Converged fabrics: Emerging technologies for simplifying data center infrastructure

technology brief

Abstract

Converged fabrics is a solution for simplifying data center management by consolidating all communication (server, storage, network, and remote management) onto a single fabric, in contrast to the more complex and costly approach of using proprietary fabrics for inter-process communication (IPC) and storage.

This technology brief explains the emerging and growing unified fabric technologies in which HP is involved. These technologies include 10 Gigabit Ethernet (10GbE) and Fibre Channel over Ethernet (FCoE) and how they work together to tie multiple network fabrics into a single, converged infrastructure.

Introduction

The complications of implementation, management, and cost of multiple network fabrics have prompted HP and other vendors to investigate converged fabric solutions, also known as unified fabric or fabric consolidation. Any network topology constructed with one or more switched network nodes can be described as a fabric. Therefore, *fabric* is the term commonly used to refer to individual network types, including communication, storage, management, and high speed networks.

Converged fabrics is a solution for simplifying data center management and reducing infrastructure costs. It consolidates all individual network types onto a single fabric, in contrast to the more complex and costly approach of using proprietary fabrics for IPC and storage. The converged fabrics solution is the common interconnect and can be either Infiniband (IB) or 10GbE.

HP expects converged fabrics technology to significantly lower infrastructure and management costs without sacrificing performance. This technology brief focuses on how 10GbE will be used as the common interconnect for converged fabrics.

Traditional data center topology

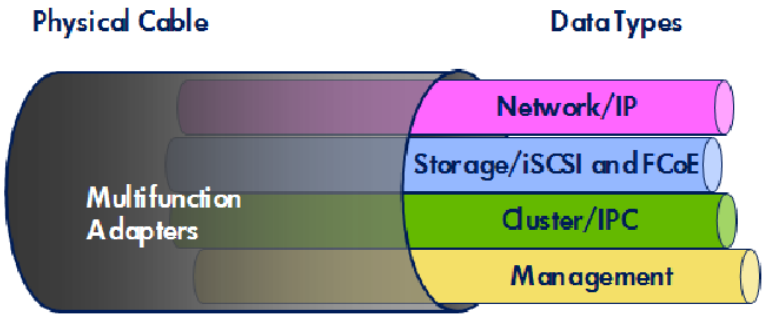
Some common characteristics of current data center infrastructure are underused capacity, inflexible single-purpose resources, and costly management. The reason is that data center topology typically includes separate, heterogeneous network appliances to manage different types of data. Each of these appliances adds to the complexity, cost, and management of the data center. Multiple networks require unique switches, network adapters for each server, and network management systems targeted for each network.

Businesses may support as many as four unique networks to manage Storage Area Network (SAN) data, IPC clustering data, remote management data, and Ethernet communications data. Each data type typically uses a different network and interconnects, which adds to the complexity of network architectures and to unifying network fabrics.

Converged fabrics with 10 Gigabit Ethernet

One obstacle to using Ethernet as a basis for converged fabrics has been its limited bandwidth. As 10GbE technology becomes more widely implemented, HP expects 10GbE network components to fulfill the needs of applications that require either the 10GbE bandwidth or its low-latency benefits. With the emergence of 10GbE, a unified Ethernet switching fabric for all data center applications is expected to serve as the basis for future data center consolidation and architectural evolution. With Ethernet and Internet Protocol (IP) as the unified switching fabric, administrators will also have maximum flexibility in selecting network management tools. As Ethernet bandwidth increases, fewer physical links can carry more data (Figure 1).

Figure 1. All traffic types sharing the same link



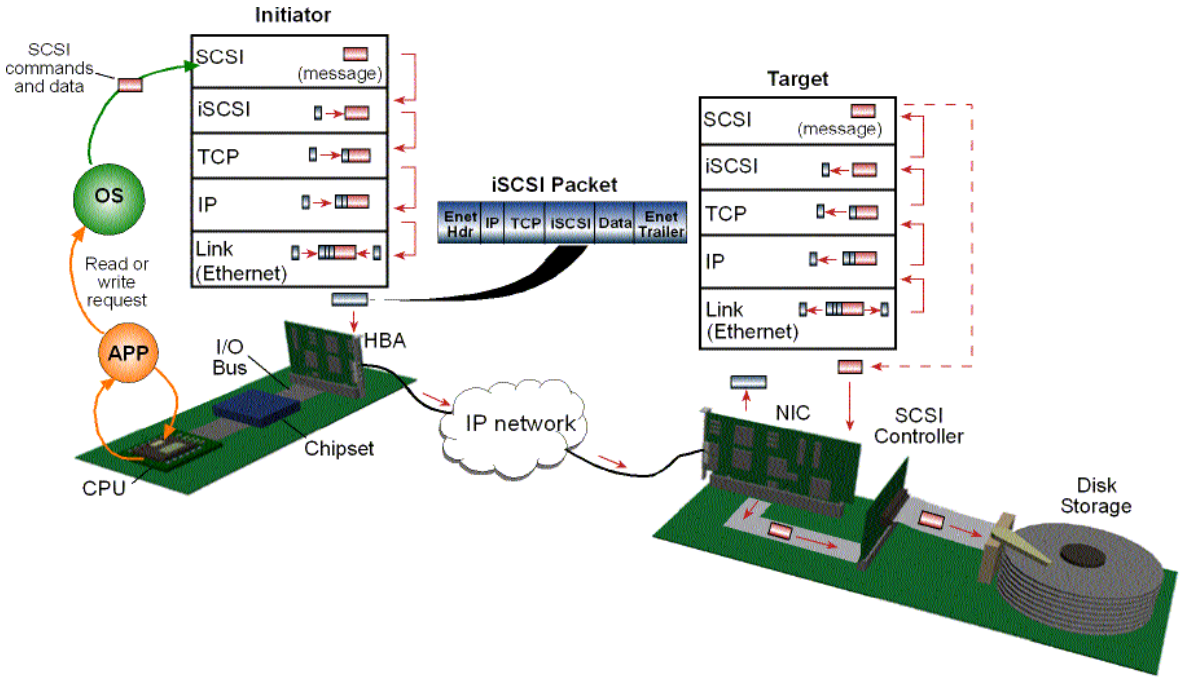
Currently, two of the most promising transport standards for converged fabrics are Internet Small Computer System Interface (iSCSI) and FCoE.

iSCSI

iSCSI follows the SCSI architectural model, which is based on message exchange between an initiator and a storage resource, or target. iSCSI initiators access targets using the iSCSI protocol. While the target is usually a drive enclosure or another computer, it can also be any other storage device that supports the iSCSI protocol, such as a tape drive. The iSCSI stack at both ends of the path encapsulates SCSI block commands into Ethernet packets for transmission over IP networks.

Figure 2 illustrates a message exchange between an initiator and a target. The process begins when an application sends a request to the operating system (OS) to read or write data. The OS generates the appropriate SCSI commands and data request in the form of a message. Before the message can be sent over an IP network, it is processed through iSCSI to encapsulate the request into the Transmission Control Protocol over Internet Protocol (TCP/IP) protocol stack (attaching routing, error checking, and control information) for transmission over the network. This can be accomplished using driver-level or OS-level software, or it can be offloaded to the host bus adapter (HBA). The HBA transmits the packets over the IP network. When the packets reach the target device, they go through a reverse process to reassemble the data, which is then moved to the SCSI controller. The SCSI controller fulfills the request by writing data to or reading data from the target device. For a read transaction, the target returns data to the initiator using the iSCSI protocol.

Figure 2. Message exchange between an initiator and target using the iSCSI protocol mode



Initiators include software initiators and HBAs. Software initiators require CPU resources to manage the protocol stack. A more efficient approach is to offload protocol management to an iSCSI HBA, such as the HP NC373i Integrated Multifunction Gigabit server adapter.

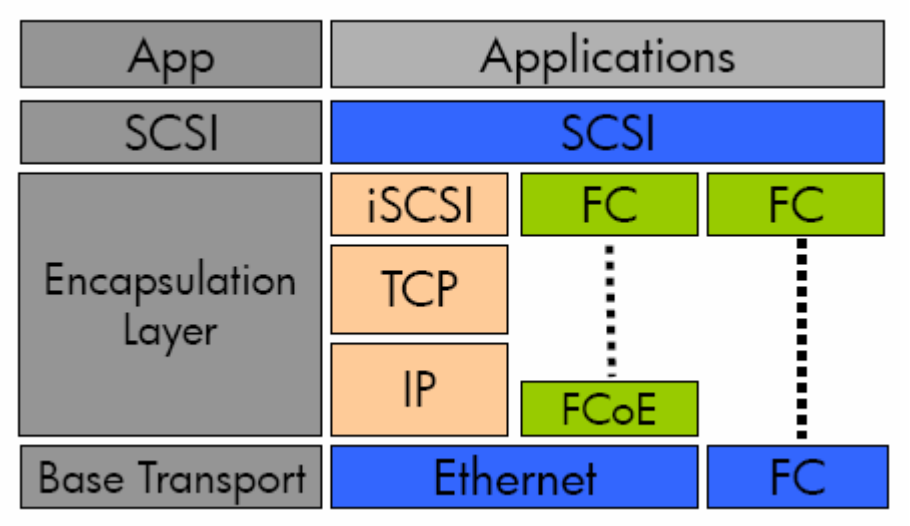
iSCSI allows Ethernet server adapters to function both as network adapters and as storage HBAs. This flexibility provides access to storage boxes and to servers over a single connection, which simplifies network infrastructure. HP multifunction adapters support hardware-assisted, or accelerated iSCSI operation—that is, offloading the iSCSI software to the server adapter for improved system performance.

Fibre Channel over Ethernet

FCoE is an emerging technology being developed by the International Committee for Information Technology Standards (INCITS) T11 technical committee. FCoE relies on flow control to recognize that a buffer is almost full and to request that the sender stop transmission until the buffer has emptied and transmission can start again. FCoE is based on Enhanced Ethernet (EE), sometimes referred to as Data Center Ethernet (DCE) or Converged Enhanced Ethernet (CEE). The Institute of Electrical and Electronics Engineers (IEEE) is working to enhance the IEEE 802 Ethernet standard to allow FC to run efficiently over Ethernet. Two recent Ethernet improvements are lossless Ethernet, which prevents lost packets, and pause-based flow control, which allows the network to selectively pause different classes of traffic.

FCoE encapsulates Fibre Channel (FC) frames within the Ethernet fabric and uses the same Open Systems Interconnection (OSI) layer as IP networks (Figure 3). In a real application environment, logic dictates using one common FCoE connection between the server and network. FCoE will use the same common driver stacks, cabling, and management applications being used today.

Figure 3. Protocol stack comparison



There are several advantages to FCoE:

- FCoE uses FC drivers, switches, and other infrastructure.
- Existing FC security and management applications are unchanged.
- FCoE uses a 10GbE fabric with no data loss.
- Existing SAN management tools can be used to access storage.
- FCoE can use enhanced Ethernet features such as traffic priority and flow control.

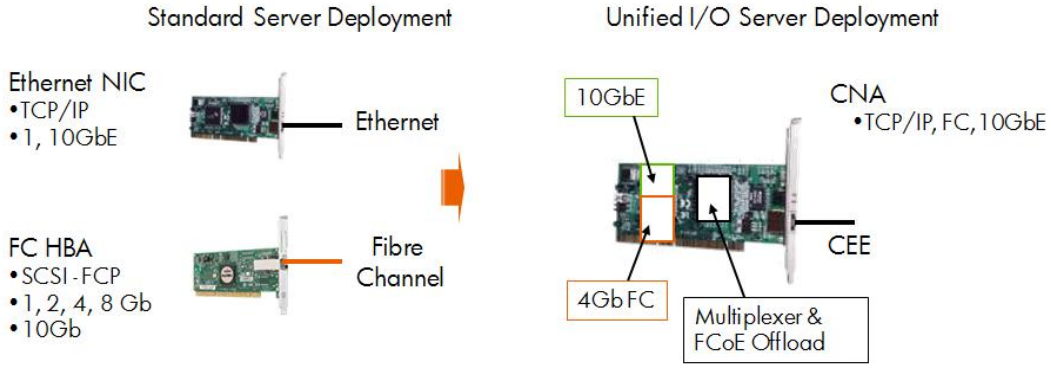
FCoE for convergence in the data center

HP is enthusiastic about the possibilities FCoE offers for network convergence. FCoE has the potential to scale in most IT environments, using the same physical Ethernet layer and common management applications in use today.

Migrating to converged fabrics

The traditional data center model uses multiple HBAs and NICs in each server to communicate with various networks. The converged model will consolidate the HBAs and NICs in a server into a single converged network adapter (CNA) to provide a unified I/O server deployment (Figure 4).

Figure 4. Standard compared to unified I/O server deployment



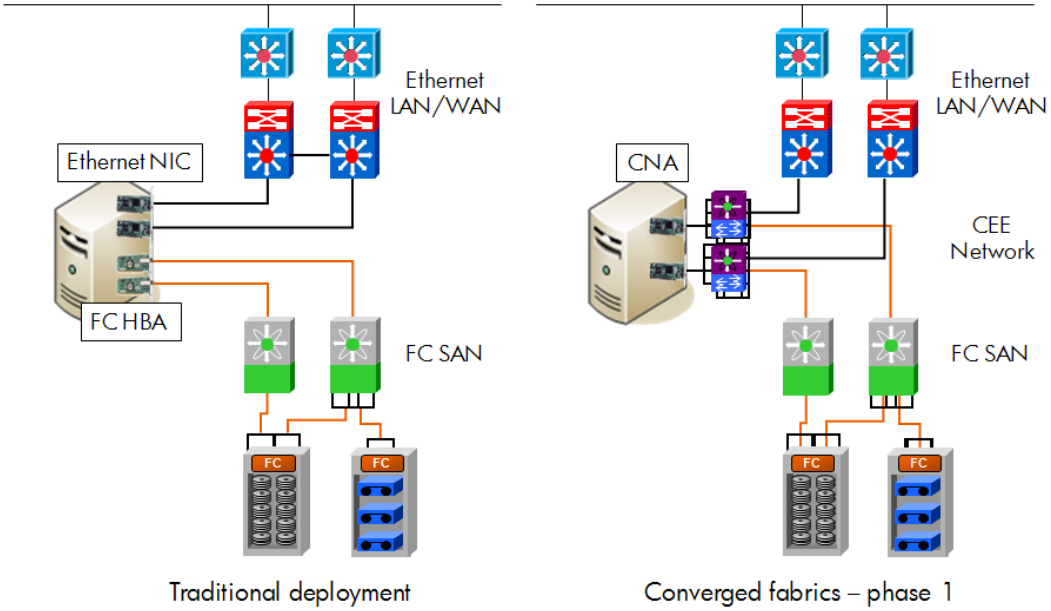
Data center bridging

Data center bridging (DCB), which is another name for DCE or CEE, prioritizes traffic types and bandwidth sharing within an Ethernet link. It provides the following enhancements to the existing Ethernet 802.1 bridge specifications that will be used in converged network deployments where all applications can be run over a single Ethernet link:

- Congestion Notification (CN), which will provide congestion management to reduce data loss
- Priority-based Flow Control (PFC), which will provide a link-level flow control mechanism to prevent loss due to congestion
- Enhanced Transmission Selection (ETS), which will provide a common management framework to assign bandwidth to different traffic classes
- A discovery and capability exchange protocol to ensure consistent configuration across the network

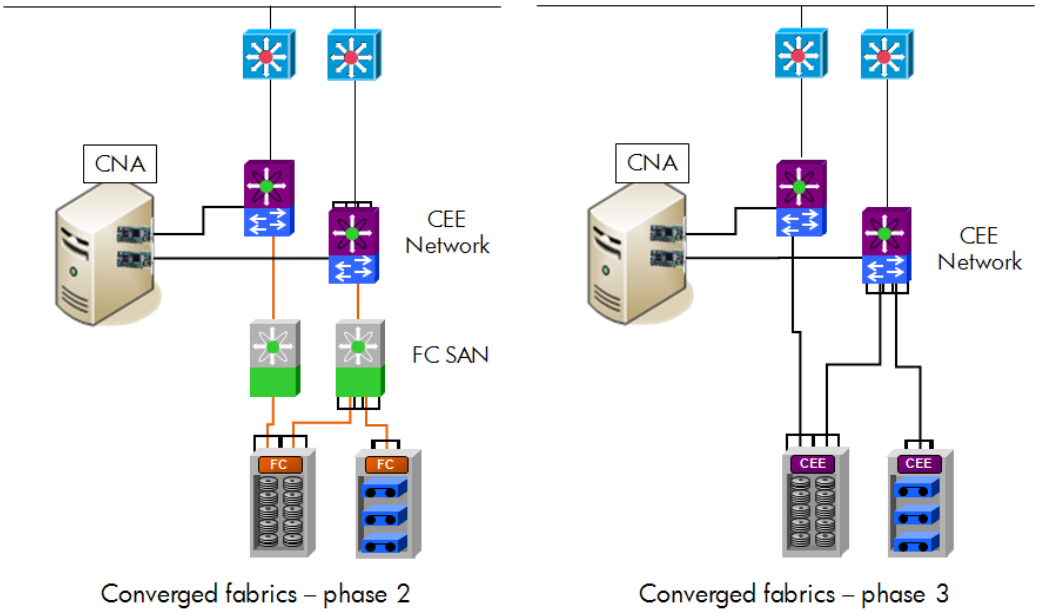
In the first phase of migration to converged fabrics, CNAs and a converged fabric bridge will provide connectivity between servers, the LAN/WAN, and the SAN (Figure 5).

Figure 5. Traditional deployment and converged fabrics, phase 1



The next phases of deployment will occur as existing data centers are updated or as new ones are built. Eventually only a single pair of redundant CNAs will be required per server. Separate FC, 10GbE, and IB switches will be replaced by converged network switches (Figure 6).

Figure 6. Converged fabrics deployment, phases 2 and 3



Virtualization in converged networks

As enterprises attempt to better utilize existing computing resources, server and network virtualization is growing rapidly. HP Virtual Connect (VC) is a set of interconnect modules and embedded software for HP BladeSystem c-Class Enclosures that simplifies server connection setup and administration. HP VC includes the HP 1/10G Virtual Connect Ethernet Module for c-Class BladeSystem, the HP Virtual Connect Manager, and the HP Flex-10 Ethernet Interconnect Module. VC uses c-Class BladeSystem adapters within the server and a new class of Ethernet interconnect modules to simplify connecting those server NICs to the data center fabric. VC also extends the capability of standard server NICs by securely administering Ethernet MAC addresses.

Server virtualization

As more virtual machines are loaded onto a physical server, the requirement for additional bandwidth increases. VMware best practice calls for six 1-Gb NICs in each physical server running virtual machines. Using this VMware guideline, just two physical servers loaded with virtual machines could fully utilize a single 10-Gb NIC. More information about virtual machine technology is available in the technology brief entitled "Server virtualization technologies for x86-based HP BladeSystem and HP ProLiant servers" at http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01067846/c01067846.pdf.

NIC virtualization

For HP BladeSystem c-Class server blades, the HP VC Ethernet Modules allow IT administrators to interconnect multiple modules and define uplinks to their data center Ethernet switches. The VC Ethernet Modules allow the administrator to select which server NIC ports will be connected to each external network.

Flex-10 for Virtual Connect

To help fully utilize 10GbE connection bandwidth, HP introduced Flex-10 technology in the BladeSystem c-Class architecture. Using Flex-10, administrators can partition the bandwidth of a single 10-Gb pipeline into multiple "FlexNICs." The bandwidth of each partition can be set to a user-defined portion of the total 10-Gb connection. Speed can be set from 100 megabits per second to 10 gigabits per second in 100 megabit increments.

There are advantages to partitioning a 10GbE pipeline:

- More NIC connections per server, which is especially important in a virtual machine environment
- Ability to match bandwidths to network functions (for example, virtual machine migration, management console, and production data)

More information about Flex-10 technology is available in the technology brief entitled "HP Flex-10 technology" at http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01608922/c01608922.pdf.

Summary

Converged fabrics is an emerging technology for simplifying data center management by consolidating server, storage, and network traffic over a common interconnect. HP is introducing intelligent, dynamic technologies such as Flex-10 to maximize the use of 10GbE connections and facilitate the transition to converged fabrics in the data center. HP expects converged fabrics implementation in the data center to significantly lower infrastructure and management costs without reducing performance. For more information

For more information

For additional information, refer to the resources listed below.

Resource description	Web address
HP Multifunction Networking Products	http://h18004.www1.hp.com/products/servers/proliant-advantage/networking.html
HP ProLiant networking Ethernet network adapters	http://h18004.www1.hp.com/products/servers/networking/index-nic.html
"Server virtualization technologies for x86-based HP BladeSystem and HP ProLiant servers" technology brief	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01067846/c01067846.pdf
"10 Gigabit Ethernet technology for industry-standard servers" technology brief	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01608915/c01608915.pdf
"HP Flex-10 technology" technology brief	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01608922/c01608922.pdf
HP Virtual Connect Technology web page	http://isscontent.americas.hpqcorp.net/products/blades/virtualconnect/

Call to action

Send comments about this paper to TechCom@HP.com.

TC090205TB, February 2009

invent