安装驱动程序。

```
> esxcli 软件 vib 安装 -d <路径>/<bundle_file>
```

例子:

> esxcli 软件 vib 安装 -d /tmp/MLNX-NATIVE-ESX-ConnectX-4-5_4。16.8。8-10EM-650.0。 0.4240417.zipesxcli

- 3.重新启动机器。
- 4. 验证驱动程序是否安装成功。

sxcli 软件 vib 列表 | grep nmlx mlx5-core

mlx5-rdma 4.16。8.8-10EM。650.0。0.4240417

4.16。8.8-1OEM。650.0。0.4240417 合作伙伴支持2017-01-31 合作伙伴支持

2017-01-31

安装过程完成后,所有内核模块将在启动时自动加载。

5.3.3 删除早期版本的 NVIDIA 驱动程序

删除之前请先卸载以前安装的驱动程序。

要删除所有驱动程序:

- 1.使用 root 权限登录 ESXi 服务器。
- 2.列出所有现有的 NATIVE ESXi 驱动程序模块。(请参阅为 VMware vSphere 安装 NATIVE ESXi 驱动程序。)
- 3. 移除各个模块:

> esxcli 软件 vib 删除-n nmlx5-rdma > esxcli 软件 vib 删除-n nmlx5-core

要删除模块,您必须按照与上例相同的顺序运行命令。

4.重新启动服务器。

5.3.4 固件编程

1. 从以下位置下载 VMware 可启动二进制映像 v4.6.0固件工具 (MFT) 站点。

a.ESXi 6.5 文件: mft-4.6.0.48-10EM-650.0.0.4598673.x86_64.vib

b. MD5SUM: 0804cffe30913a7b4017445a0f0adbe1

2. 按照*MFT 用户手册*。

以下过程需要从 USB 设备下载、安装和启动自定义启动映像。

6 更新适配器固件

每张适配卡在制造时都附带了最新版本的合格固件。但是,NVIDIA 偶尔会发布固件更新,提供新功能和错误修复。要检查您的卡是否使用最新的可用固件版本进行编程,请下载 mlxup 固件更新和查询实用程序。该实用程序可以查询可用的 Mellanox 适配器并指示哪些适配器需要固件更新。如果用户确认,mlxup 将使用嵌入的图像升级固件。最新的 mlxup 可执行文件和文档可在mlxup——更新和查询实用程序。

固件更新示例

[服务器1]# ./mlxup

查询 Mellanox 设备固件...设备类型:

ConnectX-7

零件编号: MCX753436MC-HEAB

描述: NVIDIA ConnectX-7 OCP3.0 SFF 适配卡、200GbE(默认模式)/NDR200 IB、双端口 QSFP112、多主机和

支持 Socket Direct、PCIe 5.0 x16、加密启用、安全启动启用、翼形螺钉(拉环)支架 PSID:

MT_2190110032

PCI 设备名称: 0000: 06: 00.0 基本

GUID: e41d2d0300fd8b8a

版本: 当前的 可用的

固件 16.00.0000 16.00.0000

地位: 最新

设备类型: ConnectX-7 零件编号: MCX753436MC-HEAB

描述: NVIDIA ConnectX-7 OCP3.0 SFF 适配卡、200GbE(默认模式)/NDR200 IB、双端口 OSFP112、多主机和

支持 Socket Direct、PCIe 5.0 x16、加密启用、安全启动启用、翼形螺钉(拉环)支架 PCI 设备名称: 0000: 07: 00.0

基本 MAC 地址: 0000e41d2da206d4

版本: 当前的 可用的

FW 16.20.1000 16.24.1000

地位: 需要更新

执行 FW 更新? [y/N]: y 设备 #1: 最新

设备 #2: 更新固件...完成

需要重新启动才能使更新生效。日志文件:/var/log/mlxup/

mlxup-yyyymmdd.log

7设置高速端口的链路类型

下表列出了 ConnectX-7 InfiniBand/以太网适配卡的默认网络端口配置。

开放网络	数据传输率	默认模式
MCX753436MC-HEAB	NDR200/HDR和 200GbE	200GbE
MCX753436MS-HEAB	NDR200/HDR和 200GbE	200GbE
MCX75343AMC-NEAC	NDR和 400GbE	400GbE
MCX75343AMS-NEAC	NDR和 400GbE	400GbE

要更改网络高速端口模式,您可以使用mlnx配置或者UEFI工具。

UEFI 可用于在操作系统启动前配置适配卡设备,而 mlnxconfig 可用于操作系统启动后配置。

根据您的喜好,使用以下工具之一:

7.1 mlnx配置

mlxconfig 工具允许用户更改某些设备配置,而无需重新烧录固件。重置后配置也会保留。默认情况下,mlxconfig 会显示下次启动时将加载的配置。有关更多信息和说明,请参阅mlnx配置。

7.2 UEFI

PreBoot 驱动程序首先初始化适配器设备,检测端口协议(以太网或 InfiniBand),然后启动端口。然后,它连接到 DHCP 服务器以获取其分配的 IP 地址和网络参数,并获取要从中启动的内核/操作系统的源位置。DHCP 服务器指示 PreBoot 驱动程序通过 TFTP 服务器、iSCSI 目标或其他服务访问内核/操作系统。有关更多信息和说明,请参阅<u>UEFI</u>。

8 故障排除

8.1 常规故障排除

服务器无法找到适配 器	 确保适配器放置正确 确保适配器插槽和适配器兼容将适配器安装在不同的 PCI Express 插槽中 使用适配器附带的驱动程序或下载最新的 确保你的主板具有最新的 BIOS 尝试重启服务器
适配器不再起作用	・ 将适配器重新插入其插槽或其他插槽(如有必要)・ 尝试使用另一根电缆・ 重新安装驱动程序,因为网络驱动程序文件可能已损坏或被删除・ 重启服务器
安装另一个适配器后适配 器停止工作	・ 尝试移除并重新安装所有适配器・ 检查电缆连接是否正确・ 确保你的主板具有最新的 BIOS
链接指示灯不亮	・ 尝试交换机上的另一个端口・ 确保电缆已牢固连接・ 检查您使用的电缆是否正确,且长度不超过建议长度・ 验证您的交换机和适配器端口是否兼容

链接灯亮,但无通信	· 检查是否已加载最新驱动程序
已确立的	• 检查适配器及其链路是否设置为相同的速度和双工设置
收到電量不足事件信息	· 当 [适配器的当前功耗] > [PCIe 插槽公布的功率限制] 时,服务器的系统甚至日志中都会出现一条警告消息(例如 dmesg:"检测到 PCIe 电源不足")
	· 建议使用能够提供足够电源的 PCIe 插槽。
	· 如果出现以下格式的消息 - "mlx5_core 0003:01:00.0:port_module:254:(pid 0):端口模块事件 [错误]:模块 0,电缆错误,由于 PCIe 插槽上的功率不足/未通告,一个或多个网络端口已断电"请升级适配器的固件。
	・ 如果消息仍然存在 - 请考虑从有源光缆(AOC)或收发器切换到直接连接铜(DAC)连接。

8.2 Linux 故障排除

	猫/等/问题 uname -a cat /proc/cupinfo grep '型号名称' uniq ofed_info -s 配置文件 ip 链接显示 ethtool <接口> ethtool -i <mellanox 接口端口号=""> ibdev2netdev</mellanox>
卡片检测	lspci grep -i Mellanox

Mellanox 固件工具 (MFT)	下载并安装MFT: MFT 文档 请参阅用户手册了解安装说明。安装后,运行: 启动 mst 状态 flint -d <mst_device> q</mst_device>
端口信息	统计工具 ibv_devinfo
固件版本升级	要下载最新固件版本,请参阅NVIDIA 更新和查询实用程序。
收集日志文件	猫/ var /日志/消息 dmesg >> 系统日志 journalctl(适用于新操作系统)cat /var/log/ syslog

8.3 Windows 故障排除

环境信息	从 Windows 桌面选择"开始"菜单并运行:msinfo32 要将系统信息导出到文本文件,请从文件菜单中选择导出选项。指定文件名并保存。
Mellanox 固件工具 (MFT)	下载并安装MFT: MFT 文档 请参阅用户手册了解安装说明。安装后,打开 CMD 窗口并运行: WinMFT
	启动 mst 状态 flint -d <mst_device> q</mst_device>

端口信息	状态	
固件版本升级	使用 PSID/板 ID 下载最新固件版本 <u>这里</u> .flint -d <mst_device> -i <固件bin文件> b</mst_device>	
收集日志文件	 事件日志查看器 MST 设备日志: a 启动 b mst 状态 flint -d <mst_device> dc > dump_configuration.log</mst_device> mstdump <mst_device> dc > mstdump.log</mst_device> 	

9 规格

请确保将 ConnectX-7 卡安装在能够提供所需电源和气流的 PCIe 插槽中,如下表所示。

在待机模式下,只有端口0可用。

9.1 MCX75343AMC-NEAC / MCX75343AMS-NEAC 规格

身体的	适配卡尺寸 : 高小型 (TSFF) OCP 3.0,4.52 英寸 x 2.99 英寸(115.00 毫米 x 76.00 毫米) 适配卡 高度: 0.7 英寸(17.8 毫米)				
	支架类型 :翼形螺钉(拉环)支架				
接口	看支持的接口				
	PCI Express Gen 5.0/4.0	SERDES @ 16/32GT	/s,x16 通道(兼容 Gen 3.0)		
	网络端口:单端口 OSFP InfiniBan	d 和以太网			
协议支持	数据速率	无限带宽	NDR/NDR200/HDR/HDR100/EDR/FDR/SD	R	
		以太网	400/200/100/50/25/10 Gb/秒		
			200(2 通道 x 每通道 100Gb/s)、HDR(4 通道 x 每通道 /s)、1X/2X/4X SDR(每通道 2.5Gb/s)	〔50Gb/s)、HDR100(2 通	
	以太网: 400GAUI-4 C2M,400GBASE-CR4、200GAUI-4 C2M、200GBASE-CR4、100GAUI-2 C2M、100GBASE-CR4、100GBASE-CR2、50GAUI-2 C2M、50GAUI-1 C2M、50GBASE-CR、40GBASE-CR4、25GBASE-R、10GBASE-R、10GBASE-CX4、1000BASE-CX、CAUI-4 C2M、25GAUI C2M、XLAUI C2M、XLPPI、SFI				
功能	MCX75343AMC-NEAC	启用安全启动、启用	用加密、支持多主机和套接字直连		
	MCX75343AMS-NEAC	启用安全启动、禁用	加密、支持多主机和套接字直连		
电气和热能	电压 : 12V_边缘,+3.3V_边缘				
规格	PCIe Gen 5.0 x16 中采用无源电缆的典型功率		MCX75343AMC-NEAC	25.9 瓦	
			MCX75343AMS-NEAC	24.4 瓦	
	通过 OSFP 端口可提供的最大功率: 每端口 17W(不支持散热)				
	完整的电气和热规格参见 " <i>NVIDIA ConnectX-7 电气和热规格</i> "该文件可在 NVOnline 登录后获取。				
环境的	温度	操作	0°C 至 55°C		
		无法运作	-40°C 至 70°Cь		
	湿度	操作	10% 至 85% 相对湿度		
		无法运作	10% 至 90% 相对湿度		
	海拔(运行)	3050米			
监管	安全: CB / cTUVus / CE				
	电磁兼容 (EMC): CE/FCC/VCCI/ICES/RCM/KC				
	有害物质限制指令: 符合 RoHS 规定				

笔记: a. ConnectX-7 适配器 补充了 IBTA 自动协商规范,以获得更好的误码率和更长的电缆传输距离。此补充功能仅在连接到另一个 NVIDIA InfiniBand 产品时启动。

b. 非操作存储温度规格适用于不带包装的产品。

9.2 MCX753436MS-HEAB / MCX753436MC-HEAB 规格

身体的	适配卡尺寸 :小型 (SFF), 4.52 英寸 x 2.99 英寸(115.00 毫米 x 76.00 毫米) 适配 卡高度: 0.59 英寸(15.1 毫米)			
	支架类型 :翼形螺钉(拉环)支	と架		
接口	看支持的接口			
	PCI Express Gen 5.0/4.0	SERDES @ 16/32	GT/s,x16 通道(兼容 4.0 和 3.0)	
	网络端口 : 双端口 QSFP112 In	finiBand 和以太网		
协议支持	数据速率	无限带宽	NDR200/HDR/HDR100/EDR/FDR/	SDR
		以太网	200/100/50/25/10 Gb/秒	
	EDR(每通道 25Gb/s)、FDR(每通	通道 14.0625Gb/s)、1X/2	<u>'</u>	·
	以太网: 200GAUI-4 C2M、200GAUI-4 C2C、200GBASE-CR4、200GBASE-KR4、200GBASE-CR4、200GBASE-KR4、100GBASE-CR4、100GBASE-KR4、100GBASE-SR4、100GBASE-CR2、100GBASE-KR2、100GBASE-SR2、50GBASE-R2、50GBASE-R4、40GBASE-CR4、40GBASE-KR4、40GBASE-LR4、40GBASE-ER4、40GBASE-R2、25GBASE-R、20GBASE-KR2、10GBASE-LR、10GBASE-ER、			
	10GBASE-CX4、10GBASE-CR、10GE MCX753436MC-HEAB		BE-CA、1000BASE-KA、10GBASE-SK B用加密、支持多主机和套接字直连	
-20 00	MCX753436MS-HEAB		等用加密、支持多主机和套接字直连	
电气和	电压: 12V_边缘,+3.3V_边缘	,		
热的	通过 QSFP112 端口提供的最大功率:每个端口 17W(不支持散热)			
规格	PCIe Gen 5.0 x16 中采用无源电缆的典型功率		MCX753436MC-HEAB	25.9 瓦
			MCX753436MS-HEAB	24.5 瓦
	完整的电气和热规格参见"NVIDIA ConnectX-7 电气和热规格"该文件可在 NVOnline 登录后获取。			
环境的	温度	操作	0°C 至 55°C	
		无法运作	-40°C 至 70°Cь	

湿度	湿度	操作	10% 至 85% 相对湿度	
		无法运作	10% 至 90% 相对湿度	
	海拔(运行)	3050 米		
监管	安全: CB / cTUVus / CE			
	电磁兼容 (EMC): CE/FCC/VCCI/ICES/RCM/KC			
	有害物质限制指令: 符合 RoHS 规定			

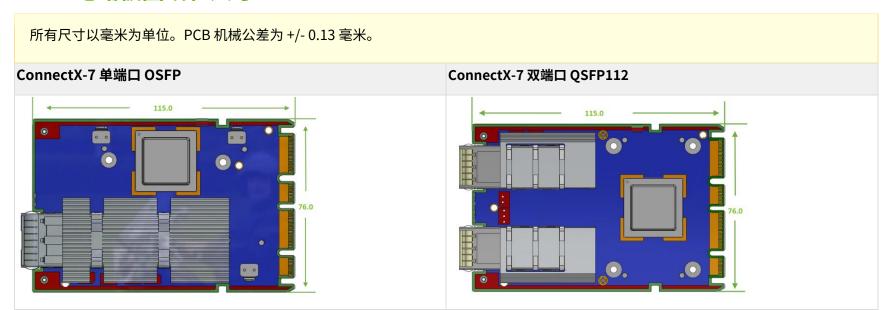
笔记: a. ConnectX-7 适配器 补充了 IBTA 自动协商规范,以获得更好的误码率和更长的电缆传输距离。此补充功能仅在连接到另一个 NVIDIA InfiniBand 产品时启动。

b. 非操作存储温度规格适用于不带包装的产品。

9.3 机械图纸和尺寸

有关该卡的 3D 模型,请参阅 PID。

9.3.1 电路板图纸和尺寸



9.3.2 支架图纸及尺寸

所有尺寸以毫米为单位。支架的机械公差为 +/- 0.25 毫米。
单端口 OSFP 翼形螺钉支架
82.50
17.80
17.80

10 监控

10.1热传感器

该适配卡采用 ConnectX IC, 其工作温度范围为 0°C 至 105°C。

三个热阈值定义影响整个系统的运行状态:

- 警告-105℃: 仅适用于受管系统: 当设备超过 105℃ 阈值时,管理软件会发出警告阈值消息,向系统管理部门表明卡已超过警告阈值。请注意,此温度阈值不需要也不会导致硬件采取任何操作(例如适配器卡关闭)。
- · 批判的-115°C: 当设备超过此温度时,固件会自动关闭设备。
- 紧急情况-130°C: 如果固件在超过临界阈值时无法关闭设备,则设备在超过紧急(130°C)阈值时会自动关闭。

可以通过系统的 SMBus 读取卡的热传感器。用户可以读取这些热传感器,并根据读数和上述 IC 热传感器的需求调整系统气流

10.2散热器

ConnectX-7 IC 上装有散热器以散热。ConnectX-7 IC 具有热关断安全机制,在发生高温事件、热耦合不当或散热器拆除时,可自动关闭 ConnectX-7 卡。

有关每个 OPN 所需的气流 (LFM),请参阅规格。

下表列出了每个 OPN 的散热器特性。

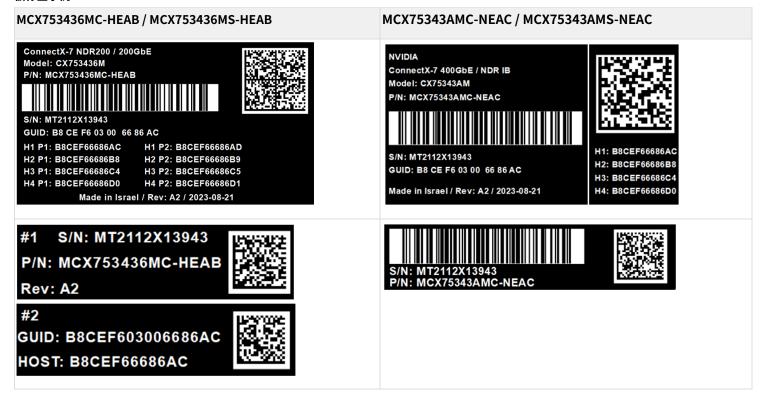
开放网络	散热器
MCX753436AS-HEAB 和 MCX753436MS-HEAB	使用两个弹簧图钉进行连接。
MCX75343AMC-NEAC 和 MCX75343AMS-NEAC	使用两个弹簧图钉和三个L形机械支架进行连接,如上面的机械图所示。
	这些 L 形机械支架将 PCB 连接至散热器以形成加强筋。

11 查找适配卡上的 MAC 地址和序列号

每个适配卡的标签上都印有不同的标识符:序列号、以太网协议的卡 MAC 和 InfiniBand 协议的卡 GUID。VPI 卡既有 GUID,又有 MAC(源自 GUID)。只有 IB 卡才有 InfiniBand 协议的 GUID。

下图中标签上标明的产品修订不一定代表卡的最新修订版本。

板标签示例





12 文档修订历史

日期	变更描述
2023年12月	更新监管标签查找适配卡上的 MAC 地址和序列号
2023年10月	・ 更新 <u>查找适配卡上的 MAC 地址和序列号</u> ・ 更新了支架尺寸 <u>规格</u>
2023年8月	更新查找适配卡上的 MAC 地址和序列号
2023年6月	更新规格 具有典型的功率数字
2023年5月	更新规格 包括非操作存储温度规范
2023年1月	更新了 400Gb/s 支持的以太网协议 <u>规格</u>
2022年12月	首次发布适用于所有 ConnectX-7 OCP 3.0 适配卡的综合用户手册

注意

本文档仅供参考,不应视为对产品的特定功能、状况或质量的保证。NVIDIA Corporation 及其任何直接或间接子公司和附属公司(统称"NVIDIA")均不对本文档中包含的信息的准确性或完整性做出任何明示或暗示的陈述或保证,并且不 对其中的任何错误承担任何责任。NVIDIA 对此类信息的后果或使用,或因使用此类信息而导致的任何专利或其他第三方权利侵权不承担任何责任。本文档不承诺开发、发布或交付任何材料(定义如下)、代码或功能。

NVIDIA 保留随时对本文档进行更正、修改、增强、改进和任何其他更改的权利,恕不另行通知。

客户应在下订单之前获取最新的相关信息,并应验证此类信息是最新且完整的。

NVIDIA 产品的销售须遵守订单确认时提供的 NVIDIA 标准销售条款和条件,除非 NVIDIA 授权代表与客户签署的单独销售协议("销售条款")另有约定。NVIDIA 在此明确反对将任何客户一般条款和条件应用于本文件中提及的NVIDIA 产品购买。本文件不直接或间接构成任何合同义务。

NVIDIA 产品的设计、授权或保证不适用于医疗、军事、飞机、太空或生命支持设备,也不适用于 NVIDIA 产品故障或失灵可能导致人身伤害、死亡或财产或环境损害的应用。NVIDIA 对在此类设备或应用中纳入和/或使用 NVIDIA 产品不承担任何责任,因此此类纳入和/或使用由客户自行承担风险。

NVIDIA 不保证基于本文档的产品适合任何特定用途。NVIDIA 不一定会测试每种产品的所有参数。客户应自行负责评估和确定本文档中包含的任何信息的适用性,确保产品适合客户计划的应用程序,并对应用程序进行必要的测试,以避免应用程序或产品出现故障。客户产品设计中的缺陷可能会影响 NVIDIA 产品的质量和可靠性,并可能导致本文档中未包含的额外或不同的条件和/或要求。NVIDIA 不承担与任何违约、损害、成本或问题相关的任何责任,这些违约、损害、成本或问题可能基于或归因于:(i) 以任何违反本文档的方式使用 NVIDIA 产品或 (ii) 客户产品设计。

本文件项下的任何 NVIDIA 专利权、版权或其他 NVIDIA 知识产权均未明示或暗示授予任何许可。NVIDIA 发布的有关第三方产品或服务的信息并不构成 NVIDIA 使用此类产品或服务的许可或此类产品或服务的保证或认可。使用此类信息可能需要根据第三方的专利或其他知识产权获得第三方许可,或根据 NVIDIA 的专利或其他知识产权获得 NVIDIA 许可。

仅当事先获得 NVIDIA 书面批准、未经修改地复制且完全遵守所有适用的出口法律和法规并附带所有相关条件、限制和声明时,才允许复制本文档中的信息。

本文档以及所有 NVIDIA 设计规范、参考板、文件、图纸、诊断、列表和其他文档(统称或单独称为"材料")均按"原样"提供。NVIDIA 不就材料作出任何明示、暗示、法定或其他形式的保证,并明确否认所有关于非侵权、适销性和适用于特定用途的暗示保证。在法律允许的范围内,NVIDIA 在任何情况下均不对因使用本文档而造成的任何损害负责,包括但不限于任何直接、间接、特殊、偶发、惩罚性或后果性损害,无论该等损害是如何造成的,也无论责任理论如何,即使 NVIDIA 已被告知存在此类损害的可能性。无论客户因何种原因可能遭受任何损害,NVIDIA 对客户就本文所述产品承担的累计责任应根据产品销售条款进行限制。

商标

NVIDIA、NVIDIA 徽标和 Mellanox 是 NVIDIA Corporation 和/或 Mellanox Technologies Ltd. 在美国和其他国家/地区的商标和/或注册商标。其他公司和产品名称可能是其所属各自公司的商标。

版权

© 2023 NVIDIA Corporation 及其附属公司。保留所有权利。