WHITEPAPER

March 2008

Importance of Unified I/O in VMware® ESX Servers

1.0 Why Unified I/O

More and more virtual servers (or virtual machines) can be now packed into single physical servers – thanks to advances in multi-core CPU technologies and efficient use of them by virtualization software makers like VMware. The result is the need for more I/O capacity per server and VMware best practices recommend use of multiple I/O adapters to meet the needs of today's virtualized servers. The only way to reduce I/O adapter-per-server and cabling sprawl in the data center infrastructure is to use higher performance and unified I/O solutions. Such solutions can reduce I/O related capital and operational expenses and make data centers greener.

1.1 Unified I/O Options – Two Worlds

When it comes to unifying I/O on the servers, there are only two options – 10GigE NICs or InfiniBand HCAs. What should you deploy, especially in VMware ESX server environments? Depending on your take on the following decisions criteria, the answer can be either. The decision criteria is driven by (a) your choice of SAN (storage area network), (b) you desire to deploy solutions today versus wait for future technologies, and (c) your price-performance objectives. Use of SANs is a requirement to avail of VMware virtual infrastructure value adds like virtual machine migration for increasing server utilization, high availability and disaster recovery. Whether you use Fibre Channel SANs or iSCSI SANs has a bearing on what unified I/O technology is practically deployable today. Per port cost (adapter and switch port) and maturity of available unified I/O technologies is the second important consideration. This paper discusses these three criteria.

1.2 Delivering Unified I/O with Fibre Channel SANs Today

Fibre Channel (FC) SANs are widely deployed in fortune 500 and large data centers in virtualized server envioronments using VMware virtualization software. VMware Virtual Infrastructure functions that span entire data centers, across geographical sites, rely heavily on

FC SANs to deliver functions such as high availability, VMotion® and Storage VMotion. On the host side, where VMware ESX Server 3.5 runs, FC adapters (HBAs) cannot unify I/O. The choice is between 10GigE and InfiniBand. FC SAN services require a reliable fabric (without data or packet loss) for carrying critical storage data. Ethernet, being a best effort medium, cannot provide such reliable services making it highly unsuitable for carrying storage traffic in mission critical applications in data centers. InfiniBand, on the other hand, provides the required reliable services at significantly higher bandwidths than FC. So, in this scenario, if unified I/O on servers is a must, if high-performance I/O is needed to meet the I/O demands of multiple virtual machines on VMware ESX Server 3.5, and if FC SAN connectivity is mandatory, InfiniBand is the only deployable solution today.

End-to-end FC SAN connectivity over InfiniBand is being deployed at many data centers today. I/O Director or gateway solutions are available from many InfiniBand solution vendors. These solutions include multiple InfiniBand ports that connect to servers. On the infrastructure side, they contain line cards or I/O modules that provide Ethernet and FC ports for connectivity to LAN and FC SAN respectively. By doing so, on the server side, they enable "one wire" deployments to reduce cabling complexity (up to 70%), and "one adapter" deployments to reduce I/O cost and power significantly (30 – 50%). At the same time, such solutions expose flexible virtual NICs and virtual HBAs that allow applications in VMs in VMware ESX Server 3.5 run transparently (they run over legacy NIC and HBA interfaces on ESX VMs). I/O provisioning for VMs and virtual infrastructure functions using VMware VirtualCenter is kept transparent the same way, enabling end-to-end Ethernet and FC SAN experience and ease of deployment.

1.3 What About Fibre Channel over Ethernet (FCoE)?

The industry initiatives around FCoE further validates the deficiencies in Ethernet and the suitability of InfiniBand for unifying server I/O when FC SAN connectivity is needed. While FCoE holds promises, specifications are still in development and the ecosystem of software, hardware and solution suppliers is still in the early formative stages. On the end nodes, FCoE adapter solutions are a new breed with brand new software, and OS vendor adoption and stack maturity will take time. On the infrastructure side, switches and gateways delivering reliable services over Ethernet will have to be developed and deployed – there are no deployable products today or in the near future. When new technologies require simultaneous upgrades on the end nodes and the infrastructure, history tells us that real data center deployments can take many years.

1.4 10GigE as Unified I/O for LAN and iSCSI SAN

If you are using iSCSI SANs, VMware ESX Server 3.5, with its support for iSCSI and accelerated I/O using its NetQueue specification, offers a unified I/O solution using 10GigE NICs. Many I/O vendors have announced close to line rate throughput for LAN and iSCSI traffic from VMs. However, CPU utilization is high, requiring use of multiple VMs and 8 core CPU based servers to attain that throughput. 10GigE NICs used with VMware ESX Server 3.5 provides a viable solution for replacing multiple GigE NICs with a single adapter, while boosting I/O performance and reducing cabling complexity. Such solutions are also good for blade servers where the number of available PCI slots is limited.

1.5 Best of Both Worlds

In the ideal world, to cover all use case scenarios and to future proof, one would like to use a unified I/O solution on their VMware ESX server that deliver the best of both worlds. That is, use InfiniBand where it makes sense or else use 10GigE. Mellanox ConnectX™ I/O adapter provides this unique flexibility in VMware ESX Server 3.5 environments. ConnectX I/O adapters are protocol agile in that the same adapter can be configured to operate as a 10/20/40 Gb/s InfiniBand HCA or a 10GigE NIC. Protocol agility is implemented through simple switch of firmware in the adapter card which results in the appropriate I/O device drivers and protocols taking effect in the VMware ESX Server 3.5 environment. ConnectX InfiniBand software for VMware ESX Server 3.5 is available today. ConnectX 10GigE software for VMware ESX Server 3.5 will be available in Q2 2008.

1.6 Scaling Virtual Machines, Boosting Server Utilization

ConnectX 10GigE NICs used with VMware ESX Server 3.5 delivers scaling of number of VMs in physical servers without compromising I/O performance. For example, 10GigE wire-speed can be maintained when the number of VMs is scaled (in 8-core CPU based servers for example) from five to sixteen VMs. 30-40% of the CPU cycles can be left aside for applications in VMs while maintaining close to line rate for each VM. Overall, Mellanox 10GigE NICs can help increase overall server utilization by eliminating I/O bottleneck and enabling more virtual machines per server.

1.7 Price-Performance Considerations

InfiniBand is the most mature 10 Gb/s+ I/O technology in the market today, having shipped close to 3M ports as of December 2007. That dwarfs 10GigE by a large margin. And that reflects on the fact that InfiniBand products today are in their fourth generation, having gone through significant feature upgrades, optimizations and integration. The proof is in the pudding - InfiniBand per port cost (switch and adapter) stands at about 50% that of 4 Gb/s FC and 30% that of 10GigE.

ConnectX InfiniBand used with VMware ESX Server 3.5 delivers end-to-end FC SAN connectivity and best price-performance. It delivers 4 times SAN I/O throughput (1.5 GB/s) from ESX VMs compared to 4 Gb/s FC. It delivers 3-4 times LAN I/O throughput from ESX VMs compared to GigE NICs. As such, a single ConnectX InfiniBand adapter can replace multiple GigE and FC adapters, reducing I/O cost by 50% and I/O power by up to 30%.

1.8 Complete Future Proofing

For those who want future proofing or want to wait for FC SAN consolidation over 10GigE, ConnectX 10GigE NICs support FCoE, with vital SCSI and FC related CPU intensive functions offloaded to the adapter hardware to deliver higher performance at lower CPU utilization.

For those who want I/O solutions that can support direct VM access using evolving standards such as PCI SIG I/O Virtualization – Single Root (SR) and reduce CPU utilization further, ConnectX adapters implement PCI SIG defined virtual functions (VFs). VFs allow the I/O adapter resources to be divided and directly accessed by the VMs, like in native OS environments. ConnectX complements those features by enabling multiple send and receive queues per VF, advanced memory management and flexible receive slide scaling and filtering capabilities to provide full isolation, protection and performance to VMs. Such hardware-based I/O virtualization solutions will be supported when they are enabled in with future virtualization software solutions.

1.9 Summary

Unified I/O in VMware ESX servers deliver many cost and power savings advantages. 10GigE and InfiniBand are two options, but they pose unique opportunities and challenges that need to be given due considerations. When it comes to unifying server I/O with best price-performance and end-to-end FC SAN connectivity, InfiniBand is the only solution that is deployable today and in the near future. 10GigE adapters are useful when unifying server I/O in iSCSI SAN environments and can provide high throughput for VMs. Mellanox's protocol agile ConnectX adapters deliver on promises in both the InfiniBand and 10GigE worlds, now! And with advanced features such as FCoE and PCI SIG IOV SR built in, ConnectX delivers peace of mind with complete future proofing.

Mellanox TECHNOLOGIES

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