Optimize your Cloud with the Right Network Adapters

Contents

Today's Clouds

The exponential rise in data volumes comes along with a growing need for on-demand computing that requires larger, high performing, and more efficient clouds, whether public or private.

The variety of cloud workloads and technologies are increasing the load on CPU utilization. Overlay networks protocols, OVS processing, massive storage access with a variety of new protocols, and other high performance workloads require intensive processing. This wastes expensive CPU cycles, clogs the path to the network and ultimately leaves a lot of bandwidth unutilized. The end result is that application efficiency is limited and the cloud as a whole becomes inefficient as well.

Because of these challenges, data center administrators now look to implement intelligent, flexible, networks that can provide enough bandwidth for the application and storage requirements, alleviate the CPU loads, and enable the cloud to scale efficiently. Intelligent networks can share the load by offloading as many networking tasks as possible, thereby freeing CPU resources to serve more users and process more data.


Network Adapters Value for the Cloud

Network adapters are the basis for an effective and high performing cloud. So the applications running on the top layers will be able to scale with minimum limitations.

Their advantages for the cloud can be aligned along three main dimensions:

- **High Performance:** the growth in demand for cloud services require increasing rates of bandwidth. Ranging from 10Gbps and up to 100Gbps from one port. Backward- and future-compatibility, ability to grow per need, and ability to incorporate in one data center various speed rates, are critical for adequately optimizing the cloud efficiency.

- **Network Intelligence:** a new trend in data center design today is sharing the processing workloads between various components. It is no longer a CPU-only task, but the modern network is required of advanced acceleration capabilities ranging from RDMA (for a variety of use cases and applications) to protocol-specific offloads, such as TCP, overlay networks, and storage offloads. Implementing them on the network layer will free expensive CPU cycles for user applications while providing an improved user experience.

- **Scale-Out Efficiency:** unlike in the past, the network layer today has a significant role in the ability of the cloud to scale efficiently. Networking acceleration can help to increase the number of VMs per server and other new innovative technologies such as Multi-Host can enable building the most cost-, space-, and power-efficient clouds.
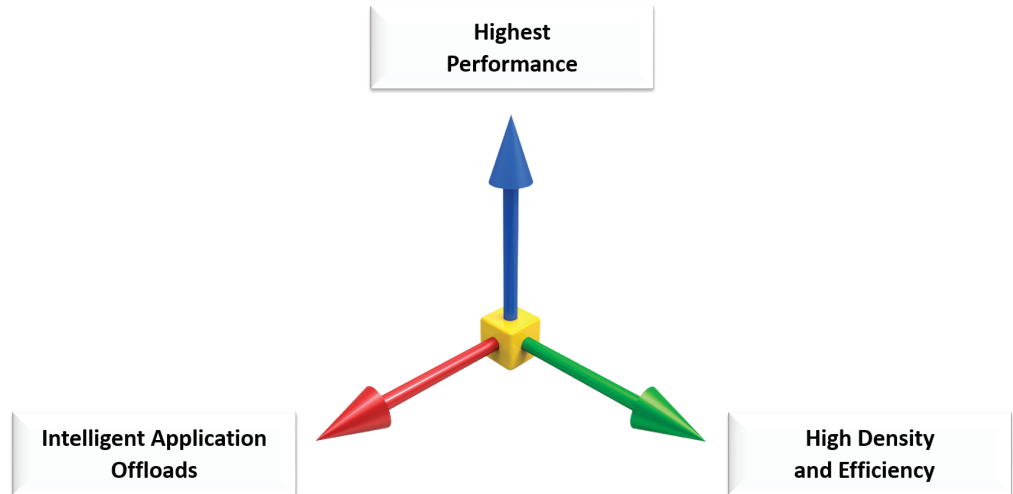
Highest Speed Rates and Flexibility

Placing more VMs on cloud compute servers, and having applications that server more users and access more data, require bandwidth as high as 100G and a flexible data center design including various speed rates.

It is no longer one type of connectivity, but clouds today require a range of solutions according to the size of the compute node and the storage servers.

Mellanox ConnectX-4 is the first 100G adapter in the market and also the first 25G and 50G adapter as well. Speed rates that are optimal for the new traffic intensive cloud and Web 2.0 data centers.

Future Proof Solution

The ConnectX-4 adapter family enables the cloud to grow as needed, from 10GbE to 25GbE, without replacing the hardware or the software.

Speed Rate / Adapter Family	ConnectX-4	ConnectX-4 Lx
10 GbE	Yes	Yes
25 GbE	Yes	Yes
40 GbE	Yes	Yes
50 GbE	Yes	Yes
100 GbE	Yes	-

## Increased VM Density by Offloading Networking Tasks to the NIC

An intelligent network will offload networking and storage tasks and in so doing, provides two main benefits: first, CPU cycles are free to serve more VMs, and second, fast networks offer more available bandwidth. Thus, more VMs can be served efficiently.

One example of this is in the implementation of overlay network protocols such as VXLAN, NVGRE, or GENEVE.

In a traditional implementation, as packets are encapsulated and their format changes, the adapter can no longer perform network offloads on the flowing packets, leaving this work to the CPU. This results in a massive number of CPU cycles being spent on networking tasks, instead of user workloads.

This eventually limits the overall bandwidth and the maximum number of VMs that can be placed on any one server.

Mellanox adapters can parse and understand the network overlay protocols (VXLAN, NVGRE, and GENEVE) and thus, able to offload their network processing. This enables more bandwidth, greatly improved VM density, and a significantly better optimized CPU utilization - which leads to massive cost benefits.
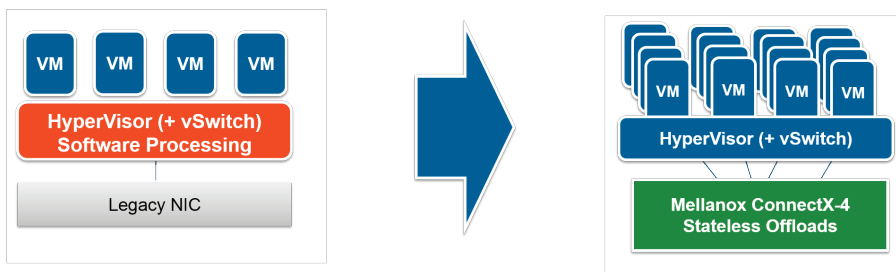
Traditional Model: Hypervisor Processing
Max Bandwidth: 17.6Gbps
CPU Utilization: 61%
VMs per Server: 17

ConnectX-4: Hardware Offload
Max Bandwidth: 36.5Gbps
CPU Utilization: 27%
VMs per Server: 36

Figure 1. 40GbE Adapter with VXLAN traffic example: almost double the VM density and bandwidth, while CPU utilization is reduced by 35%

Bare Metal-Like Virtual Machine (VM) Performance with SR-IOV

Hypervisor networking software takes a heavy toll on VM performance. Therefore, for high performance demanding applications, cloud administrators traditionally used dedicated bare metal servers as virtual machines were not considered to provide the right networking performance.

But, with the right networking technologies, such as SR-IOV and OVS Offload, it is now possible to run applications that require low latency and high bandwidth, such as financial and HPC workloads, from virtual machines.

Mellanox enables direct VM access to the network by enabling SR-IOV. Mellanox SR-IOV technology also provides RDMA access from the VM to other VMs or physical hosts on the network thus enabling bare metal like latency and performance from virtual machines.
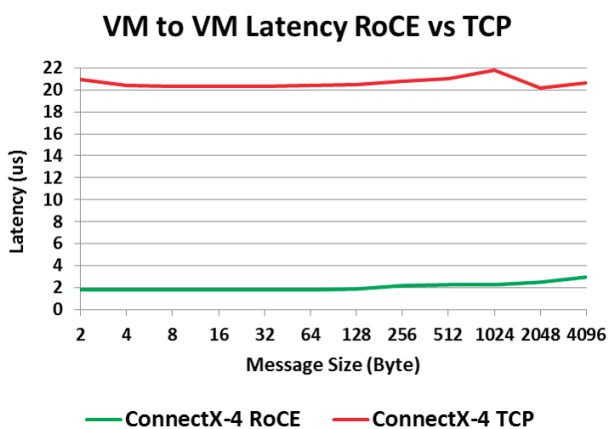

Figure 2. 10X latency advantage of RoCE over SR-IOV vs. TCP SR-IOV

High Storage Performance and Massive COGS Savings with RDMA

By enabling RDMA (Remote Direct Memory Access) and RoCE (RDMA over Converged Ethernet) connectivity on storage access, cloud users can now benefit from the lowest latency and highest bandwidth without wasting any additional load on the CPU.

This achieves line rate performance while saving on compute and storage costs.

Some of the major storage protocols and platforms are RDMA-enabled today. Examples include iSER (iSCSI over RoCE), SMB Direct, iSER within OpenStack's Cinder, Ceph RDMA, and other.

iSER and SMBDirect performance advantages below:

High Performance Storage iSER

Using iSER (iSCSI Extensions for RDMA) for storage communication provides significant improvement in latency, IoPS, bandwidth, and CPU utilization.

The advantages of using Mellanox adapters with iSER instead of regular iSCSI can be seen in the Figure 3.

iSER is now embedded in multiple storage solutions and software packages, such as OpenStack Cinder.
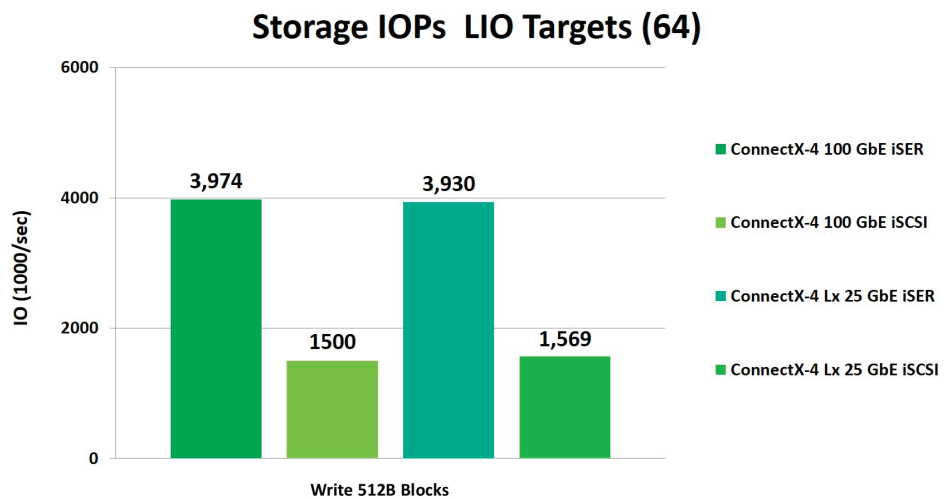
Storage IOPs LIO Targets (64)

Figure 3. *Mellanox iSER performance vs. regular iSCSI*

RDMA with SMB Direct

Another example of RDMA-enabled storage is Microsoft SMBDirect. Figure 4 shows the high performance achieved when using SMB Direct over RoCE.
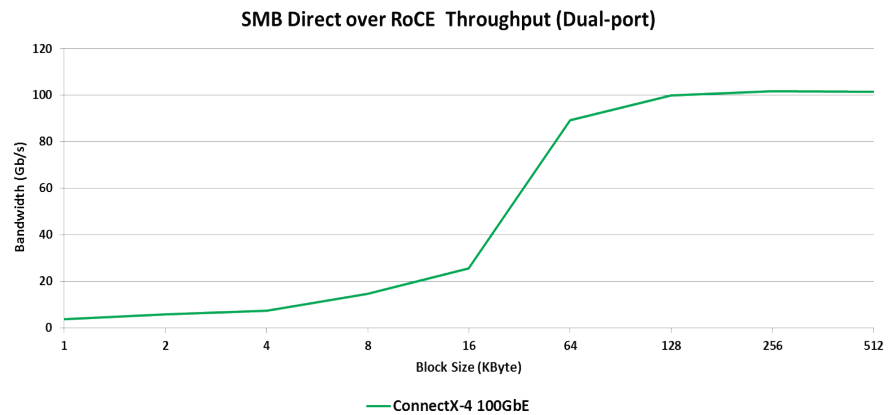
SMB Direct over RoCE Throughput (Dual-port)

Figure 4. *High performance when using SMB Direct over RoCE*

Improve CPU Utilization with Advanced Open vSwitch (OVS) Offload

Implementing networking tasks such as OVS switching in software taxes the host CPU with networking tasks instead of concentrating on running the user's applications. This results in massive cost burdens and performance inefficiencies.

Mellanox ConnectX-4 can implement hardware offloading of VM switching, which reduces CPU overhead and saves costs.

ConnectX-4 features a granular per-flow offload policy. A decision can be made for each flow whether to offload it or to process it within the CPU. If the flow is processed in the hardware, advanced policies and offloads can be implemented, such as overlay network protocols offloading, access control lists, forwarding, and so on.

The OVS offload enables the user to continue working with OVS without changing application behavior, while still receiving the benefits from SR-IOV's high performance and ConnectX-4's hardware offloads, thereby increasing bandwidth availability, achieving high packet rate, and improving CPU utilization.
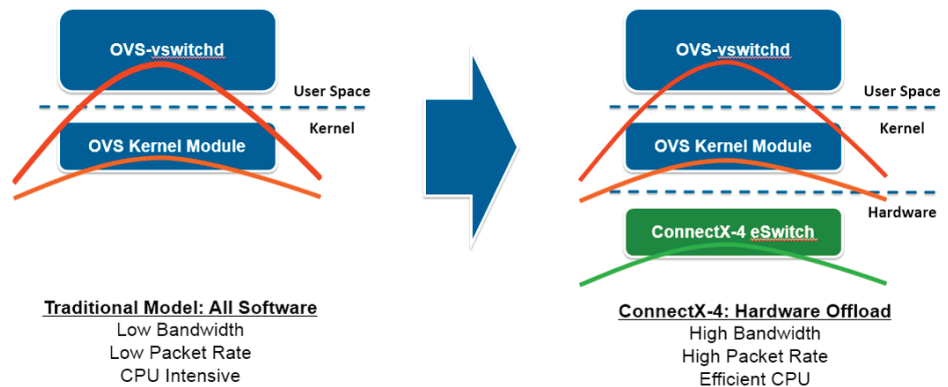

***Figure 5**. OVS offload frees the CPU and provides performance benefits*

Data Center Density and Heterogeneity

In order to increase efficiency, data center administrators are constantly looking into innovative designs that provide better compute density, lower power and space footprint, and lower complexity overall.

Mellanox Multi-Host™ Technology is the next step in high data center density and efficient rack design. The Multi-Host technology allows multiple compute or storage hosts to be connected into a single interconnect adapter by separating the adapter PCIe interface into multiple independent PCIe interfaces.

This reduces the number of adapters needed, reduces cables and switch ports.

Multi-Host technology is also used for providing different CPUs or Sockets direct access to the fabric, lowering latency and freeing-up CPU cycles.

Multi-Host supports a heterogeneous data center architecture; the various hosts connected to the single adapter can be x86, Power, GPU, ARM, FPGA, or memory, thereby removing any limitations in passing data or communicating between compute elements.
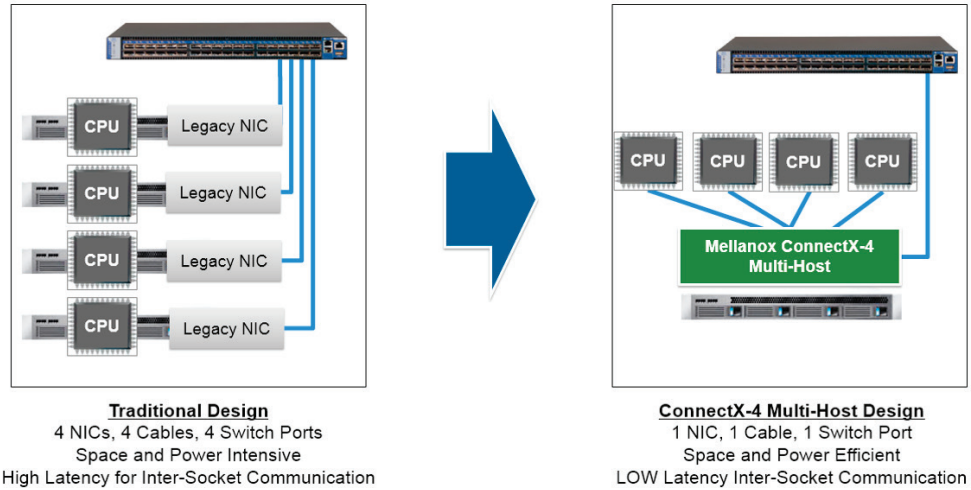
Traditional Design
4 NICs, 4 Cables, 4 Switch Ports
Space and Power Intensive
High Latency for Inter-Socket Communication

ConnectX-4 Multi-Host Design
1 NIC, 1 Cable, 1 Switch Port
Space and Power Efficient
LOW Latency Inter-Socket Communication

Figure 6. Mellanox Multi-Host saves on hardware and power costs and improves performance

Ease of Deployment

Many times, network related deployment in the cloud is overlooked, thus creating unnecessarily high deployment and maintenance costs.

In order for the cloud operation to work smoothly with the networking equipment, one should take into consideration aspects as in the below diagram.

Figure 7. Ease of Deployment Considerations

Mellanox solutions for the cloud, feature a high level of integration in upstream O/S distributions and in commercial cloud products, plus Mellanox management and automation tools and expertise.

These advantages enable cloud administrators to deploy full production environments in 1 to 3 weeks with standard off-the-shelf software and hardware.

Inbox Drivers – Mellanox drivers are inbox in the major distributions releases. For additional required functionality, Mellanox OFED drivers are publicly available on the Mellanox website.

http://www.mellanox.com/page/products_dyn?product_family=26&mtag=linux_sw_drivers

Cloud Distribution Integration – Mellanox is constantly working to make its solution the default in OpenStack distributions such as Mirantis, RDO, Canonical, and others.

Provisioning and Management Platform – The NEO management platform is offered to Mellanox customers for automating cloud deployment and provisioning, and for enabling day-by-day network monitoring and operations.

NEO features an extensive REST API, enabling it to integrate in any existing management framework seamlessly.

Professional Services – With a large variety of clouds already deployed, whether with service providers, private deployments, or HPC, Mellanox Professional Services are very experienced in helping customers effectively deploy cloud environments.

Community – For those who would like to handle deployment and operation by themselves, Mellanox provides extensive documentation and community discussion on the Mellanox Community website:

https://community.mellanox.com/docs/DOC-2288

CloudX Reference Architectures for Leading Cloud Platforms

In order to maximize ROI and protect existing and costly infrastructure investments, it is crucial that the networking solution take part in the overall cloud deployment, provisioning, and operations processes and automation.

OpenStack Integration

Mellanox contributes a great deal of code to various OpenStack projects, as part of:

1. Automating the cloud management process in relation to networking devices
2. Optimizing their networking-related performance

The following integrations are available today:

- OpenStack Neutron and Nova – for cloud networking automation
- OpenStack Cinder – high latency storage solution with iSER
- OpenStack Ceilometer – for network metering

The full reference architecture document can be found at the following link:

http://www.mellanox.com/openstack/pdf/mellanox-openstack-solution.pdf

VMware vCloud Integration

VMware clouds deployed with Mellanox technology can take advantage of the Virtual eXtensible Local Area Network (VXLAN) offload capabilities of Mellanox ConnectX®-3 Pro (and above) adapters to provide both the scalability and efficiency of VXLAN without the typical associated penalties (significant drop off in network throughput and increase in CPU utilization).

In addition, VMware based clouds can benefit from Mellanox high-speed adapters by enabling high performance to the VM, and increasing the VM density on the hypervisor.

The full reference architecture document can be found at the following link:

http://www.mellanox.com/related-docs/whitepapers/CloudX_VMware_5_5_with_vShield_Reference_Guide.pdf

Windows Azure Pack Integration

Mellanox enables organizations to build the most efficient clouds, running on Windows Azure Pack. Utilizing technologies such as SR-IOV, Network Overlay offloads, and RDMA, Windows clouds built with Mellanox adapters achieve the highest levels of performance, efficiency, and scalability.

Microsoft's Hyper-V hypervisor uses NVGRE network virtualization for tenant traffic. This is efficiently accelerated by Mellanox's ConnectX®-3 Pro (and above) network adapters, which reduces CPU overheads caused by network communication and produces higher density of virtual machines.

In Windows Server 2012 R2, Hyper-V utilizes RDMA/RoCE for live migration. This technology significantly reduces the time required for virtual machine migration, and especially the CPU overhead caused by live migrations. As the migration costs are greatly reduced, operators can load balance the cloud traffic faster. This allows the cluster to be operated with smaller resource reserves, without harming any SLA.

The cloud storage uses Scale-Out File Server enhanced by Microsoft Server Message Block (SMB)

Protocol Version 3.0 on Microsoft Storage Spaces. The SMB 3.0 file servers use SMB Direct over RoCE to reduce CPU overhead that arises with storage access, ensuring the best possible performance.

The full reference architecture document can be found at the following link:

http://www.mellanox.com/related-docs/applications/Windows_Azure_Pack_(WAP)_CloudX_Reference_Guide_v1.0.pdf

Mellanox Adapters Address Your Cloud Requirements

Mellanox adapters provide the most unique combination of advantages for cloud deployments, for a wide range of needs and use cases.

Whether the priority is for the highest raw data rate, running HPC applications, low latency storage, or VM density, Mellanox adapters provide the most advanced and most efficient solution on the market today.

Advantages of Mellanox Adapters for Cloud Use Cases	
10G / 25G / 40G / 50G /100G Highest Performance	✓
Enhanced VM Density	✓
Run Low Latency Applications such as HPC in the VM	✓
Accelerate Storage with No CPU Effort	✓
Free Up CPU with Intelligent Offloads	✓
Overlay Networks Stateless and Encap/Decap Offloads	✓
Enhance Rack Density and Reduce Cable and Port Complexity	✓
Integrate with Leading Cloud Vendors (Open Source and Commercial)	✓
Easy to Deploy	✓
Proven Field Solution in Large Clouds	✓

Mellanox TECHNOLOGIES

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