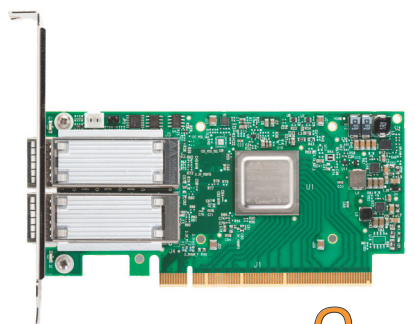


ConnectX[®]-5 EN

100Gb/s 以太网适配器卡



ConnectX[®]-5

支持 RDMA 的智能网络适配器卡，具有针对高性能计算、Web2.0、云和存储平台的高级应用卸载功能

ConnectX-5 EN 支持具有 100Gb 以太网连接的两个端口，同时可提供小于 600 纳秒延迟和超高消息速率以及 PCIe 交换机和 NVMe over Fabric 卸载。ConnectX-5 可为要求最苛刻的以下应用和市场提供最高性能和最灵活的解决方案：机器学习、数据分析等。

HPC 环境

ConnectX-5 为高性能、数据密集型和可扩展的计算与存储平台提供高带宽、低延迟和高计算效率。ConnectX-5 通过提供 MPI、SHMEM/PGAS 和集合标记匹配卸载、针对无序 RDMA 写入和读取操作的硬件支持以及额外的网络原子和 PCIe 原子操作支持，为 HPC 基础架构提供增强功能。

ConnectX-5 EN 利用 RoCE（基于融合以太网的 RDMA）技术，提供低延迟和高性能。

ConnectX-5 通过完成交换机自适应路由功能和支持无序传输的数据，同时保持有序的完成语义，为所有网络拓扑（包括 DragonFly 和 DragonFly+）提供多路径可靠性和高效支持，从而增强了 RDMA 网络功能。

ConnectX-5 还支持针对后台检查点的突发缓冲区卸载，而不会干扰主 CPU 操作，并支持动态连接传输 (DCT) 这一创新型传输服务，以确保计算和存储系统获得极致的可扩展性。

存储环境

NVMe 存储设备越来越受欢迎，提供非常快速的存储访问。不断发展的 NVMe over Fabric (NVMe over Fabric) 协议利用 RDMA 连接进行远程访问。通过提供 NVMe over Fabric 目标卸载功能，ConnectX-5 实现了进一步的增强，可以在无 CPU 干预的情况下进行非常高效的 NVMe 存储访问，从而提高性能和降低延迟。

此外，嵌入式 PCIe 交换机使客户能够构建独立存储或机器学习设备。与前几代 ConnectX 适配器一样，标准块和文件访问协议可利用 RoCE 进行高性能存储访问。与多架构网络相比，将计算和存储进行整合的网络可实现显著的性价比优势。

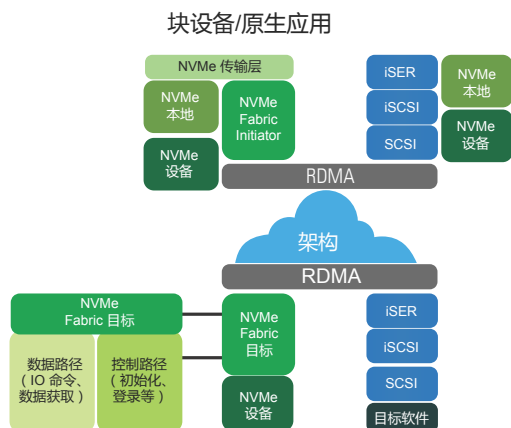
亮点

新功能特性

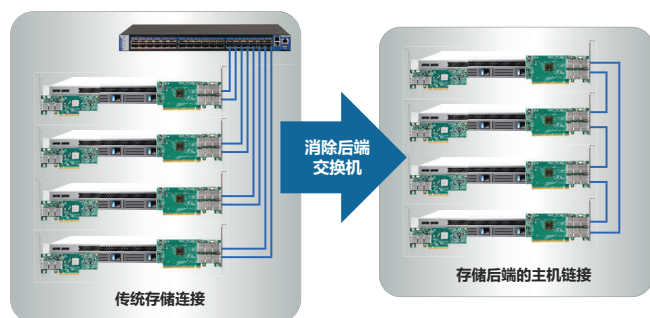
- 标记匹配和集合卸载
- 可靠传输的自适应路由
- 针对后台检查点的突发缓冲区卸载
- NVMe over Fabric (NVMe over Fabric) 卸载
- 通过主机链接消除后端交换机
- 嵌入式 PCIe 交换机
- 增强虚拟交换机/虚拟路由器卸载
- 灵活管线
- 叠加 (Overlay) 网络的 RoCE
- PCIe Gen 4 支持

优势

- 每端口最高 100 Gb/s 连接速度
- 行业领先的吞吐量、低延迟、低 CPU 利用率和高消息速率
- 利用多主机技术最大限度地提高数据中心的 ROI
- 基于主机链接技术的创新型存储和机器学习机架设计
- 基于 x86、Power、ARM 和 GPU 的计算和存储平台的智能互连
- 包括 NVMe over Fabric 卸载的高级存储功能
- 支持灵活管线可编程能力的智能网络适配器
- 虚拟化网络中的尖端性能，包括网络功能虚拟化 (NFV)
- 促成高效的服务链能力
- 高效的 I/O 整合，降低数据中心的成本和复杂性



ConnectX-5 支持创新的存储机架设计——主机链接，通过它不同的服务器可以直接互连，而不涉及架顶式 (ToR) 交换机。或者，可以使用最初由 ConnectX-4 引入的多主机技术。Mellanox Multi-Host™ 技术通过将 PCIe 接口分隔为多个独立的接口，从而将多个主机连接到单个适配器。借助多种新的机架设计替代方案，ConnectX-5 通过减少资本开支（线缆、网卡和交换机端口费用）和运营开支（削减交换机端口管理和整体功耗成本），降低了数据中心的总体拥有成本 (TCO)。



云和 Web2.0 环境

云和 Web2.0 客户在软件定义网络 (SDN) 环境中开发其平台，利用其服务器操作系统的虚拟交换功能来实现最大的灵活性。

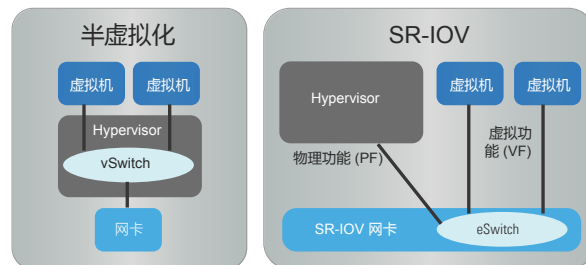
开放式虚拟交换机 (OVS) 是虚拟交换机的一个示例，允许虚拟机相互之间以及与外部世界进行通信。虚拟交换机传统上驻留在 Hypervisor 中，而交换是基于流上的十二元组匹配。基于虚拟交换机或虚拟路由器软件的解决方案是 CPU 密集型，会影响系统性能，并妨碍充分利用可用带宽。

利用 Mellanox 加速交换和数据包处理 (ASAP²) Direct 技术，可以通过在网卡硬件中处理数据层面来卸载虚拟交换机/虚拟路由器，同时保持控制层面不被修改。因此，显著提高了虚拟交换机/虚拟路由器的性能，而不会产生相关的 CPU 负载。

ConnectX-5 支持的虚拟交换机/虚拟路由器卸载功能包括叠加 (Overlay) 网络（例如 VXLAN、NVGRE、MPLS、GENEVE 和 NSH）标头封装和解封，以及内部数据包的无状态卸载及支持 NAT 功能的数据包头重写等。

此外，可对 ConnectX-5 智能灵活管线功能（包括灵活解析器和灵活的匹配操作表）进行编程，从而实现针对未来协议的硬件卸载。

ConnectX-5 SR-IOV 技术为服务器中的虚拟机 (VM) 提供专用适配器资源以及有保证的隔离和保护。此外，利用 ConnectX-5 网络功能虚拟化 (NFV)，可以将虚拟机用作虚拟设备。利用完整数据路径操作卸载以及 Hairpin 硬件功能和服务链，虚拟设备能够以极低的 CPU 使用率处理数据。



利用这些功能，数据中心管理员可以受益于更出色的服务器利用率，同时降低成本、功耗和线缆复杂性，从而在相同硬件上承载更多虚拟设备、虚拟机和租户。

兼容性

PCI Express 接口

- PCIe Gen 4
- PCIe Gen 3.0，兼容 1.1 和 2.0
- 2.5、5.0、8、16GT/s 链路速率
- 自动协商为 x16、x8、x4、x2 或 x1 通道
- PCIe 原子
- TLP（事务层数据包）处理提示 (TPH)
- 嵌入式 PCIe 交换机：最多 8 个分支

- PCIe 交换机下行端口控制 (DPC) 启用以实现 PCIe 热插拔
- 用于对等安全通信的访问控制服务 (ACS)
- 高级错误报告 (AER)
- 进程地址空间 ID (PASID) 地址转换服务 (ATS)
- IBM CAPI v2 支持（一致的加速器处理器接口）
- 支持 MSI/MSI-X 机制

操作系统/发行版*

- RHEL/CentOS
- Windows
- FreeBSD
- VMware
- OpenFabrics 企业发行版 (OFED)
- OpenFabrics Windows 发行版 (WinOF-2)

连接

- 与以太网交换机（最高 100GbE）的互操作性
- 具有 ESD 保护的无源铜缆
- 用于光纤和有源线缆支持的供电连接器

功能特性

以太网

- 100GbE/50GbE/40GbE/25GbE/10GbE / 1GbE
- IEEE 802.3bj、802.3bm 100 Gb 以太网
- IEEE 802.3by、以太网联盟 25、50 Gb 以太网，支持所有 FEC 模式
- IEEE 802.3ba 40 Gb 以太网
- IEEE 802.3ae 10 Gb 以太网
- IEEE 802.3az 高效以太网
- 基于 IEEE 802.3ap 的自动协商和 KR 启动
- IEEE 802.3ad、802.1AX 链路聚合
- IEEE 802.1Q、802.1P VLAN 标记和优先级
- IEEE 802.1Qau (QCN) – 拥塞通知
- IEEE 802.1Qaz (ETS)
- IEEE 802.1Qbb (PFC)
- IEEE 802.1Qbg
- IEEE 1588v2
- 巨型帧支持 (9.6KB)

增强的功能特性

- 基于硬件的可靠传输
- 集合操作卸载
- 矢量集合操作卸载
- PeerDirect™ RDMA (亦称 GPUDirect®) 通信加速

- 64/66 编码
- 扩展可靠连接传输 (XRC)
- 动态连接传输 (DCT)
- 增强的原子操作
- 高级内存映射支持，允许用户模式内存注册和重新映射 (UMR)
- 按需分页 (ODP)
- MPI 标记匹配
- 集合协议卸载
- 支持自适应路由的无序 RDMA
- 突发缓冲区卸载
- 网络内内存免注册 RDMA 内存访问

CPU 卸载

- 基于融合以太网的 RDMA (RoCE)
- TCP/UDP/IP 无状态卸载
- LSO、LRO 校验和卸载
- RSS (也在封装的数据包上)、TSS、HDS、VLAN 和 MPLS 标记插入/剥离、接收流转向
- 面向内核旁路应用程序的数据层面开发工具包 (DPDK)
- 使用 ASAP² 的开放虚拟交换机 (OVS) 卸载
 - 灵活的匹配动作流表
 - 隧道封装/解封
- 智能中断合并

- 支持 NAT 路由器硬件卸载的包头重写

存储卸载

- 针对目标机器的 NVMe over Fabric 卸载
- 纠删码卸载 – 卸载 Reed Solomon 计算
- T10 DIF – 针对进口和出口流量的线速签名移交操作
- 存储协议：SRP、iSER、NFS RDMA、SMB Direct、NVMe

叠加 (Overlay) 网络

- 基于叠加 (Overlay) 网络的 RoCE
- 用于叠加 (Overlay) 网络隧道协议的无状态卸载
- VXLAN、NVGRE 和 GENEVE 叠加 (Overlay) 网络的封装和解封装硬件卸载

基于硬件的 I/O 虚拟化

- 单根 IOV
- 地址转换和保护
- VMware NetQueue 支持
- SR-IOV：最多 1K 虚拟功能
- SR-IOV：每主机最多 16 个物理功能
- 虚拟化层次结构 (例如 NPAR 和多主机，启用时)
 - 虚拟化物理端口上的物理功能
 - 每个物理功能上的 SR-IOV

- 用户可编程且可配置的 QoS
- 保证虚拟机的 QoS

HPC 软件库

- Open MPI、IBM PE、OSU MPI (MVAPICH/2)、Intel MPI
- Platform MPI、UPC、Open SHMEM

管理和控制

- NC-SI over MCTP over SMBus 和 NC-SI over MCTP over PCIe
- 基板管理控制器接口
- 用于监控和控制 DSP0248 的 PLDM
- 用于固件更新 DSP0267 的 PLDM
- 用于管理 eSwitch 的 SDN 管理接口
- 用于设备控制和配置的 I²C 接口
- 通用 I/O 引脚
- SPI 闪存接口
- JTAG IEEE 1149.1 和 IEEE 1149.6

远程启动

- 通过以太网远程启动
- 通过 iSCSI 远程启动
- 统一可扩展固件接口 (UEFI)
- 预执行环境 (PXE)

* 此部分介绍硬件特性和功能。有关提供的功能特性，请参阅驱动程序和固件版本说明。

表 1 - 部件号和说明

OPN	说明	不含支架的尺寸
MCX512A-ACAT	ConnectX [®] -5 EN 网络接口卡，25GbE 双端口 SFP28，PCIe3.0 x8，高支架，ROHS R6	14.2cm x 6.9cm (薄型)
MCX515A-GCAT	ConnectX [®] -5 EN 网络接口卡，50GbE 单端口 QSFP28，PCIe3.0 x16，高支架，ROHS R6	
MCX516A-GCAT	ConnectX [®] -5 EN 网络接口卡，50GbE 双端口 QSFP28，PCIe3.0 x16，高支架，ROHS R6	
MCX515A-CCAT	ConnectX [®] -5 EN 网络接口卡，100GbE 单端口 QSFP28，PCIe3.0 x16，高支架，ROHS R6	
MCX516A-CCAT	ConnectX [®] -5 EN 网络接口卡，100GbE 双端口 QSFP28，PCIe3.0 x16，高支架，ROHS R6	
MCX516A-BDAT	ConnectX [®] -5 Ex EN 网络接口卡，40GbE 双端口 QSFP28，PCIe 4.0 x16，高支架，ROHS R6	
MCX516A-CDAT	ConnectX [®] -5 Ex EN 网络接口卡，100GbE 双端口 QSFP28，PCIe4.0 x16，高支架，ROHS R6	

注：所有高支架适配器在发运时都安装了高支架，并提供一个短支架作为附件。



北京市朝阳区望京东园七区保利国际广场 T1 15 层
电话：010-5789 2000
www.mellanox.com

