We already have 8 or even 16 cores on one CPU chip, hardware-based CPU virtualization, servers with hundreds of gigabytes of memory and Numa architectures with endless memory bandwidth (hundred GB/s memory traffic with standard server), and even the disks are now much faster with SSD technology. So it seems like we can now efficiently consolidate our applications to much fewer physical servers, or are we missing anything?

With all those new capabilities, why do users experience slow application response time? Why isn’t the application performance predictable? And can we guarantee real isolation between virtual machines (VMs)?

The answer is simple, the current bottleneck is in the I/O and the network. If we run 10 VMs on a server, they generate 10 times or more I/O traffic. Virtualized infrastructure adds even more traffic over the network to satisfy VM migration, access to remote shared storage, high-availability, etc.

When multiple VMs share the same network adapter, or when different traffic types (like VM networking, storage, migration, fault tolerance) run over the same wire, one can easily interfere with the other. An important message can wait behind far less critical burst of data, or one user can run a bandwidth consuming benchmark that will deny service from other users.
To eliminate the I/O problem we need to address the following:

1. Provide faster network and storage I/O with lower CPU/Hypervisor overhead
2. Deliver flatter/lower-latency interconnects with more node to node bandwidth (and less blocking)
3. Guarantee isolation between conflicting traffic types and virtual machines
4. Optimize power, cost, and operation efficiency at scale

A critical element in the solution is to integrate the I/O virtualization and provisioning with the overall cloud management and hypervisor framework, so it becomes seamless to the end user.

Mellanox products and solutions are uniquely designed to address the Virtualized infrastructure challenges, delivering best in class and highest performance server and storage connectivity to various demanding markets and applications. Providing true hardware-based I/O isolation and network convergence, with unmatched scalability and efficiency. Mellanox solutions are designed to simplify deployment and maintenance through automated monitoring and provisioning and seamless integration with the major cloud frameworks.

The following document covers Mellanox I/O virtualization solution and its benefits.

**Mellanox I/O Virtualization Features and Benefits**

**Highest I/O Performance**

Mellanox provides the ConnectX® family of I/O adapters. These are the industry’s fastest, and support dual port 10/40 Gigabit Ethernet and/or 40/56 Gb/s InfiniBand. Using Mellanox ConnectX, we can drive far more bandwidth out of each node, while offloading, acceleration, and RDMA features can greatly reduce the CPU overhead, leading to better performance and higher efficiency.

![Figure 1. VM network performance using Mellanox ConnectX vs. alternative](image-url)

Figure 1 demonstrates how using the connect adapter with 40GbE interface a user can deliver much faster I/O traffic than using multiple 10GbE ports from competitors.

Some demanding applications such as databases, low-latency messaging and data intensive workloads may require bare metal performance and would need to bypass the hypervisor completely when issuing I/O. Mellanox ConnectX® supports multiple physical functions on the same adapter and SR-IOV (Single Root I/O Virtualization), allowing direct mapping of VMs to I/O adapter resources, bare metal performance and zero hypervisor CPU overhead. Note that some hypervisors may block live VM migration capability.
when direct mapping is used (cold migration is still supported), which is probably a valid tradeoff for performance demanding applications.

**Figure 2.** Message latency with SR-IOV enabled compared to traditional NIC solution. SR-IOV enables the same performance across virtual and virtual infrastructure.

**Hardware Based I/O Isolation**

Providing very fast I/O access can also be dangerous. VMs can generate loads of traffic, consuming the entire network, creating up stream or resource congestion problems, and denying service from other more mission critical applications.

In addition, the hypervisor must also have a guaranteed chunk of traffic for its own management, storage, and VM migration traffic. If a certain amount of I/O traffic cannot be guaranteed for the hypervisor use, the system can malfunction or even fail. That is why many users don’t take any chances and just install multiple adapters for different traffic classes, however installing multiple adapters drive much higher costs (require multiple adapters, switches, cables, etc.) and complexity.

Mellanox ConnectX adapters and Mellanox switches provide high degree of traffic isolation in hardware, allowing true fabric convergence without compromising service quality and without taking additional CPU cycles for the I/O processing. Mellanox solutions provide end-to-end traffic and congestion isolation for fabric partitions, and granular control of allocated fabric resources.

**Figure 3.** Mellanox ConnectX providing hardware enforced I/O Virtualization, isolation, and Quality of Service (QoS)
Every ConnectX adapter can provide thousands of I/O channels (Queues) and more than a hundred virtual PCI (SR-IOV) devices, which can be assigned dynamically to form virtual NICs and virtual storage HBAs. The channels and virtualized I/O can be controlled by an advanced multi stage scheduler, controlling the bandwidth and priority per virtual NIC/HBA or a group of virtual I/O adapters. Thereby ensuring traffic streams are isolated, and that traffic is allocated and prioritized according to application and business needs.

**Accelerating Storage Access**

In addition to providing better network performance, ConnectX’s RDMA capabilities can be used to accelerate hypervisor traffic such as storage access, VM migration, data and VM replication. The use of RDMA pushes the task of moving data from node-to-node to the ConnectX hardware, yielding much faster performance, lower latency/access-time, and lower CPU overhead.

![Figure 4. Example using RDMA based storage access vs. traditional I/O](image)

Figure 4 demonstrates the storage throughput achieved when using RDMA-based iSCSI (iSER) compared to traditional TCP/IP based iSCSI, and how RDMA can provide 10X faster performance.

When deploying a rack of servers, each hosting a couple of dozen virtual machines, the result is 1,000 VMs generating independent I/O transactions to storage. Because of the high density and random I/O patterns involved, most traditional storage systems cannot cope with the load and produce extremely slow performance and response time. Leading to underutilized CPU resources and degraded application performance.

Mellanox’s award winning Storage Accelerator (VSA) software or hardware appliance (installed on a standard server) is designed to deliver parallel storage access across multiple 40/56Gb Ethernet or InfiniBand fabric to interconnect, and eliminate the storage bottleneck. It provides the following key features:

- Use internal fast SSD or Flash for boot, temporary data, or storage cache
- Significantly increases VM performance and the VMs per server consolidation ratio
- Delivers easy-to-use FC I/O port virtualization, enabling fabric convergence
- Provide simple management and monitoring of all storage traffic

©2012 Mellanox Technologies. All rights reserved.
Automated Fabric and I/O Provisioning

The Fabric is a key element in any cloud solution; Virtual Machines can be assigned to multiple tenants and share the same fabric, but each VM may want to have its own isolated domain and private networks. Each VM may also require a certain amount of allocated network resources (bandwidth, priority, VLANs, etc.). Since VMs are deployed dynamically and can migrate from one server to another, it is key to have a dynamic network virtualization and resource management solution, which will integrate with the overall cloud management framework.

Mellanox Unified Fabric Manager™ (UFM™) provides service oriented network provisioning, virtualization, and monitoring. It utilizes industry standard Quantum REST API (part of OpenStack) for network and I/O provisioning in virtualized environments, and it is integrated with a variety of cloud frameworks. This allows seamless operation of the cloud while ensuring network isolation, security, and SLA management.

Figure 5. Mellanox VSA Based Storage Acceleration Solution

Figure 6. Cloud solution with seamless integration of server, storage, and network provisioning
In today’s virtualized data center, I/O is the key bottleneck leading to degraded application performance and poor service levels. Furthermore, infrastructure consolidation and a cloud model mandate that I/O and network resources be partitioned, secured, and automated.

Mellanox products and solutions enable high-performance and an efficient cloud infrastructure. With Mellanox, users do not need to compromise their performance, application service level, security, or usability in virtualized environments. Mellanox provides the most cost effective cloud infrastructure.

Our solutions deliver the following features:

- Fastest I/O adapters with 10, 40, and 56 Gb/s per port and sub 1us latency
- Low-latency and high-throughput VM-to-VM performance with full OS bypass and RDMA
- Accelerated storage access with up to 6GB/s throughput per VM
- Hardware-based I/O Virtualization and network isolation
- I/O consolidation of LAN, IPC, and storage over a single wire
- Cost-effective, high-density switches and fabric architecture
- End-to-end I/O and network provisioning, with native integration into key cloud frameworks

Mellanox solutions address the critical elements in the cloud; bringing performance, isolation, operational efficiency, and simplicity to the new enterprise data center, while dramatically reducing the total solution cost. Please visit us at www.mellanox.com to learn more about Mellanox products and solutions.