Data Center Solutions

Implementing a 10 Gb/s Physical Infrastructure to Achieve I/O Consolidation in the Data Center

WHITE PAPER

Introduction

Traditional concerns of data center stakeholders are keeping their applications running and their networks scalable, available, and secure. However, over the last five years, demand for processing power has surged and server/storage equipment densities have increased dramatically. It has become increasingly difficult to specify cost-efficient power and cooling solutions that keep up with new applications.

Technology evolution in the data center is challenging IT managers to effectively budget and develop reliable, high-performance, secure, and scalable network infrastructures. Advances in networking and processing require innovative physical layer solutions to support these systems. 10 Gb/s transport rates over Ethernet are a reality, with higher rates in development that likely will be available by 2010. Server consolidation through virtualization, the growing density of network storage systems, and the bigger and multiple workloads allowed by multicore architectures all drive demand for higher-bandwidth network connections.

This paper discusses 10 Gigabit readiness for the physical infrastructure of the data center, in the form of I/O consolidation at the rack level using a Top of Rack (ToR) networking configuration. This architecture leverages 10 Gigabit Ethernet (10 GbE) switches that are equipped with the enhanced Ethernet capabilities to support Fibre Channel over Ethernet (FCoE), an encapsulation of Fibre Channel protocols into Ethernet. The paper then reviews the advanced physical layer infrastructure elements to support the switches, including cabinets, cabling media, and cable and thermal management devices. These infrastructure elements are designed to handle demanding network requirements and must be able to scale over the life of the data center.

Are You Ready for I/O Consolidation?

A key trend with data centers is consolidation. IT organizations are consolidating multiple applications onto a smaller number of more powerful servers using virtualization technology. The benefits range from a reduced need for data center space to reduced power and cooling expenses. Forward-looking enterprises are preparing for the next wave: input-output (I/O) consolidation, which leverages a unified 10 Gb/s fabric at rack level to run LAN, SAN, and interprocess communication (IPC) protocols over the same cabling infrastructure.

Many data centers operate multiple parallel networks: one for IP networking, one for block I/O, and a third for server-to-server IPC protocols used by high-performance computing (HPC) applications. Running multiple networks has a tremendous impact on capital expenditures. Several attempts at merging I/O into one consolidated physical infrastructure have been adopted for some data center applications, such as InfiniBand for cluster and grid computing, but these technologies can be more costly than 10 Gigabit Ethernet.

An enabling consolidation option is 10 Gigabit Ethernet, which brings high performance at a competitive cost to the data center. Extremely fast transmission rates are typically first available over fiber connectivity, and become available over copper as the market for silicon-derived components increases. Today 10 Gb/s transports are available over copper and fiber cabling, and both Ethernet and FCoE standards are ready for the transition to 10 Gb/s transmission speeds.

With 40 Gb/s and 100 Gb/s data rates on the horizon for even faster next-generation networking, it is clear that Ethernet is an effective, high-performance I/O consolidation platform now and well into the future.

©2008 *PANDUIT* Corp. All rights reserved.

Improving Data Center Performance through a Unified I/O Fabric

The goal of I/O consolidation is to create a data center environment that provides anytime, anywhere access to content over a single cabling infrastructure. Enhancements to 10 Gigabit Ethernet represent a significant opportunity to improve data center efficiencies.

Fibre Channel over Ethernet

A set of Ethernet enhancements termed Data Center Ethernet (DCE, also known as Converged Enhanced Ethernet, or CEE) uses the bandwidth of 10 GbE technologies to make Ethernet a unified fabric for I/O consolidation in the data center. Backed by major industry heavyweights including Cisco, Intel, and *PANDUIT*, the features of DCE make the Ethernet approach compelling:

- 10 Gb/s and higher data rate
- Lossless, easy-to-manage transport via flow control and congestion management
- Layer 2 multipathing for unicast and multicast traffics
- DCBCXP (Data Center Ethernet Capabilities Negotiation)
- Enables larger Layer 2 network domains

One primary data center application made possible by DCE is FCoE, a standards-based encapsulation of Fibre Channel onto an Ethernet fabric. The FCoE specification transports native Fibre Channel frames over an Ethernet infrastructure, allowing existing Fibre Channel management modes to stay intact when used over Ethernet networks.

FCoE requires the underlying network fabric to be lossless (i.e., no packet drops). While this can be implemented using the existing Ethernet PAUSE mechanism, DCE features enhance the ability for the network to carry multiple protocols, each with different requirements. Priority Flow Control supports class-of-service based flow control with the ability to implement lossless transmission on a per flow basis. DCE congestion management techniques help to prevent congestion from reaching the network core, and bandwidth management supports traffic shaping that comes into play only when resources become scarce. FCoE is based on the Fibre Channel model, so standard fabric services (including host-to-switch and switch-to-switch behavior) and features (in-order delivery, Fabric Shortest Path First [FSPF] load balancing) remain unchanged.

Consolidated Adapter Technology

A new generation of network adapters function like a separate Ethernet Network Interface Card (NIC) and Fibre Channel Host Bus Adapter (HBA) to the host operating system, transparently encapsulating Fibre Channel alongside IP traffic over the same Ethernet channels before passing both traffic flows to the network. These adapters can use existing operating system drivers, making integration seamless with the use of existing Fibre Channel software and management models.

The FCoE encapsulation process is accomplished by using one of two types of FCoE-capable adapters. The first type of adapter is based on Application-Specific Integrated Circuit (ASIC) technology: ASIC-based adapters called Consolidated Network Adapters (CNAs) are capable of outstanding encapsulation performance in addition to advanced virtualization when used with Cisco Nexus 5000 Series switches. CNAs based on this technology use pluggable transceivers in the form factor of either SFP+ or XFP to support future standards.

©2008 *PANDUIT* Corp. All rights reserved.

Another approach to supporting FCoE is using a standard 10 GbE NIC combined with a driver capable of performing the FCoE encapsulation. The FCoE software stack is based on an OpenFC open source project, and the underlying Ethernet layer is handled by a 10 GbE ASIC such as the Intel 92598-based (Oplin) 10 Gigabit PIC-Express NIC. The FCoE stack remains the same and comprises the OpenFC layer and an FCoE layer, with both of these layers below the Fibre Channel upper layers.

Enhanced Ethernet Technologies Achieve New Physical Layer Efficiencies

Table 1 illustrates how deploying 10 GbE technologies at the rack level helps to simplify the physical layer infrastructure. For example, data center stakeholders might decide to unify their LAN and SAN traffic over Ethernet by deploying a top-of-rack Cisco Nexus 5000 Series switch. Such a deployment would consolidate both LAN and SAN traffic onto a single network link, reducing the number of cables used within the cabinet by a factor of two (see Table 1).

Table 1. Impact of I/O Consolidation on Data Center Infrastructure

Without I/O Consolidation

16 Servers	Ethernet	FC	Total
Adapters*	16	16	32
Switches	2	2	4
Cables	32	32	64

* Using dual port adapters

- Servers have two adapters: one Ethernet dual port NIC and one FC dual-port HBA
- Each adapter uses two cables
- Each server requires four cables
- Each cabinet hosts four switches

With I/O Consolidation

16 Servers	Ethernet	FC	Total
Adapters*	16	0	16
Switches	2	0	2
Cables	32	0	32

* Using dual port adapters

- The FC switches have been eliminated
- The FC HBAs in servers have been eliminated
- The Ethernet NICs in servers have been replaced by FCoE CNAs
- Only two cables are needed to connect a server in a redundant deployment design instead of four
- Two more rack positions are freed up for use by servers

I/O consolidation via DCE and FCoE can also have a direct impact on capital expenditures. FCoE eliminates the need for separate switches, cabling, adapters, and transceivers for each class of traffic, dramatically reducing power consumption and helping reduce both capital and operational expenses.

When used to consolidate the cables within the cabinet, the savings around cabling can be ever greater, because the number of network interface cards within the server is high (see Table 2). Reducing cabling density minimizes the amount of cable management needed and frees up valuable rack space for additional equipment.

With Ethernet transport becoming lossless, I/O convergence becomes a reality: one unified infrastructure supports both Ethernet and I/O traffic.

Table 2. Cabling Reductions Achieved Using FCoE Architecture

Consolidated Connections per Server		Reduced Number of Cables
LAN Connections	SAN Connections	
2	2	50%
4	2	67%
6	2	75%

©2008 *PANDUIT* Corp. All rights reserved.

Top of Rack (ToR) Deployment

I/O consolidation over 10 Gigabit channels brings a new set of requirements and specifications to the data center, and the physical deployment must be planned accordingly. ToR deployments are particularly useful in the data center: these layouts can be used to either reduce the density of cabling in the data center or increase equipment density to achieve a cost-effective, high-performance network infrastructure.

Mapping Physical Layer Design onto ToR Architecture

When designing a ToR deployment it is critical to specify an infrastructure that will support the addition of a switch in each cabinet. Understanding the logical network topology as it relates to the physical room can improve your design significantly.

Figure 1 shows a logical ToR network topology as it correlates to 10 GbE switching technology. This configuration deploys the Cisco Nexus 5000 Switch at the access layer, and can utilize any 10 Gb/s core switch for the aggregation layer (such as the Cisco Nexus 7000 Switch or the Cisco Catalyst 6500 Switch). The Cisco Nexus 5000 Series provides 10 GbE links back to the aggregation layer as well as 10 GbE links to each compute node.

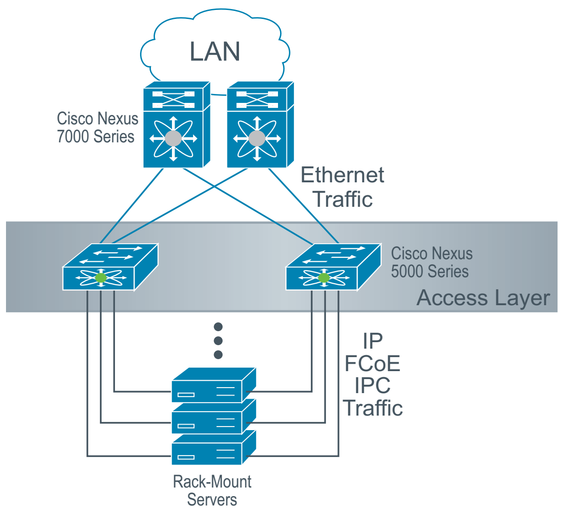

Figure 1. Logical ToR Network Topology

After the logical topology is defined, the physical design can be mapped directly to the model. One way to simplify the design and simultaneously incorporate a scalable layout is to design and divide the physical infrastructure into manageable sub-areas called pods. Each pod within the data center contains access-layer switches that are located within each cabinet.

When choosing a ToR configuration it is important to understand the number of network connections needed for each compute resource. Any connectivity requirement above 48 connections requires an additional access layer switch in each cabinet. For example, if you are deploying 30 1RU servers within a 45RU cabinet and each server requires two SFP+ connections (for a total of 60), an additional Cisco Nexus 5000 access layer switch must be added within the cabinet to support the higher density.

Once the physical design is mapped, the next step is to incorporate the hardware information into the physical configuration. In a ToR layout, the Cisco Nexus 5000 access switch is typically located toward the top of the cabinet. This allows for heavier equipment, such as servers, to occupy the bottom of the cabinet. Dedicated pathways, either underfloor or overhead, are needed to support fiber connectivity for links to the aggregation layer, and horizontal cable managers are used to route short-haul copper links within the cabinets. Cisco Nexus 5000 switches use either the SFP+ Direct Attached 10 Gigabit Copper cabling based on Twinax technology, or SFP+ short-reach optical transceivers and fiber cabling. A typical configuration would use copper media (Twinax) for all connections within racks, and fiber media for longer cable runs between the top-of-rack switches and the aggregation layer.

©2008 *PANDUIT* Corp. All rights reserved.

Fiber and Copper Cabling Capabilities

Investing in the optimal cabling media for your 10-Gigabit-ready data center applications involves striking a balance between factors of bandwidth, flexibility, and scalability (see Table 3). The first 10 GbE switch supporting DCE and FCoE, the Cisco Nexus 5000 Series, currently supports short-reach optical SFP+ transceivers supporting 10GBASE-SR and OM3 fiber, and an integrated SFP+ transceiver and Twinax cable solution.

Fiber. Several grades of high-bandwidth laser-optimized fiber are available for use in high-speed network installations, each with a different reach and data rate capability:

- 62.5/125µm (OM1) fiber, designed to achieve 10 Mb/s and 100 Mb/s data rates, is now largely a legacy fiber
- 50/125µm (OM2) fiber, used to achieve 1 Gb/s data rates
- **50/125µm (OM2+, OM3, and OM3+) fiber**, used to achieve 10 Gb/s data rates; OM2+ and OM3+ fiber grades offer nearly double the bandwidth of their parent fibers ('+' represents extended reach OM2 and OM3 fiber).

Copper. Four types of copper cabling find common use in the data center:

- Category 6 cabling specifies bandwidth up to 250 MHz and is designed to support 1 Gb/s over a 100 m channel
- **Category 6A copper** cabling was developed in conjunction with the 10GBASE-T standard to achieve reliable 10 Gb/s operation over 100 m copper twisted-pair channels. Category 6A shielded and unshielded products are designed to extend usable bandwidth up to 500 MHz and to drastically reduce alien crosstalk interference.
- **1X-based Twinax copper** cabling has been optimized for differential pair applications to support 10 Gb/s signaling. Twinax employs a unique homogeneous construction with 100% shielding that enables completely new levels of data rates over a single line with virtually no crosstalk. Both delay and amplitude skew are minimized, because the one integral dielectric eliminates material variations and forces a nearly identical distance between conductors to be maintained.

Table 3. Fiber and Copper 10 Gb/s Ethernet Cabling Characteristics

Fiber Type	Core/ Cladding (µm)	minEMBc* (MHz•km)	Max. Reach at 10 Gb/s (m)**
OM1 (legacy)	62.5/125	200	33
OM2	50/125	500	82
OM2+***	50/125	950	150
OM3	50/125	2000	300
OM3+***	50/125	4700	550

* Minimum Calculated Effective Modal Bandwidth.

** Data based on use with 850 nm VCSEL-based serial transceivers.

*** OM2+ and OM3+ extended-reach fiber grades are not currently listed in ISO standards.

Copper Type	Diameter (inches)	Bandwidth (Hz)	Max. Channel Length at 10 Gb/s (m)
Category 6	0.22–0.25	250 M*	37*
Category 6A (UTP)	0.29–0.35	500 M	100
Category 6A (Shielded)**	0.29–0.31	500 M	100
Twinax (1X)	0.18–0.24	Up to 11 G	10

* Category 6 cabling may support 10 Gb/s up to 37 m as long as channel bandwidth achieves 500 MHz.

** Provides exceptional suppression of alien crosstalk and EMI/RFI noise.

©2008 *PANDUIT* Corp. All rights reserved.

10 Gigabit Transceivers and SFP+

10 GbE transceivers are available in several formats: XENPAK, X2, XFP and SFP+ (see Table 4). All perform at a rate of 10 Gb/s, with size and form factor varying with their power consumption. XFP and SFP+ are smaller than XENPAK and X2, and consume less power because they remove the Serializing/Deserializing (SERDES) circuitry and XAUI interface that are built into the XENPAK and X2 modules. The XFP embeds clock and data recovery (CDR), where the latter is pulled out from the SFP+ to the system side. Smaller form factor transceivers also achieve higher densities at line cards and switches.

Table 4. 10 Gigabit Transceivers

Form Factor	Electrical Interface	Fiber Connector	Copper Connector	Protocol Agnostic*
XENPAK	XAUI parallel 4 x 3.125 Gb/s	SC	CX4	No
X2	XAUI parallel 4 x 3.125 Gb/s	SC	CX4	No
XFP	XFI serial 1 x 10.3 Gb/s	LC or SC	Direct-attach Twinax	Yes
SFP+	SFI serial 1 x 10.3 Gb/s	LC**	Direct-attach Twinax	Yes

* Cisco XENPAK and X2 currently support only 10 Gigabit Ethernet. XFP and SFP+ do not include a Mux/Demux function, and therefore are protocol agnostic (i.e., can support 10 Gigabit Fibre Channel or 10 Gigabit Ethernet).

** Generally, 10 Gb/s cabling systems are recommended because they are backward compatible with 1 Gb/s in addition to their ability to handle rapid levels of growth in network traffic.

Active equipment that supports 10GBASE-T operating over 100 m of Category 6A UTP cabling is expected to be available on the market near the end of 2008. In the meantime, SFP+ interface-enabled active equipment is expected to be adopted for 10 GbE applications in the data center. The Cisco Nexus 5000 Series switch uses SFP+ modules. The QLogic and Emulex CNAs use SFP+ modules to connect servers to the switch. Intel's first version of its 92598-based NIC will use SFP+ modules.

The competitive advantages of the SFP+ interface are cost and connector backward compatibility: SFP+ modules require a lower number of components, making SFP+ a cheaper alternative to other transceivers. Also, the front connector of the optical form factor remains an LC connector, which enables reuse of the existing fiber infrastructure in the data center.

SFP+ modules are widely available, including a 10 Gigabit optical transceiver available from Cisco. The availability of SFP+ electronics for both fiber and copper links is summarized in Table 5.

Table 5. SFP+ Availability Over Fiber and Copper Ethernet Links

PHY Specification	Availability	Media	Distance
10GBASE-SR	Now	MM fiber (OM3 / OM3+)	300 / 550 m
10GBASE-LR	Q2'CY08	SM fiber	2 km
10GBASE-ER	Q1'CY09	SM fiber	10 km
10G Serial Copper	Q3'CY08	1X-based Twinax	10 m

©2008 *PANDUIT* Corp. All rights reserved.

Thermal Benefits of I/O Consolidation

In many data centers, especially in virtualized environments, servers embed several network adapters to segregate traffic or to increase throughput (via port aggregation). 10 Gb/s adapters can improve server performance by increasing bandwidth and occupying fewer expansion slots, resulting in reduced power requirements and a reduced heat load. Deploying 10 Gb/s adapters also leads to increased design flexibility in the data center: the consolidation of multiple adapters and cables down to one 10 Gb/s adapter and cable reduces infrastructure and frees valuable space within the cabinet.

Removing heat from the data center environment can be a challenge. Movement of cool air through the data center is optimized through strategic layout of air conditioning units and physical layer elements. In a conventional raised-floor environment, cool air is distributed underneath the raised floor, and the air enters the room through perforated floor tiles to cool the active equipment. Data center servers and switches are positioned to face the cold aisles where perforated floor tiles are located. This allows cool air to be transferred through racks and cabinets by equipment fans and to be released as exhaust into hot aisles to the rear.

Unused space in server and switching enclosures has the potential to cause two problems—hot exhaust recirculation and cool air bypass. However, the strategic deployment of rack, cabinet, and cable management systems can have a considerable impact on data center cooling efficiency. Data center stakeholders should specify cabinets that optimize thermal performance and facilitate airflow in high-density environments through the use of passive cooling systems; these systems leverage active equipment exhaust fans to move air through the cabinet efficiently to minimize risks from overheating.

Filler panels can be deployed in horizontal and vertical spaces to prevent cold air bypass and recirculation of warm air through the cabinet and equipment. Use of wider cabinets and effective cable management can reduce static pressure at the rear of the cabinet by keeping the server exhaust area free from obstruction. Also, door perforations should be optimized for maximum cool airflow to equipment intakes. Deployment of racks and cabinets should follow standards established in TIA/EIA-310-D.

Conclusion

10 Gigabit Ethernet and its enhancements bring many performance and scalability benefits to the data center environment. The deployment of a 10 Gb/s physical infrastructure eases current bandwidth concerns and enables I/O consolidation of LAN and SAN traffic; FCoE is one feasible application of 10 Gigabit Ethernet used as a lossless transport.

The Cisco Nexus 5000 Series switch is the first switch that brings a unified 10 GbE fabric to the data center for LAN and SAN. PANDUIT cabling, rack, and cabinet systems provide the most reliable and cost-effective 10 Gb/s infrastructure solutions to support data center active equipment. When used together, Cisco Nexus Series switches and modular, scalable, and 100% tested PANDUIT cabling systems work together to enable business agility and yield best-in-class performance within your data center.

For more information on how to consolidate your physical infrastructure, download the Data Center Design Guide “Mapping Cisco Nexus, Catalyst, and MDS Logical Architectures into PANDUIT Physical Layer Infrastructure Solutions” located at http://www.panduit.com/DataCenterDesignGuides.

©2008 *PANDUIT* Corp. All rights reserved.

About *PANDUIT*

PANDUIT is a leading, world-class developer and provider of innovative networking and electrical solutions. For more than 50 years, *PANDUIT* has engineered and manufactured end-to-end solutions that assist our customers in the deployment of the latest technologies. Our global staff and expertise and strong industry relationships make *PANDUIT* a valuable and trusted partner dedicated to delivering technology-driven solutions and unmatched service. Through our commitment to innovation, quality and service, *PANDUIT* creates competitive advantages to earn customer preference.

www.panduit.com · cs@panduit.com · 800-777-3300

Cisco, Cisco Systems, Nexus, Catalyst, and MDS are registered trademarks of Cisco Technology, Inc.

©2008 *PANDUIT* Corp. All rights reserved.