

10 Gigabit Ethernet Unifying Fabric: Foundation for the Scalable Enterprise

10 Gigabit Ethernet is positioned to fulfill many expectations of Ethernet LAN technology that have been long anticipated but not yet realized by its predecessors and their alternatives. This article discusses the criteria for creating a virtual data center and how 10 Gigabit Ethernet can meet these criteria through its ability to function as a unifying fabric.

BY J. CRAIG LOWERY, PH.D.

Related Categories:

Data center technology

Data networking

Ethernet

Network fabric

Scalable enterprise

Virtual data center

Virtualization

Visit www.dell.com/powersolutions for the complete category index.

The next generation of enterprise computing infrastructure is envisioned as a collection of standard, inexpensive, even disposable computing and storage components bound by specialized software into a distributed system. Often referred to as a virtual data center, these components in aggregate potentially offer unprecedented flexibility, enhanced reliability, and reduced total cost of ownership and systems management overhead.

Technical advancements since the virtual data center was described in *Dell Power Solutions* four years ago¹ have brought this goal within reach. Key to virtual data centers is the concept of a *unifying fabric*, also known as a converged network. Today, various classes of data center traffic use their own specialized interconnect fabrics, as shown on the left side of Figure 1: storage typically uses Fibre Channel for SCSI block-level access to storage area network (SAN) devices, high-performance computing cluster nodes use a high-performance interconnect such

as InfiniBand to communicate with each other, and traditional networks typically use TCP/IP over Ethernet.

A unifying fabric, in contrast, enables all data center communications—including SAN, intra-cluster, and traditional network traffic—to use the same networking infrastructure and protocol set, as shown on the right side of Figure 1. All communication could take place through a single cable between each device and a concentrator or switch. In practical terms, various traffic classes are segregated for performance or security reasons, but they can still use the same networking technology and components. This simplification potentially leads to reduced equipment costs and a reduced number of support groups within an IT organization.

But what networking technology could serve as a foundation for building a unifying fabric? The requirements for this technology have been clearly understood for some time, and several candidates have been considered

¹ "Building the Virtual Data Center," by J. Craig Lowery, Ph.D., in *Dell Power Solutions*, February 2003, www.dell.com/content/topics/global.aspx/power/en/ps1q03_lowery.

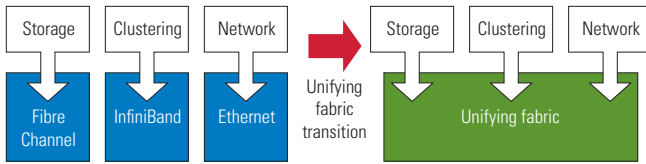


Figure 1. Traffic classes and the unifying fabric transition

and rejected. Most notable among these is InfiniBand, which has been and likely will continue to be successful in the high-performance computing inter-node fabric domain. Although it was specifically designed to be a unifying fabric,² it failed to gain acceptance as such because it was an unfamiliar, complex, expensive new technology, appearing well before its time with respect to the unifying fabric vision. And at that time, two key developments had yet to appear: virtualization and Internet SCSI (iSCSI).

Enabling convergence

Until recently, one of the primary inhibitors of unifying fabrics has been the tight coupling of software to hardware. In particular, an OS installed on a server contains drivers specific to that server’s hardware components. Booting the OS depends on a series of events occurring in the BIOS and the OS bootstrap code that are configured during installation, after which the OS is bound to that server hardware; moving the OS to different hardware (for example, by removing the hard disk from one system and inserting it in another) may not work, even if the other system is identical to the original. Even after a successful OS boot, accessing disk volumes—especially those stored as RAID sets—depends on a mutually dependent software and hardware configuration.

This tight coupling largely nullifies the promise of unifying fabrics, which derive much of their value from putting compute and storage resources in different places on the fabric and connecting them dynamically as needed. Complementary technologies that break the dependency of software on hardware (notably operating systems and their hosted applications) are, therefore, a critical requirement for a successful unifying fabric.

Virtualization is an overloaded term in the current technical lexicon. However, when defined as a means by which software is dissociated from hardware, it becomes a key enabler for a unifying fabric. Server virtualization, such as that implemented by VMware® ESX Server and Microsoft® Virtual Server, clearly separates the software stack from the server hardware by creating a virtual machine (VM) environment in which the guest operating systems execute. It effectively makes every physical server appear

to be identical to the VM guest operating systems, and therefore enables VMs to move freely to any server platform running the virtualization software.

Although server virtualization severs the ties between an OS and the underlying physical hardware, it does not provide the complete means for actually moving VMs between servers. Accomplishing this requires a type of storage virtualization in which VM state information resides on a remote storage device that can be connected to any physical hosting server participating in the fabric. The VM state information is contained in a virtual disk file—a large, flat file that mimics the functionality of a physical hard disk. To enable mobility, this file is kept on a remote storage device and accessed through the hosting virtualization software’s file system facility. The VM can then be hosted by any physical server with access to the remote storage.

Enter iSCSI

In the early releases of ESX Server, mutual access to virtual disk files was accomplished by connecting all ESX Server systems to the same Fibre Channel SAN and multiple Ethernet networks, as shown on the left side of Figure 2. However, replacing the Fibre Channel SAN with an iSCSI SAN—the method of choice for implementing IP SANs—clears the way for a unifying fabric, as shown on the right side of Figure 2.

In addition to having a physical server host VMs that reside on an iSCSI SAN, a diskless server can boot from an iSCSI target logical unit (LUN). This method helps simplify boot-time logistics for servers because they only have to support one standard—iSCSI boot—instead of several potentially proprietary hard disk controllers (especially RAID controllers). Multiple diskless servers can iSCSI

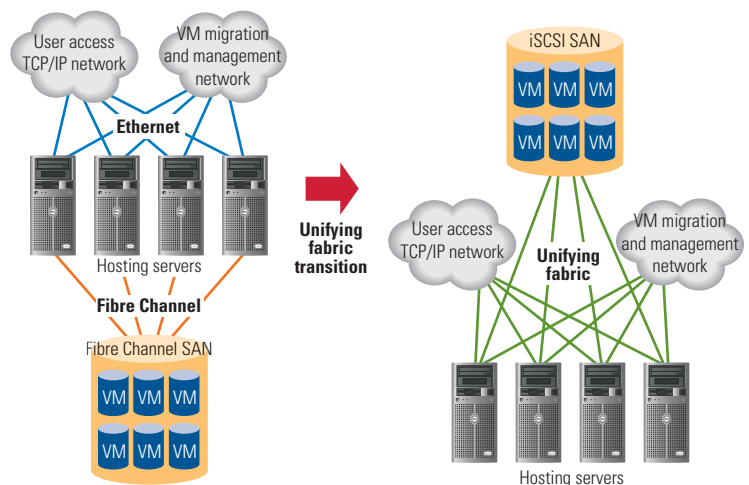


Figure 2. Virtual hosting infrastructure and the unifying fabric transition

² For more information, see “The Promise of Unified I/O Fabrics,” by J. Craig Lowery, Ph.D., and David Schmidt, in *Dell Power Solutions*, February 2005, www.dell.com/downloads/global/power/ps1q05-20040191-Lowery.pdf.

boot a common virtualization software image, such as ESX Server, and then run multiple VMs—all across the same unifying fabric.

Creating a 10 Gigabit Ethernet unifying fabric

Figure 3 shows the conceptual building blocks of the Dell scalable enterprise based on a 10 Gigabit Ethernet unifying fabric. The Dell scalable enterprise seeks to standardize core data center elements to help simplify operations, improve resource utilization, and enable cost-effective scaling as needed. Virtualization is key to that vision, including VMs supported by server and storage virtualization technologies built on iSCSI and iSCSI booting, which in turn require TCP/IP. With its attendant optimization and offload technologies, 10 Gigabit Ethernet can serve as a unifying fabric foundation to help realize the scalable enterprise vision.

Given that a unifying fabric must support storage, clustering, and traditional network traffic, and that server and storage virtualization technologies depend on iSCSI and TCP/IP protocols, a unifying fabric must meet the following requirements:

- Uses technology based on industry standards
- Becomes familiar to and trusted by the market
- Supports TCP/IP
- Coexists and interoperates transparently with existing interconnect technologies
- Provides sufficient performance (low latency and high throughput) to support iSCSI SANs and high-performance computing cluster interconnects
- Provides fault tolerance
- Offers cost-effectiveness

10 Gigabit Ethernet has emerged as the leading candidate to fulfill these requirements. As the next iteration of the established IEEE 802.3 Ethernet standards, 10 Gigabit Ethernet has a well-understood foundation on which to build; it also includes TCP/IP support and is designed to interoperate with previous generations of 10/100/1,000 Mbps Ethernet. SAN hardware from companies such as EMC provide both Fibre Channel and iSCSI capabilities, and legacy Fibre Channel enclosures can have a server front end that presents the Fibre Channel storage as iSCSI targets on the unifying fabric. Regardless, Ethernet gateways for Fibre Channel and InfiniBand already exist and can be further optimized for use with 10 Gigabit Ethernet.

Although Ethernet itself is not optimized for use as a unifying fabric, offloading technologies such as TCP/IP Offload Engine (TOE) and iSCSI offload help 10 Gigabit Ethernet provide sufficient performance for many application environments. Network interface card teaming, failover, and the robustness of the TCP/IP and iSCSI

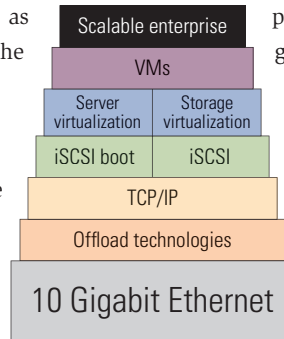


Figure 3. 10 Gigabit Ethernet as the foundation for the Dell scalable enterprise

protocols can provide fault tolerance. As with previous generations of Ethernet, cost is likely to decrease as the technology is deployed in commodity volumes, eventually making it a ubiquitous and cost-effective technology.

Moving from vision to reality

As a unifying fabric candidate, 10 Gigabit Ethernet has received widespread positive support from the IT industry, and has been deployed since 2004 as a foundational networking technology with optical and heavy-duty copper-based interconnects for high-performance computing, where it has proven to be sufficient for a majority of inter-node communication tasks.

PCI Express, protocol offloads, and multi-core processors can all contribute to removing bottlenecks that have previously prevented efficient use of 10 Gigabit Ethernet. A transition to 10 Gigabit Ethernet BaseT is on the horizon, at which time familiar CAT6 and CAT7 twisted-pair cable plants will likely be the cabling method of choice. 10 Gigabit Ethernet could become the true workhorse of enterprise computing, even supporting traffic types such as voice-over-IP telephony, videoconferencing, and remote computing.

Although it holds great promise, 10 Gigabit Ethernet as a unifying fabric requires much more than simply network adapters, switches, and cables. Management tools, methods, and best practices for iSCSI SANs and iSCSI boot (including boot image management) are needed. Refinement of protocol offloads, including integration into OS stacks and matching appropriate offloads to workloads, is still in the early stages of development. But even so, the consensus across the IT industry seems to be that a unifying fabric is a foregone conclusion, and that—with 10 Gigabit Ethernet identified as the key foundational technology—it is coming quickly. The transition may take several years, but 10 Gigabit Ethernet can take the industry one major step closer to realizing the virtual data center and the scalable enterprise. ☞

J. Craig Lowery, Ph.D., is an advanced solutions technology strategist in the office of the CTO in the Dell Product Group. He currently focuses on strategic planning for enterprise solutions, particularly those that realize the Dell vision of the scalable enterprise. Craig has a B.S. in Computing Science and Mathematics from Mississippi College and an M.S. and Ph.D. in Computer Science from Vanderbilt University.

FOR MORE INFORMATION

Internet Engineering Task Force IP Storage Working Group:
www.iETF.org/html.charters/ips-charter.html

IEEE 802.3 Working Group:
grouper.ieee.org/groups/802/3