Datacenter Networking Convergence –
Trends and Directions

Bob Nusbaum
Cisco Mainframe Solutions

March 15, 2012
Session Number 8196
Agenda

- What’s the Problem?
- Technology and Standards for Convergence
- Solution Components
- Key Take-aways
Data Center Application Architecture

The Data Center Evolution

<table>
<thead>
<tr>
<th>Phase 1</th>
<th>Phase 2</th>
<th>Phase 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mainframe</td>
<td>Client-Server and Distributed Computing</td>
<td>Service Oriented and Web 2.0 Based</td>
</tr>
<tr>
<td>Centralized</td>
<td>Decentralized</td>
<td>Virtualized</td>
</tr>
</tbody>
</table>

- What's driving the need to evolve Data Center Application Architecture?

- The Data Center Evolution:
  - Phase 1: Mainframe
  - Phase 2: Client-Server and Distributed Computing
  - Phase 3: Service Oriented and Web 2.0 Based

Network importance

IT Relevance and Control
Challenges in Today’s Data Centers

Customer Requirements:

- Reduce:
  - Power and Cooling Requirement
  - Cables
  - Architecture Complexity
  - Administrators

- Increase:
  - Speed
  - Utilization
  - Agility
  - Application Roll-out
  - Speed & Efficiency
Scale of Current Problems Demands a New Approach

**Infrastructure Scalability**
- By 2009, 50% of large businesses will spend more on power and cooling then on new servers (Gartner, 2006)
- Quad-cores and octal-cores will drive significantly more traffic
- Storage is expected to continue to grow at a 40-70% CAGR (Gartner, 2006)

**Transport Flexibility**
- Continued deconstruction of the server increases demands on the network
- Market transitions between transport technologies and application architectures

**Operational Continuity**
- Expectation of 24x7 application availability
- 54% of network downtime is caused by human error (Uptime Institute, 2007)
Data Center Infrastructure
Operations & Maintenance Now ~80% of IT Budgets and Growing

Virtualization will only make things worse

Source: IDC 2009
FICON: A Precursor to Convergence

- Simpler network
- Reduction in cost per port
- Asset utilization – 1.4 x performance (4:1 consolidation)
  - Cost of FICON features very small when performance enhancement is considered
- Application effectiveness
  - Up to 100% improvement
- Additional throughput
  - Roughly 6x bandwidth of ESCON
  - Shortens file maintenance windows by roughly 25%
- Competitively positions company for future
  - ESCON will have no new development
  - ESCON will not be supported on future System z models
10 Gigabit Ethernet to the Server
Impacting DC access layer cabling architecture

- Multicore CPU architectures
- **Virtual Machines** driving Increased I/O bandwidth per server
  - Increased business agility
- Increased network bandwidth demands
- Consolidation of networks
  - Segmentation & Unified Fabrics / UIO

Future Proofing - Network, Cable Plant
The Case for Converged Fabric

Reduce overall Data Center power consumption by up to 8%. Extend the lifecycle of current data center.

Wire hosts once to connect to any network - SAN, LAN, HPC. Faster rollout of new apps and services.

Every host is SAN-enabled. Drive storage consolidation and improve utilization.

Rack, Row, and X-Data Center VM portability become possible.
Increased Efficiency, Simpler Operations

Converged Fabric

Mgmt Network

Front-End Network

Backup Network

Storage Network

Back-End Network

Converged Fabric and I/O
Converged Fabric Overview
Fibre Channel over Ethernet (FCoE)

**FCoE**
- Mapping of FC Frames over Ethernet
- Enables FC to Run on a Lossless Ethernet Network

**Benefits**
- Fewer Cables
  - Both block I/O & Ethernet traffic co-exist on same cable
- Fewer adapters needed
- Overall less power
- Interoperates with existing SAN’s
  - Management SAN’s remains constant
- No Gateway
Converged Fabric
Why?

- Fewer CNAs (Converged Network adapters) instead of NICs, HBAs and HCAs
- Limited number of interfaces for Blade Servers / Rack Mounted Servers

All traffic goes over 10GE
What is Fibre Channel over Ethernet?

- From a Fibre Channel standpoint it’s
  FC connectivity over a new type of cable called… an Ethernet cloud

- From an Ethernet standpoints it’s
  Yet another ULP (Upper Layer Protocol) to be transported, but… a challenging one!

- And technically…

FCoE allows the extension of Fibre Channel onto a Lossless Ethernet fabric
FCoE Enablers

- 10Gbps Ethernet
- Lossless Ethernet
  - Matches the lossless behavior guaranteed in FC by B2B credits
- Ethernet jumbo frames
  - Max FC frame payload = 2112 bytes

Normal Ethernet frame, ethertype = FCoE

Same as a physical FC frame

Control information: version, ordered sets (SOF, EOF)
FCoE Advantages

FCoE is managed like FC at initiator (CH), target (CU), and switch level

- Completely based on the FC model
- Same host-to-switch and switch-to-switch behavior of FC
- In order delivery or FSPF load balancing
- WWNs, FC-IDs, hard/soft zoning, Name server, RSCN

FCoE is Fibre Channel

Standard Approved June 4th 2009
FCoE versus 8Gig Fibre Channel Comparisons - Test results

FCoE technology is capable of providing up to 250,000 IOPs

As 8G FC reaches its throughput limits; FCoE can still provide additional bandwidth, up to 25% for larger block sizes.
The FC-BB-5, FCoE, draft standard was unanimously approved as the final standard - June 4th 2009

1) FC-BB_IP: The FC-BB_IP model defines the means by which Fibre Channel networks interface with and connect across an IP network.

1) FC-BB_GFPT: The FC-BB_GFPT model defines the means by which FC physical links may be extended over any WAN. Transport infrastructure for which GFP mapping is defined.

2) FC-BB_PW: The FC-BB_PW model defines the means by which FC physical links may be extended over a wide area MPLS network.

3) FC-BB_E: The FC-BB_E model defines the means by which Fibre Channel frames are transported over a Lossless Ethernet network.
# Nexus 5000 Enhanced Ethernet Fabric Features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Priority Flow Control</strong></td>
<td>Provides class of service flow control by enabling PAUSE functionality on IEEE 802.1p lanes</td>
</tr>
<tr>
<td><strong>IEEE 802.1Qbb</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Data Center Bridging Exchange</strong></td>
<td>Auto-negotiation of Enhanced Ethernet capabilities DCBX (switch to NIC)</td>
</tr>
<tr>
<td><strong>IEEE 802.1AB</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Bandwidth Management</strong></td>
<td>Enhanced Transmission Selection - manage bandwidth and assign priorities to groups of IEEE 802.1p lanes based on class of traffic</td>
</tr>
<tr>
<td><strong>IEEE 802.1Qaz</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Congestion Management</strong></td>
<td>This standard specifies protocols, procedures and managed objects that support congestion management of long-lived data flows within network domains of limited bandwidth delay product (BCN/QCN)</td>
</tr>
<tr>
<td><strong>IEEE 802.1Qau</strong></td>
<td></td>
</tr>
<tr>
<td><strong>L2 Multipathing</strong></td>
<td>Layer-2 multipathing eliminates standby uplinks increasing the available uplink bandwidth. In the NX5K this is enabled by Ethernet Host Virtualizer Mode.</td>
</tr>
<tr>
<td><strong>“IETF TRILL”</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Transparent Interconnect of Lots of Links</strong></td>
<td></td>
</tr>
</tbody>
</table>
Data Center Ethernet: PFC & Bandwidth Management

**Priority Flow Control**

- Enables lossless behavior for each class of service
- PAUSE sent per virtual lane when buffers limit exceeded

**CoS based Bandwidth Management**

- Enables Intelligent sharing of bandwidth between traffic classes
- Control of bandwidth
- 802.1Qaz Enhanced Transmission

**10 GE Realized Traffic Utilization**

- 3G/s
- HPC Traffic 3G/s
- 2G/s
- 3G/s
- Storage Traffic 3G/s
- 3G/s
- LAN Traffic 4G/s
- 5G/s

**Offered Traffic**

- 3G/s
- 4G/s
- 6G/s

- t1
- t2
- t3
Priority Flow Control
Fibre Channel over Ethernet Flow Control

- Enables lossless Ethernet using PAUSE based on a COS as defined in 802.1p
- When link is congested, CoS assigned to FCoE will be PAUSEd so traffic will not be dropped
- Other traffic assigned to other CoS will continue to transmit and rely on upper layer protocols for retransmission
Enhanced Transmission Standard

Bandwidth Management -- IEEE 802.1Qaz

- Required when consolidating I/O – It’s a QoS problem
- Prevents a single traffic class of “hogging” all the bandwidth and starving other classes
- When a given load doesn’t fully utilize its allocated bandwidth, it is available to other classes
- Helps accommodate for classes of a “burtsy” nature
Will FCoE Work for FICON?

Key capabilities are there:

- Class 2
- Loop Back
- Credit Pooling

Performance and Error Statistics

- Number of Words Transmitted
- Number of Words Received
- Number of Frames Transmitted
- Number of Frames Received
- Number of Frames Discarded
- Time that port is unable to receive frames from xN_Port due to zero credit
- Time that port is unable to send frames to xN_Port due to zero credit
- General error statistics appropriate to the port protocol:
  - Number of Code Violation Errors
  - Number of CRC/checksum errors
  - Number of times Loss of Sync was detected
  - Number of Link Failures

Standards work is in process in T11 for FICON over DCB
What are the necessary pieces?
IBM System X Converged I/O Adapters and Blade Center H and HT
Hardware Converged Network Adapter Solution

- 10 Gbps maximum throughput for high-bandwidth storage (SAN) and networking (LAN) traffic
  - Dual 10GE/FCoE ports
- Support for native drivers and utilities
  - Customer certified stacks
- Replaces multiple adapters per server/blade
- Consolidates 10GE and FC on a single interface
- Blade Center Support for Blade Open Fabric Manager for BIOS, UEFI, and Fcode
CNA Technology Rapidly Evolving

First generation
Mid-2008

Second generation
Mid-2009

Near future

• Multiple components
• Full height/length
• Single chip
• Half height/length
• Less than half the power
• Same support as HBAs

And other NIC vendors
View from Operating System

- CNAs are multi-function devices
- Standard drivers
- Same management
- Operating System sees:
  - Dual port 10 Gigabit Ethernet adapter
  - Dual Port Fibre Channel HBAs
Cost Effective 10G Server Connectivity Today

SFP+ USR – ‘Ultra Short Reach’
- 100M on OM3 fiber, 30M on OM2 fiber
- Support on all Cisco Catalyst and Nexus switches

SFP+ Direct Attach
- 1, 3, 5 and 7M and 10M on Twinax active cables
- Support across all Nexus Switches
## Evolution of Ethernet Physical Media

### Role of Transport in Enabling 10GE Technology

<table>
<thead>
<tr>
<th>Technology</th>
<th>Distance</th>
<th>Power (each side)</th>
<th>Transceiver Latency</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFP+ Cu Copper</td>
<td>Twinax 5m</td>
<td>~0.1W</td>
<td>~0.2μs</td>
</tr>
<tr>
<td>SFP+ USR</td>
<td>MM OM2 30m</td>
<td>1W</td>
<td>~0.1μs</td>
</tr>
<tr>
<td></td>
<td>MM OM3 100m</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MM OM2 30m</td>
<td>1W</td>
<td>~0.1μs</td>
</tr>
<tr>
<td></td>
<td>MM OM3 100m</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10GBASE-T</td>
<td>Cat6a/7 100m</td>
<td>~8W</td>
<td>2.5μs</td>
</tr>
<tr>
<td></td>
<td>Cat6a/7 30m</td>
<td>~4W</td>
<td></td>
</tr>
</tbody>
</table>

- **Twinax**: ~0.25μs, ~0.1W, 5m SFP+ CU Copper
- **MM OM1**: ~0.1μs, 33m SFP+ SR
- **Cat6a/7**: ~0.1μs, ~8W, 30m 10GBASE-T

- **SFP+ Cu to SFP+**: Low cost, low power and latency, up to 5m in rack and adjacent cabling.
Unified Fabric
What is it? Really?
Critical Infrastructure for Data Center Networking

<table>
<thead>
<tr>
<th>Converged Fabric and I/O Interfaces</th>
<th>Enables storage, mainframe, Ethernet, IP, and HPC traffic to converge to a single network</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lowers overall data center power draw</td>
</tr>
<tr>
<td><strong>Cisco® Nexus Switching Platforms</strong></td>
<td>Multi-Terabit platform with ops-centric design</td>
</tr>
<tr>
<td></td>
<td>Delivers unified fabric and I/O deployments</td>
</tr>
<tr>
<td></td>
<td>Designed for the most stringent availability needs</td>
</tr>
<tr>
<td><strong>NX-OS Operating System</strong></td>
<td>First multi-protocol DC-class operating system</td>
</tr>
<tr>
<td></td>
<td>Virtualized control plane and scalable design sets new standard for usability</td>
</tr>
<tr>
<td><strong>Cisco MDS Storage Switching</strong></td>
<td>The industry-leading features and capabilities of the Cisco MDS 9000 Family provide superior performance and efficiency for deploying Unified I/O in Data Center Networking</td>
</tr>
</tbody>
</table>
Selecting the Unified Fabric Mix

Add MDS 9000 for
- New or existing servers with HBAs
- New or existing FC or FCoE-only storage ports
- Replacing DCX, SilkWorm or McDATA switches
- SAN Extension over MAN or WAN links
- Advanced storage services (migration, encryption, etc.)

Add Nexus 7000 for
- SAN core for multi-protocol Ethernet storage devices:
  - FCoE and
  - iSCSI / NAS
- Highly available access layer for mission critical servers with CNAs
- Aggregation layer for fully converged fabrics

Add Nexus 5000 for
- Cost-effective server access using CNAs
- Small Unified Fabrics containing FC and FCoE (SAN & LAN in one box)
- Server access where the mix of HBAs and CNAs will change over time
<table>
<thead>
<tr>
<th>Attribute</th>
<th>MDS 9500</th>
<th>Nexus 7000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physical Redundancy</td>
<td>Supervisors, Power Supplies, Fans</td>
<td></td>
</tr>
<tr>
<td>Logical Redundancy</td>
<td>Port Channels, Load Balancing</td>
<td></td>
</tr>
<tr>
<td>Non Disruptive Upgrades</td>
<td>Full ISSU</td>
<td></td>
</tr>
<tr>
<td>Segmentation/Isolation</td>
<td>VSANs, VDCs</td>
<td></td>
</tr>
<tr>
<td>Maximum Scale</td>
<td>528 ports</td>
<td>768 ports</td>
</tr>
<tr>
<td>Performance</td>
<td>256-Gbps/Slot</td>
<td>550-Gbps/Slot</td>
</tr>
<tr>
<td>FCoE modules</td>
<td>8-port</td>
<td>32-port ➔ 48</td>
</tr>
</tbody>
</table>
8-Port 10G FCoE Module for MDS 9500

Enables integration of existing FC infrastructure into Unified Fabric

- 8 FCoE ports at 10GE full rate
- 80-Gbps front panel bandwidth
- SFP+ SR, LR, CX-1 optics support

FCoE connectivity from MDS 9500 Directors to:

- Nexus 5000 and Nexus 7000
- FCoE Storage Arrays
Cisco Nexus 5000
Now available through IBM!
Unified Ports
Dynamic and Efficient Port Allocation

**Availability**
- 16-port Expansion Module on the Nexus 5548 and 5596
- All Ports on the Nexus 5596

**Benefits**
- Simplify switch purchase - remove ports ratio guess work
- Increase design flexibility
- Remove specific protocol bandwidth bottlenecks

**Use-cases**
- Flexible LAN & storage convergence based on business needs
- Service can be adjusted based on the demand for specific traffic

Lossless Ethernet – FCoE, iSCSI, NAS
Summary & Take-Aways

- Fully standardized
- Major cost savings
- Products are there
- Deploy now for distributed systems
- FICON – “watch this space”

The bottom line....
“It’s One Network!”
Thank you.