



ConnectX[®]-4 Lx EN Adapter Card for Open Compute Project (OCP)



Single-port 10/25/40/50 Gigabit Ethernet Adapter Cards supporting Multi-Host Technology, RDMA, Overlay Networks and more

ConnectX-4 Lx EN network controller with a 10/25/40/50Gb/s Ethernet interface delivers high-bandwidth, low latency and industry-leading Ethernet connectivity for Open Compute Project (OCP) server and storage applications in Web 2.0, Enterprise Data Centers and Cloud infrastructures.

With ConnectX-4 Lx EN, server and storage applications achieve significant throughput and latency improvements resulting in faster access, real-time response and more virtual machines hosted per server. ConnectX-4 Lx EN for Open Compute Project (OCP) improves network performance by increasing available bandwidth while decreasing the associated transport load on the CPU especially in virtualized server environments.

ConnectX-4 Lx EN adapter cards are available in various form factors to meet the needs of every data center including OCP Specifications 2.0 and 3.0.

Moreover, ConnectX-4 Lx EN introduces Multi-Host technology, which enables a new innovative rack design that achieves maximum CAPEX and OPEX savings without compromising on network performance.

MELLANOX MULTI-HOST™ TECHNOLOGY

Mellanox's ConnectX-4 Multi-Host provides high speed interconnect to up to four separate hosts from a single NIC with no performance degradation on any of them. Furthermore, ConnectX-4 Lx EN offers four fully-independent PCIe buses, lowering total cost of ownership in the data center by reducing CAPEX requirements from four cables, NICs, and switch ports to only one of each, and by reducing OPEX by cutting down on switch port management and overall power usage.

Each host can be active or inactive at any time, independent of the other hosts, and receives bandwidth of its own. Bandwidth is split between the hosts, either evenly (default) or based on configurable differentiated Quality of Service (QoS), depending on the data center's needs. Multi-Host technology features uncompromising independent host management, with full independent NC-SI/MCTP support to each host and to the NIC. IT managers can remotely control the configuration and power state of each host individually, such that management of one host does not affect host traffic performance or the management of the other hosts, guaranteeing host security and isolation. To further lower the total cost of ownership, ConnectX-4 Lx EN supports management of the multiple hosts using a single BMC, with independent NC-SI/MCTP management channels for each of the managed hosts.

Multi-Host also supports a heterogeneous data center architecture; the various hosts connected to the single adapter can be x86, Power, GPU, or Arm, thereby removing any limitations in passing data or communicating between CPUs.

HIGHLIGHTS

NEW FEATURES

- Multi-Host technology
- Connectivity to up to 4 independent hosts
- OCP Specification 2.0 and 3.0, as applicable
- 10/25/40/50Gb/s speeds
- Virtualization
- Accelerated Switch and Packet Processing (ASAP²)
- Low latency RDMA over Converged Ethernet
- Hardware offloads for NVGRE and VXLAN encapsulated traffic
- CPU offloading of transport operations
- Application offloading
- Mellanox PeerDirect™ communication acceleration
- End-to-end QoS and congestion control
- Hardware-based I/O virtualization
- Erasure Coding offload
- RoHS-R6

BENEFITS

- 10/25/40/50GbE connectivity for servers and storage
- Open Compute Project form factor
- Industry-leading throughput and low latency for Web access and storage performance
- Maximizing data centers' return on investment (ROI) with Multi-Host technology
- Smart interconnect for x86, Power, Arm, and GPU-based compute and storage platform
- Cutting-edge performance in virtualized Overlay Networks (VXLAN and NVGRE)
- Efficient I/O consolidation, lowering data center costs and complexity
- Virtualization acceleration
- Power efficiency

I/O VIRTUALIZATION

ConnectX-4 Lx EN SR-IOV technology provides dedicated adapter resources and guaranteed isolation and protection for virtual machines (VMs) within the server. I/O virtualization with ConnectX-4 Lx EN gives data center administrators better server utilization while reducing cost, power, and cable complexity, allowing more Virtual Machines and more tenants on the same hardware. ConnectX-4 Lx EN's SR-IOV capability and Multi-Host technology are mutually exclusive, and each host in a Multi-Host server can leverage an individual SR-IOV implementation.

OVERLAY NETWORKS

In order to better scale their networks, data center operators often create overlay networks that carry traffic from individual virtual machines over logical tunnels in encapsulated formats such as NVGRE and VXLAN. While this solves network scalability issues, it hides the TCP packet from the hardware offloading engines, placing higher loads on the host CPU. ConnectX-4 Lx EN effectively addresses this by providing advanced NVGRE and VXLAN hardware offloading engines that encapsulate and de-encapsulate the overlay protocol headers, and enable the traditional offloads to be performed on the encapsulated traffic for these and other tunneling protocols (GENEVE, MPLS, QinQ, and so on). With ConnectX-4 Lx EN, data center operators can achieve native performance in the new network architecture.

OVS OFFLOAD

ConnectX-4 Lx EN adds additional offloads to OpenVSwitch (OVS) functionalities, maximizing the CPU efficiency while maintaining high flexibility.

OVS allows Virtual Machines to communicate with each other and with the outside world. OVS traditionally resides in the hypervisor and switching is based on twelve tuple matching on flows. The OVS software-based solution is CPU intensive, affecting system performance and preventing full utilization of available bandwidth.

ASAP2™

Mellanox ConnectX-4 Lx EN offers Accelerated Switch and Packet Processing (ASAP²) technology to perform offload activities in the hypervisor, including data path, packet parsing, VxLAN and NVGRE encapsulation/decapsulation, and more.

ASAP² allows offloading by handling the data plane in the NIC hardware using SR-IOV, while maintaining the control plane used in today's software-based solutions unmodified. As a result, there is significantly higher performance without the associated CPU load. ASAP² has two formats: ASAP² Flex™ and ASAP² Direct™.

One example of a virtual switch that ASAP² can offload is OpenVSwitch (OVS).

RDMA OVER CONVERGED ETHERNET (ROCE)

ConnectX-4 Lx EN supports RoCE specifications delivering low-latency and high performance over Ethernet networks. Leveraging data center bridging (DCB) capabilities as well as ConnectX-4 Lx EN advanced congestion control hardware mechanisms, RoCE provides efficient low-latency RDMA services over Layer 2 and Layer 3 networks.

MELLANOX PEERDIRECT™

PeerDirect communication provides high efficiency RDMA access by eliminating unnecessary internal data copies between components on the PCIe bus (for example, from GPU to CPU), and therefore significantly reduces application run time. ConnectX-4 Lx EN advanced acceleration technology enables higher cluster efficiency and scalability to tens of thousands of nodes.

STORAGE ACCELERATION

Storage applications will see improved performance with the higher bandwidth ConnectX-4 Lx EN delivers. Moreover, standard block and file access protocols can leverage RoCE for high-performance storage access. A consolidated compute and storage network achieves significant cost-performance advantages over multi-fabric networks.

DISTRIBUTED RAID

ConnectX-4 Lx EN delivers advanced Erasure Coding offloading capability, enabling distributed Redundant Array of Inexpensive Disks (RAID), a data storage technology that combines multiple disk drive components into a logical unit for the purposes of data redundancy and performance improvement.

ConnectX-4 Lx EN's Reed-Solomon capability introduces redundant block calculations, which, together with RDMA, achieves high performance and reliable storage access.

STANDARD & MULTI-HOST MANAGEMENT

Mellanox's host management technology for standard and multi-host platforms optimizes board management and power, performance and Firmware update management via NC-SI, MCTP over SMBus and MCTP over PCIe, as well as PLDM for Monitor and Control DSP0248 and PLDM for Firmware Update DSP0267.

SOFTWARE SUPPORT

All Mellanox adapter cards are supported by Windows, Linux distributions, VMware, FreeBSD, and Citrix XENServer. ConnectX-4 Lx EN supports various management interfaces and has a rich set of tools for configuration and management across operating systems.

COMPATIBILITY

PCI Express Interface

- PCIe Gen 3.0 compliant, 1.1 and 2.0 compatible
- 2.5, 5.0, or 8.0GT/s link rate x8
- Auto-negotiates to x8, x4, x2, or x1
- Support for MSI/MSI-X mechanisms

Operating Systems/Distributions*

- RHEL/CentOS
- Windows
- FreeBSD
- VMware
- OpenFabrics Enterprise Distribution (OFED)
- OpenFabrics Windows Distribution (WinOF-2)

Connectivity

- Interoperable with 10/25/40/50Gb Ethernet switches
- Passive copper cable with ESD protection
- Powered connectors for optical and active cable support

FEATURES

Ethernet

- 50GbE / 40GbE / 25GbE / 10GbE
- 25G Ethernet Consortium 25, 50 Gigabit Ethernet
- IEEE 802.3ba 40 Gigabit Ethernet
- IEEE 802.3ae 10 Gigabit Ethernet
- IEEE 802.3az Energy Efficient Ethernet
- IEEE 802.3ap based auto-negotiation and KR startup
- Proprietary Ethernet protocols (20/40GBASE-R2, 50/56GBASE-R4)
- IEEE 802.3ad, 802.1AX Link Aggregation
- IEEE 802.1Q, 802.1P VLAN tags and priority
- IEEE 802.1Qau (QCN) – Congestion Notification
- IEEE 802.1Qaz (ETS)
- IEEE 802.1Qbb (PFC)
- IEEE 802.1Qbg
- IEEE 1588v2
- Jumbo frame support (9.6KB)

Multi-Host

- Up to 4 separate PCIe interfaces to 4 independent hosts
- Two PCIe x4 to two hosts or four PCIe x4 to four hosts or four PCIe x2 to four hosts
- Independent NC-SI SMBus interfaces
- Independent stand-by and wake-on-LAN signals

Enhanced Features

- Hardware-based reliable transport
- Collective operations offloads
- Vector collective operations offloads
- PeerDirect RDMA (aka GPUDirect®) communication acceleration
- 64/66 encoding
- Extended Reliable Connected transport (XRC)
- Dynamically Connected transport (DCT)
- Enhanced Atomic operations
- Advanced memory mapping support, allowing user mode registration and remapping of memory (UMR)
- On demand paging (ODP)
- Registration free RDMA memory access

Offloads

- RDMA over Converged Ethernet (RoCE)
- TCP/UDP/IP stateless offload
- LSO, LRO, checksum offload
- RSS (can be done on encapsulated packet), TSS, HDS, VLAN insertion/stripping, Receive flow steering
- Intelligent interrupt coalescence

Storage Offloads

- RAID offload - erasure coding (Reed-Solomon) offload

Overlay Networks

- Stateless offloads for overlay networks and tunneling protocols
- Hardware offload of encapsulation and decapsulation of NVGRE and VXLAN overlay networks

Hardware-Based I/O Virtualization

- Single Root IOV
- Multi-function per port
- Address translation and protection
- Multiple queues per virtual machine
- Enhanced QoS for vNICs
- VMware NetQueue support

Virtualization

- SR-IOV: Up to 256 Virtual Functions
- SR-IOV: Up to 16 Physical Functions per port
- Virtualization hierarchies (e.g. NPAR and Multi-Host)
 - Virtualizing Physical Functions on a physical port
 - SR-IOV on every Physical Function
- 1K ingress and egress QoS levels
- Guaranteed QoS for VMs

Protocol Support

- OpenMPI, IBM PE, OSU MPI (MVAPICH/2), Intel MPI
- Platform MPI, UPC, Open SHMEM
- TCP/UDP, MPLS, VxLAN, NVGRE, GENEVE
- iSER, NFS RDMA, SMB Direct
- uDAPL

Management and Control Interfaces

- NC-SI, MCTP over SMBus and MCTP over PCIe - Baseboard Management Controller interface
- PLDM for Monitor and Control DSP0248
- PLDM for Firmware Update DSP0267.
- SDN management interface for managing the eSwitch
- I²C interface for device control and configuration
- General Purpose I/O pins
- SPI interface to Flash
- JTAG IEEE 1149.1 and IEEE 1149.6

Remote Boot

- Remote boot over Ethernet
- Remote boot over iSCSI
- PXE and UEFI

* This section describes hardware features and capabilities. Please refer to the driver and firmware release notes for feature availability.

Table 1 - Part Numbers and Descriptions

OPN	Description	OCP Specification
MCX4421A-XCQN	ConnectX [®] -4 Lx EN network interface card for OCP with Host Management, 10GbE dual-port SFP28, PCIe3.0 x8, no bracket, ROHS R6	OCP Spec 2.0
MCX4411A-ACAN	ConnectX [®] -4 Lx EN network interface card for OCP, 25GbE single-port SFP28, PCIe3.0 x8, no bracket, ROHS R6	
MCX4421A-ACAN	ConnectX [®] -4 Lx EN network interface card for OCP, 25GbE dual-port SFP28, PCIe3.0 x8, no bracket, ROHS R6	
MCX4411A-ACQN	ConnectX [®] -4 Lx EN network interface card for OCP with Host Management, 25GbE single-port SFP28, PCIe3.0 x8, no bracket, ROHS R6	
MCX4421A-ACQN	ConnectX [®] -4 Lx EN network interface card for OCP with Host Management, 25GbE dual-port SFP28, PCIe3.0 x8, no bracket, ROHS R6	
MCX4431A-GCAN	ConnectX [®] -4 Lx EN network interface card for OCP with Host Management, 50GbE single-port QSFP28, PCIe3.0 x8, no bracket, ROHS R6	
MCX4431M-GCAN	ConnectX [®] -4 Lx EN network interface card for OCP with Multi-Host and Host Management, 50GbE single-port QSFP28, PCIe3.0 x8, no bracket, ROHS R6	OCP Spec 3.0
MCX4621A-ACAB	ConnectX [®] -4 Lx EN network interface card for OCP 3.0, with host management, 25GbE Dual-port SFP28, PCIe3.0 x8, Thumbscrew bracket, ROHS R6	

NOTE: All listed speeds are the maximum supported and include all lower supported speeds as well.



350 Oakmead Parkway, Suite 100, Sunnyvale, CA 94085
 Tel: 408-970-3400 • Fax: 408-970-3403
www.mellanox.com