Cisco Server Fabric Switches

InfiniBand on Wall Street

Cisco SVBU - Product Management

26 March
Take Aways
With Ongoing Scalability Testing

1. Cisco has **nine month lead time** with large IB clusters
   - Sandia Thunderbird 4400 node cluster
   - One of three >1000 node IB clusters Cisco

2. Cisco is **only vendor with access to large IB cluster** today
   - Daily remote and direct access to 4400 node cluster
   - Enables continuous improvement of HPC Subnet Manager
   - Scalability testing with LUSTRE and IB attached storage
     - Cisco has ~1000 InfiniBand nodes inside it’s own test labs
       - Cisco’s developers in US and Asia have access around the clock
     - Cisco is **directly funding the open source community** to increase stability of large scale cluster
       - Two grants already made in 2006 for MVAPICH development

1. Cisco is already developing unified management across large scale InfiniBand and Ethernet connected clusters
Experience & Expertise

• Largest performance based cluster – 4600 servers; 1 Year in production
  Sandia National Labs
  \textit{Validated scalability of all components – host-based subnet manager, host drivers, Switch OS, Hardware}

• Largest ‘compute grid’ deployment – 900+ servers; 1+ years in production
  Wall Street Investment Bank
  14 Credit Risk applications from 8 lines of business

• Service Provider, Australia – 100+ servers; 3 months in production
Why Cisco?

Cisco Offers End-to-End DC Networking Solution

IP Network | Storage Network | InfiniBand | Security | Application Optimization | Metro Optical | Management
---|---|---|---|---|---|---
Tape | FC SAN | RAID | IDS | Cache | Content Switch | Operations
Cisco MDS 9500 | RAID | FC SAN | Internet | Metro Network | DWDM/SONET/Et | InfiniBand

CISCO Foundry
CISCO Brocade
CISCO Voltaire
CISCO Juniper
CISCO Checkpoint
CISCO F5
CISCO Nortel
CISCO ??

NO OTHER COMPANY CAN OFFER COMPREHENSIVE END-TO-END DATA CENTER NETWORKING SOLUTION
Value of Comprehensive Blade IO Portfolio

Unique Cisco Value Proposition

Simplified Management

Ethernet

Operational Ease of Use

- IOS CLI or IOS like CLI
- CiscoWorks (Ethernet & IB)
- Fabric Manager (FC)
- Subnet Manager (IB)

Intelligent Networking

Fiber Channel

Focus on

- Technology Innovation
- Resiliency
- Security
- End-to-end solutions

Proven Support & Services

InfiniBand

Focus on Customers

- World class support & service (TAC)
- Inter-operability certification
- Extensive sales coverage

Focus on Customers

- World class support & service (TAC)
- Inter-operability certification
- Extensive sales coverage

Focus on

- Technology Innovation
- Resiliency
- Security
- End-to-end solutions

Simplified Management

Ethernet

Operational Ease of Use

- IOS CLI or IOS like CLI
- CiscoWorks (Ethernet & IB)
- Fabric Manager (FC)
- Subnet Manager (IB)

Intelligent Networking

Fiber Channel

Focus on

- Technology Innovation
- Resiliency
- Security
- End-to-end solutions

Proven Support & Services

InfiniBand

Focus on Customers

- World class support & service (TAC)
- Inter-operability certification
- Extensive sales coverage
Server Networking Challenges

Growing Performance Demands

- Highly distributed X86 architectures with increased bus speeds (PCI Express)
- Increased messaging between servers
- Direct relationship of server to server communication latencies and performance
- Increased server utilization with server hypervisor and utility grid technologies
- Multi-CPU’s and multi-core on exponential supply curves

Growing Network Complexities

- Average server has 4-8 network interfaces including NICs, HBA’s and HCA’s
- Complex wiring with increased services requirements
- Blurring of server and networking divisions within blade chassis
Key Value Propositions

• **Performance – Top-line benefit**
  - Latency, Throughput, Message Rate, CPU offload
  - Ecosystem: TIBCO, 29West, Wombat, Reuters
  - No app recode, including for Java applications

• **Compute Grid – Bottom-line benefit**
  - Flexibility, Grid Applications, ‘Scale-out’
  - Ecosystem: Platform, DataSynapse, Gemstone, Gigaspaces
Key Value Propositions

• Performance – Top-line benefit
  - Latency, Throughput, Message Rate, CPU offload

• Compute Grid – Bottom-line benefit
  - Flexibility, MFIO, Grid Applications, ‘Scale-out’
Why Performance?

• Latency – *The less the better*
  
  With increased algorithmic trading, lowered latency means more profitable trades

• Throughput – *Need to keep pace and plan for future*
  
  Increased number of trades, smaller trade sizes, smaller spreads, newer instruments are some of the factors increasing the amount of market data and messaging rates from the exchanges

• Computation capability – More sophistication yields better business results
  
  Ability to crunch increasingly sophisticated models yields better pricing, better risk analysis, better hedging strategies and better working capital management

*High performance is a competitive advantage for firms!*
STAC performs vendor-independent, hands-on research into the latest technology stacks for the capital markets industry.

Cisco and Wombat engaged STAC to optimize and benchmark Wombat software on both InfiniBand and Ethernet switches from Cisco.

These slides highlight some of the key findings. Details are in a forthcoming report.

For more information about STAC, see:

www.STACresearch.com
Key findings

• Cisco IB reduces mean latency
  By up to 63%
  Reduces latency for a range of update sizes

• Cisco IB reduces disperson
  Std dev drops by about 34%
  Radical reduction in outliers caused by request bursts
  Chops 30-35 milliseconds off spikes
Wombat Testing Results

- 2/3\textsuperscript{rd}s reduction in latency and spread of latency
- Units are in microseconds
- Average of 6, 30 min runs
  - 4 Clients for concurrent access
  - 216M datapoints per run

<table>
<thead>
<tr>
<th></th>
<th>GigE</th>
<th>Cisco SFS</th>
<th>Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Latency (usec)</td>
<td>217</td>
<td>75</td>
<td>65%</td>
</tr>
<tr>
<td>Variance (usec)</td>
<td>135</td>
<td>61</td>
<td>54%</td>
</tr>
<tr>
<td>Average of Max - Min (usec)</td>
<td>7718</td>
<td>2318</td>
<td>70%</td>
</tr>
</tbody>
</table>
TIBCO-EMS Benchmarks (Consumer)

**TIBCO-EMS Performance - Consumer**

- **Message Rate (Msg/Sec)** vs **Message Size (Bytes)**
- Lines representing different protocols:
  - GigE
  - IPoIB
  - SDP
  - **4X Improvement!**

The graph shows a significant improvement in message rate for GigE compared to other protocols as the message size increases.
Cisco InfiniBand Hardware

New

IBM Bladecenter H InfiniBand 4X Switch & HCA

SFS-3012P

SFS-3012

SFS-7024P

4X DDR PCI-Ex HCA

SFS-7012P

SFS-7008P

SFS-7000P

New

New

IBM Bladecenter InfiniBand 1X Switch & HCA

SFS-3012

SFS-3012P

New

New

New

Cisco High Performance Subnet Manager

CiscoWorks - LAN Management System

Device Fault Manager

Resource Manager Essentials

New

SFS-7000D Family

New

Cisco PCI-X & PCI-Ex 2*4X HCA

Cisco PCI-X & PCI-Ex 1*4X HCA

SFS-3012

HP C-Class integration

CiscoWorks - LAN Management System

Device Fault Manager

Resource Manager Essentials

New
The Cisco SFS Product Line: InfiniBand for High Performance Computing

### Server Fabric Switch

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SFS 7000D</td>
<td>(24) 4XIB</td>
</tr>
<tr>
<td>SFS 3001</td>
<td>(12) 4XIB + 1 Gw</td>
</tr>
<tr>
<td>SFS 3012</td>
<td>(24) 4XIB + 12 Gws</td>
</tr>
<tr>
<td>SFS 7008P</td>
<td>(96) 4XIB</td>
</tr>
<tr>
<td>SFS 7012D</td>
<td>(144) 4XIB</td>
</tr>
<tr>
<td>SFS 7024D</td>
<td>(288) 4XIB</td>
</tr>
</tbody>
</table>

### Blade Server

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
</table>
| IBM BladeCenter & BC-H | Switch & HCA  
| Dell 1855   | HCA (2) 4XIB PCI-ex  
| HP C-Class  | Relationship WIP  

### HCA

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SDR &amp; DDR HCA’s</td>
<td></td>
</tr>
<tr>
<td>With/without mem</td>
<td></td>
</tr>
<tr>
<td>Various form factors</td>
<td></td>
</tr>
</tbody>
</table>

### Subnet Mgt

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Performance Subnet Manager Software</td>
<td></td>
</tr>
<tr>
<td>Embedded Subnet Manager</td>
<td></td>
</tr>
</tbody>
</table>

### Wire

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 – 28 AWG Std. IB CX4</td>
<td></td>
</tr>
<tr>
<td>SDR 1, 3, 5, 7, 10, and 15 meters</td>
<td></td>
</tr>
<tr>
<td>DDR up to 8m</td>
<td></td>
</tr>
</tbody>
</table>

**Cisco Confidential**
SFS Family - Broad Portfolio of Products

- InfiniBand Switches – SFS7000
  24 to 288 Ports; SDR & DDR

- Blade Server Products

- IO Gateways – SFS3000

- HCAs and Drivers
  - PCI-X and PCIe
  - Linux, Windows, Solaris & HPUX
driver support
  (Commercial and Open Source)

- And, of course, Ethernet and Fibre-Channel based products
I/O Consolidation & Virtualization

- Single redundant fabric for:
  - IPC
  - Client Connectivity
  - Storage connectivity

- VFrame
  - Dynamic resource mapping interface
  - Policy based resource loading
  - Availability monitoring
  - Partitioning manager

- Cisco MDS 9000 Series
- IO Gateway
- Catalyst 6500 Series

- Fibre Channel to MFIO switch for storage access
- Ethernet to switch for client and NAS access
Cisco SFS 3000 Ethernet Gateway

- IB-to-Ethernet gateway
- Six 100/1000 Ethernet ports
- Custom silicon yields wire-speed performance
- 12 Gbps throughput full duplex—10.5 million pps throughput
- Robust Cisco® Ethernet features for full Cisco certification
  - Topology transparency for transparent Layer 2 bridging
  - Cisco EtherChannel® link aggregation between ports
    - Active/passive and active/active dynamic load balancing across multiple gateways/switches
- Uses all available bandwidth efficiently and protects against failures
Cisco SFS 3000 Fibre Channel Gateway

- IB-to-Fibre Channel gateway
- Two 2-Gbps Fibre Channel ports
- Scale up I/O dynamically by hot plugging new cards
- Proven hardware interoperability
  - EMC RPQ certification
  - IBM storage proven
  - And many more
- Transparent topology supports
  - Zoning and LUN-based access controls with unique WWN per host
  - Multipathing and load balancing (e.g., EMC Powerpath, DMP, etc.)
- Gateway and port-level trunking/failover
Oracle 10g: Scope of IB Benefits

Intra RAC: IPC over IPoIB
- FC gateway: host/lun mapping

OracleNet: SDP over IB

Ethernet Gateway

Network

SAN

Network

Application Servers

Oracle 10g RAC

Consolidate and share Storage amongst servers

20% improvement in throughput
2x improvement in throughput and 45% less CPU
30% improvement in DB performance

RDS For Oracle – Coming Soon!
SFS Deployments – Typical Architecture

• High-performance, high-flexibility, high-availability environments
  - No single points of failure
  - Scale IO dynamically
  - Grow cluster as applications move onto the grid
  - Entire rack can be staged off-site and rolled in
Key Cisco Differentiators

• Market leadership in both IB and Ethernet
• Only company to offer integrated IB + Ethernet

Unified Compute Fabric
• Industry’s first subnet manager to support multivendor environments

• Most scalable clusters and best practices
• Enterprise-class resiliency/HA, serviceability and security for best cluster efficiency.

• Cisco drives OpenFabrics and OpenMPI.
## SFS High Availability Fabric Comparisons

<table>
<thead>
<tr>
<th>Features</th>
<th>Cisco</th>
<th>Competition</th>
<th>Competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fully HA Subnet Manager</td>
<td>Yes, active db-synch</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Fully HA IP Gateways</td>
<td>Yes, Active/active load balancing between gateways and switches; port merging between ports and cards</td>
<td>No, doesn’t support IPoIB, forces proprietary VNIC</td>
<td>Limited failover between chassis, especially for multicast applications</td>
</tr>
<tr>
<td>Support for storage and SM multipathing</td>
<td>Yes</td>
<td>No, ASIC lacks talk-through model</td>
<td>No, ASIC lacks talk-through model</td>
</tr>
<tr>
<td>Redundant TACACS and RADIUS Servers</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Redundant SNMP, NTP, and DNS servers</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Rapid Service Architecture with passive midplanes</td>
<td>Yes (96 port switches)</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
## SFS High Performance Subnet Manager

<table>
<thead>
<tr>
<th>Feature</th>
<th>Cisco</th>
<th>Competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leverage high performance servers</td>
<td>Yes, stand-alone</td>
<td>No, embedded processor only</td>
</tr>
<tr>
<td>Supports multi-vendor Infiniband networks</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Proven to scale &gt;1000 nodes in production</td>
<td>Yes (Discovers 4700 nodes at Sandia in &lt;60 seconds)</td>
<td>No</td>
</tr>
<tr>
<td>No single point of failure</td>
<td>Yes – active db-synch between active/standby.</td>
<td>No – failure forces resweep and failure in traffic</td>
</tr>
<tr>
<td>Integrated Performance Management</td>
<td>Yes – Performance thresholds issues SNMP alerts</td>
<td>No</td>
</tr>
<tr>
<td>Diagnoses problems in-band and out-of-band</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Multi-path/Multi-LID Support</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
## SFS Enterprise Security Features

<table>
<thead>
<tr>
<th>Features</th>
<th>Cisco</th>
<th>Competition</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSH, SNMP v3</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>SSL</td>
<td>Yes</td>
<td>No, HTTP only (SS)</td>
</tr>
<tr>
<td>Secure ACS Integration</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>RADIUS</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>TACACS</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>SCP</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Roles-based Access Controls (Authorization)</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Centralized configuration audits</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>
Why Cisco for Enterprise Grids?

- High Availability & Scalability with Cisco InfiniBand Drivers
  - Cisco InfiniBand Drivers provide “true HA” for IPoIB & SRP
  - Scalable Sockets Direct Protocol (1000’s of connections)
  - Multi-pathing for storage
  - Full Linux & Windows Support (AIX, Solaris, HP-UX also available)

- Storage Certifications
  - EMC RPQ Certification
  - IBM Storage Proven
  - HDS Certifications

- Cisco Service & Support
  - Customer support centers in each geography / timezone
  - Cluster Integration services available worldwide through Cisco Advanced Services
Unified Management – Ethernet and InfiniBand

• Leverage Cisco’s expertise in building mission critical networks
• Leverage customer expertise in managing Cisco’s products
• Integrating SFS products into Cisco’s Network Management Suite
  CLI is Cisco-like
  Resource Manager Essentials
  Device Fault Manager
Global Support & Advanced Services

• 4 Global TACs for around the clock coverage
  San Jose, RTP, Brussels, Sydney

• 900+ parts depots in 120 countries
  On-site spares, 2-hour replacement to next business day

• Advanced Services
  Expertise in bringing up and validating clusters
  Documented best practices

• Global Training
  Basic and advanced classes – can be hosted globally
Summary & Contact Information

• Cisco SFS Products provide:
  Top-line benefits with performance
  Bottom-line benefits with compute grid

• InfiniBand is ready for Wall Street
  Production proven in mission critical environments
  High availability, secure, manageable, proven scalability

• Renaud Larsen
  SBDM - Europe
  renaudl@cisco.com