SNIA™ Ethernet Storage Forum

www.snia-europe.org

10GbE comes of age

By David Flair, SNIA's Ethernet Storage Forum Board member, Intel.

The IEEE standard for 10 Gigabit Ethernet (10GbE) network technologies was ratified over ten years ago, but adoption has been slow. Today, technical and economic reasons are creating an environment in the data center primed for 10GbE adoption. So, will 2012 be the year for 10GbE?

Technical Drivers

Virtualization: Virtual machine adoption is growing rapidly with many analysts citing that virtual server shipments now exceed physical servers. As virtual machine density continues to increase on the server, 10GbE becomes more attractive.

Network Convergence: Another complementary trend in the data center is the consolidation of resources. Networked storage, such as SANs and NAS, introduced a wave of resource consolidation. So, as Ethernet continues to increase in capability to support not only iSCSI, but now Fibre Channel traffic with FCoE, the opportunity to consolidate both the LAN and the traditional storage network is a reality. 1GbE cannot support FCoE. 10GbE offers the required bandwidth for converged traffic while also introducing some significant economic benefits.

Economic Drivers

Hardware Cost: 10GbE delivers both increased performance and economic value. Fewer adapters, fewer cables, and fewer switch ports are required to support the same data traffic of previous generation products. Price reductions for 10GbE are now at the point where the cost per Gigabit of bandwidth is less for 10GbE versus 1GbE. Per port costs for 10GbE switches are dropping rapidly as demand for 10GbE is now driving volume.

Green Initiatives: Consolidating onto 10GbE from 1GbE reduces the number of adapters, cables and switch ports required to support the same I/O requirements. Reductions in equipment translate into less power and cooling requirements in additional to the reduction in equipment costs.

Why 2012 is a big year

Romley Platform: The latest Intel server and workstation platforms, the Intel® Xeon® processor E5 family will also significantly help drive broad adoption of 10GbE in 2012. These processors introduce three major advancements to facilitate high bandwidth and low latency Ethernet traffic. First, the PCI Express interface is, for the first time, on the processor itself rather than on a separate I/O Hub. This eliminates a bottleneck and a bus hop for Ethernet data to get to and from the processor. Second, the E5 family implements PCI Express® 3.0, which doubles the bandwidth per pin compared to PCI Express 2.0. PCIe3 will support four channels of 10GbE or one channel of 40GbE on a single PCIe x8 connection. Lastly, the E5 introduces Intel Data Direct I/O (DDIO), a radical re-architecting of the flow of I/O data in the processor with tremendous benefits for Ethernet traffic in terms of increased bandwidth and lower latency allowing Ethernet NICs and controllers to talk directly to the processor's last-level cache without a detour to main memory in either direction.

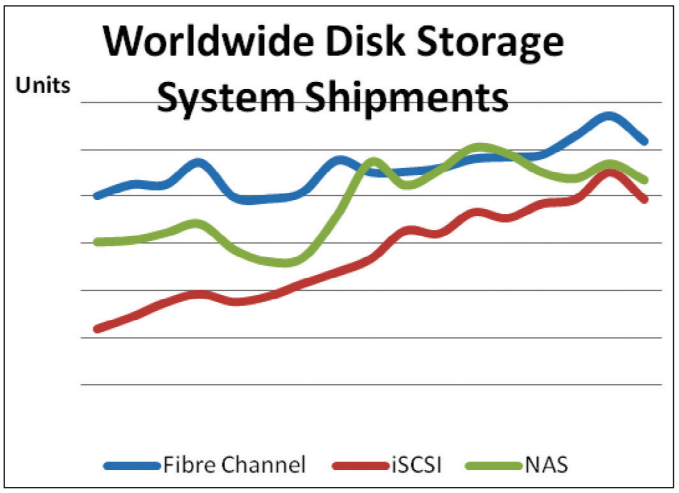

Figure 1 Source: Worldwide Quarterly Disk Storage Systems Tracker - 2012 Q1, IDC, June 2012

10GBASE-T: 10GbE technology has now penetrated 10-15% of the Ethernet connections in the data center. 10GBASE-T, the interface technology that uses the familiar RJ-45 jack and low-cost Categories 6 and 6a twisted pair cables also fuels 10GbE adoption because switches that support 10GBASE-T can also support 1GbE, data center administrators can deploy 10GbE in an evolutionary manner based on changing requirements. Tested to deliver the reliability and low bit error rates required by specifications, 10GBASE-T switches are now widely available. The first single-chip 10GBASE-T adapter silicon reached the market in 2012, driving lower power requirements and lower prices than the two-chip PHY and MAC 10GBASE-T adapter solutions from earlier designs.

Expanded LOM: Traditionally, server OEMs have offered their customers Ethernet technology in three forms: PCI Express Ethernet Network Interface Cards (NICs), an integrated circuit soldered to the server's motherboard called "LAN on Motherboard' (LOM), or lastly, for server blades only, a mezzanine or "daughter" card. A fourth option, which could be called "flexible LOM," simplifies and reduces the cost of the traditional daughter card for mainstream servers providing the option of 10GbE at LOM-like costs.

Adapter Virtualization

With the widespread adoption of server virtualization, several technologies are being developed to partition or virtualize network adapter ports to fully utilize the 10GbE bandwidth.

With NIC partitioning, multiple NIC functions are created for each physical port. These functions are equivalent to unique physical devices so there is no requirement for new virtualization capabilities in operating systems or hypervisors.

Single Root I/O Virtualization (SR-IOV) is a PCI Special Interest Group (PCI-SIG) standard for "device sharing" in virtualized servers. The SR-IOV specification allows an I/O device to appear as multiple physical and virtual devices, using the concept of physical and virtual functions that "look" like physical functions to a VM. With SR-IOV, VFs can be assigned directly to VMs, bypassing the I/O overhead in the hypervisor. SR-IOV is currently supported with Kernel Virtual Machine (KVM) in Red Hat Enterprise Linux 6 and SUSE Enterprise Linux 11 (and later). Microsoft has announced support for SR-IOV Windows Server 2012 Hyper-V. Other hypervisor providers are expected to announce SR-IOV support.

Data Center Bridging: Enhancements made to the Ethernet protocol, collectively referred to as Data Center Bridging (DCB), enable support of FCoE and converged data storage traffic over a shared Ethernet wire. These enhancements offer the ability to allocate bandwidth as well as improve management of traffic flow to ensure lossless transmission characteristics. The most commonly deployed enhancements are Enhanced Transmission Selection and Priority-based Flow Control. Data Center Bridging Exchange (DCBX) over LLDP, is implemented with DCB to facilitate discovery and configuration of features enabled between endpoints. DCB is supported in all recently introduced adapters and switches from the leading vendors.

Disruption of Flash Technology

As a new price/performance tier between DRAM and Hard Disk Drives (HDD), flash can offer between 10 and 100 times the speed of HDD. For workloads which require every read and write IO to be extremely fast, it is possible to configure storage arrays which consist entirely of flash technology with many terabytes of solid state disk drives (SSD). For most workloads, however, flash will provide a suitable complement to HDDs whereby caching algorithms will allow most high performance IOPS to be serviced from flash. These faster storage systems make demands on the network and increase the need for 10GbE.

What additional technologies will benefit from 10GbE deployment?

iSCSI: iSCSI rides on top of TCP/IP and Ethernet and benefits from all of the enhancements and improvements that come along with both. iSCSI has grown with a compounded annual growth rate (CAGR) of 92% from 2003 through full year 2011 (based upon IDC WW Storage Tracker, Q1 2012 . iSCSI is supported by all of the OS vendors with built-in support for initiators that support multi-path I/O (MPIO) to provide redundancy and network throughput between servers, the switching infrastructure and storage. TCP/IP protects against dropped packets from physical or network layer errors caused by network congestion. iSCSI doesn't require DCB to function on Ethernet (unlike FCoE), but it can take advantage of DCB offered by the 10GbE ecosystem, if deployed.

The NAS Protocols: Parallel NFS (pNFS) and SMB: Parallel NFS (pNFS) in NFS 4.1, represents a major step forward. pNFS benefits workloads with many small files, or very large files, especially those that run on compute clusters requiring simultaneous, parallel access to data. By allowing the aggregation of bandwidth, pNFS relieves performance issues that are associated with point-to-point connections. The SMB NAS protocol hasn't stood still either. SMB 3.0 Multi-Channel supports multiple connections to improve throughput for typical server workloads, such as database or virtualization.

These advancements in NAS network protocol stacks lead to greater performance. In conjunction with the increased bandwidth available with 10GbE, these NAS protocols become attractive both for HPC and data center use.

FCoE Maturity: Fibre Channel over Ethernet (FCoE) holds promise of significant reduction in data center costs, cable, switches, and power by unifying storage and data networks over Ethernet. Open FCoE is supported in Linux, Windows, and VMware and qualified for Brocade and Cisco switches and EMC and NetApp SANs. FCoE is now a proven and mature technology and will ride the ramp of 10GbE.

RDMA: Another capability that 10GbE brings to the data center's Remote Data Memory Access (RDMA) technology, historically used in HPC applications. RDMA is the capability to write data directly from the memory of one computer into the memory of another with minimal operating system engagement. RDMA enables very low-latency data transmissions. There are two RDMA-over-Ethernet technologies being deployed today over 10GbE, iWARP (internet Wide-Area RDMA Protocol) and RoCE (RDMA over Converged Ethernet). iWARP is layered on top of TCP/IP. In contrast, RoCE uses the InfiniBand transport – over an Ethernet wire. Both support the Open Fabrics Alliance software stack that will run on either iWARP or RoCE technologies. Also, Microsoft has announced that Windows Server 2012 will take advantage of RDMA capabilities, if available in the network, to support SMB Direct 3.0.

Where do we go from here and when?

With four times the capacity of 10GbE and the ability to cost-effectively migrate to 100GbE, 40GbE is the next logical step in the evolution of the data network. 40GbE is starting to be deployed today in aggregation links within data center networks, and by 2016 its scope is expected to reach the network edge with 40G access links to connect servers directly. Complementing 40GbE, 100GbE is a perfect choice for carrier service providers and core links in data centers.

Is 2012 the year of 10GbE? The key factors driving adoption are virtualization, consolidation, converged networking of storage over 10GbE spurred by the broad deployment of DCB and maturity of FCoE, plummeting 10GbE adapter and switch costs, new server platforms that enhance network performance, the emergence of 10GBASE-T, flexible LOM, and RDMA over Ethernet going mainstream in Windows. For all the reasons presented in this paper, we believe 10GbE will ramp steeply toward mass adoption in the data center. To learn more about SNIA's Ethernet Storage forum and to download the SNIA-written paper "10GbE Comes of Age, visit http://snia.org/forums/esf/resources/whitepapers.