

Case for Enhancing Ethernet Capabilities for Backplane Fabric Interconnects*

Presentation for IEEE Backplane Ethernet Study Group Meeting
March 2004

Gopal Hegde
Networking and Storage Group
Intel Corporation



* Backplane Fabric Interconnects = Switches, two end points and the backplane trace



Outline

- Key Messages
- Blade Market Trends and Emerging usage models
- Application characterizations for emerging usage models
- Emerging next generation architectures
- Possible Ethernet enhancements
- Conclusion



Key Messages

- The following market trends are requiring blades to better handle storage and Inter process communication traffic
 - Increased deployment in Back End and Mid Tier application
 - Suitability of blades for use in clusters built with commodity components
- Standard Ethernet does not meet all the needs of storage, Inter processor communication (IPC) traffic and communication workloads
- Users are evaluating various fabrics for future storage and cluster interconnect
- Enhance Ethernet capabilities for next generation blades
 - Such platforms would offer Ethernet price/performance advantages
- BESG needs to evaluate enhancements to improve Ethernet latency and congestion control capabilities

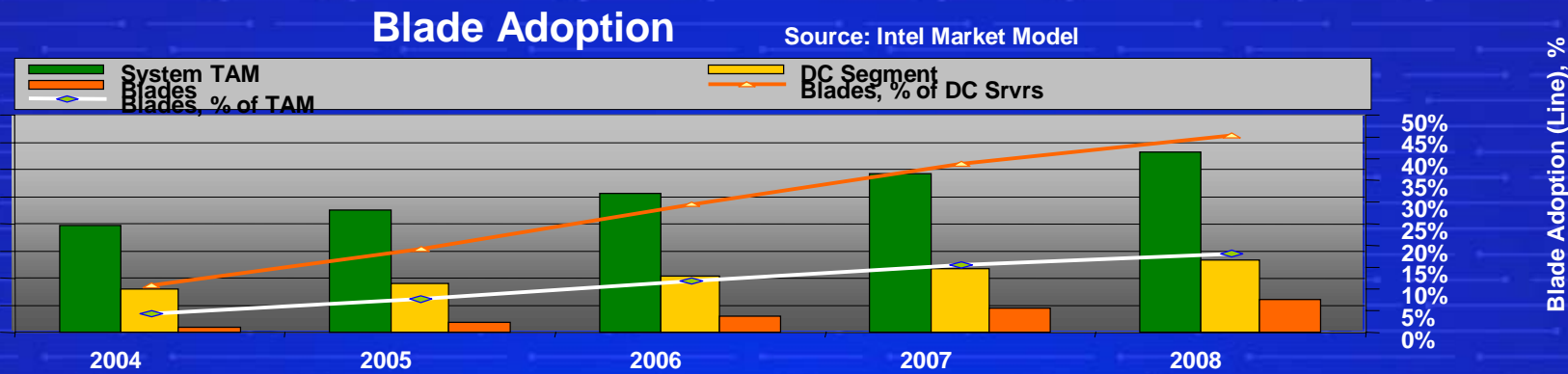


Outline

- Key Messages
- Blade Market Trends and Emerging usage models
- Application characterizations for emerging usage models
- Emerging next generation architectures
- Possible Ethernet enhancements
- Conclusion



Server Market Trends



** IDC Workload Models 2003 Model, 11/22/03

Blade Market Trends

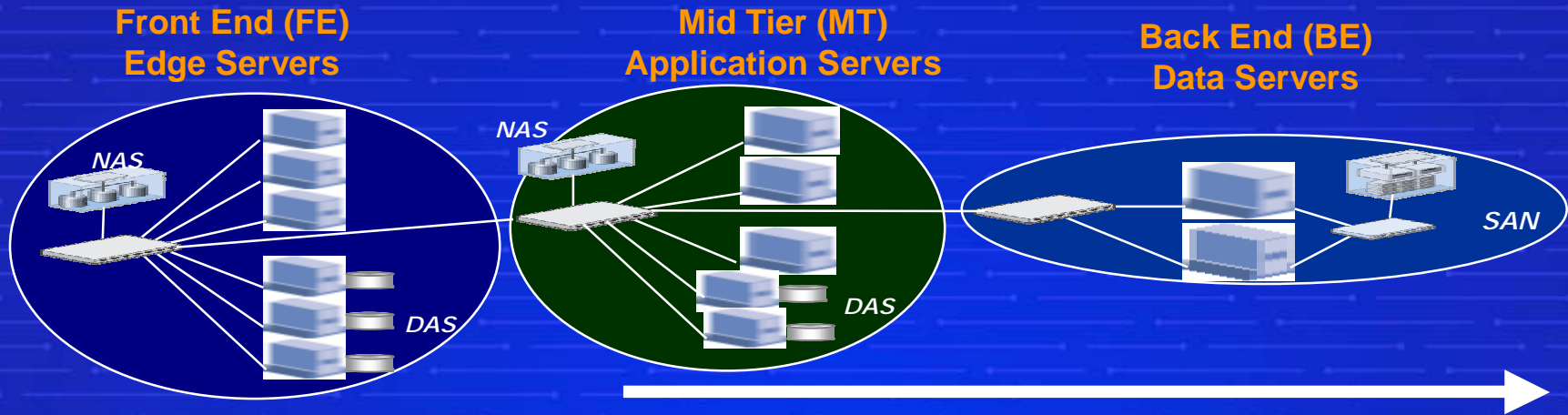
- Blades being deployed widely in large data centers
- Blades cut into Telco pie ~ 26% of Telco servers by '07 – In-Stat/MDR

High Performance Computing (HPC) Market Trends

- Highest growth in the “Technical Capacity” Servers
- Clusters built w/ commodity components gaining popularity
 - Expected to be 20% of HPC market by 2007
- InfiniBand® positioned as clustering fabric



Emerging Blade Usage Models



- Blades are increasingly being deployed in back-end & mid-tier applications
- Low cost cluster deployment is growing at the expense of high-end cluster solutions
- Server virtualization is gaining popularity
 - Server consolidation
 - Increasing server utilization



NAS = Network Attached Storage

DAS = Direct Attached Storage

SAN = Storage Area Network



Outline

- Key Messages
- Blade Market Trends and Emerging usage models
- Application characterizations for emerging usage models
- Emerging next generation architectures
- Possible Ethernet enhancements
- Conclusion

Mid-Tier (MT) Application Characterizations

- Workload → Mostly Dynamic
- Networking
 - Front End ↔ Mid Tier All networking
- Storage Traffic
 - Storage traffic is less than 10%
 - Storage architecture is DAS; trending to NAS
- Inter Processor Communication (IPC) Traffic
 - Within the MT tier
 - Limited & function of implementation (Clustered/Standalone)
 - Mid-Tier ↔ Back-End All communication tends to be IPC



NAS = Network Attached Storage

DAS = Direct Attached Storage

SAN = Storage Area Network



Back-End Application Characterizations

- All workload is Dynamic
 - Database buffers efficiently cache data
- Networking Traffic is < 200 Mbps
- Storage Traffic
 - Moderate to significant (2-4 Gbps)
 - Storage architectures mixed SAN and NAS
- Inter processor communication (IPC) Traffic
 - Within database tier
 - Significant, especially in distributed databases (2-4 Gbps)
 - Small messages (<1KByte); Latency sensitive
 - Between Back-End & Mid-Tier all communication is IPC
 - Latency is critical for IPC traffic
 - “Latency directly affects cluster scalability” -- Oracle



Source: Achieving Mainframe-Class performance on Intel servers using Infiniband – Oracle White Paper, April 2003



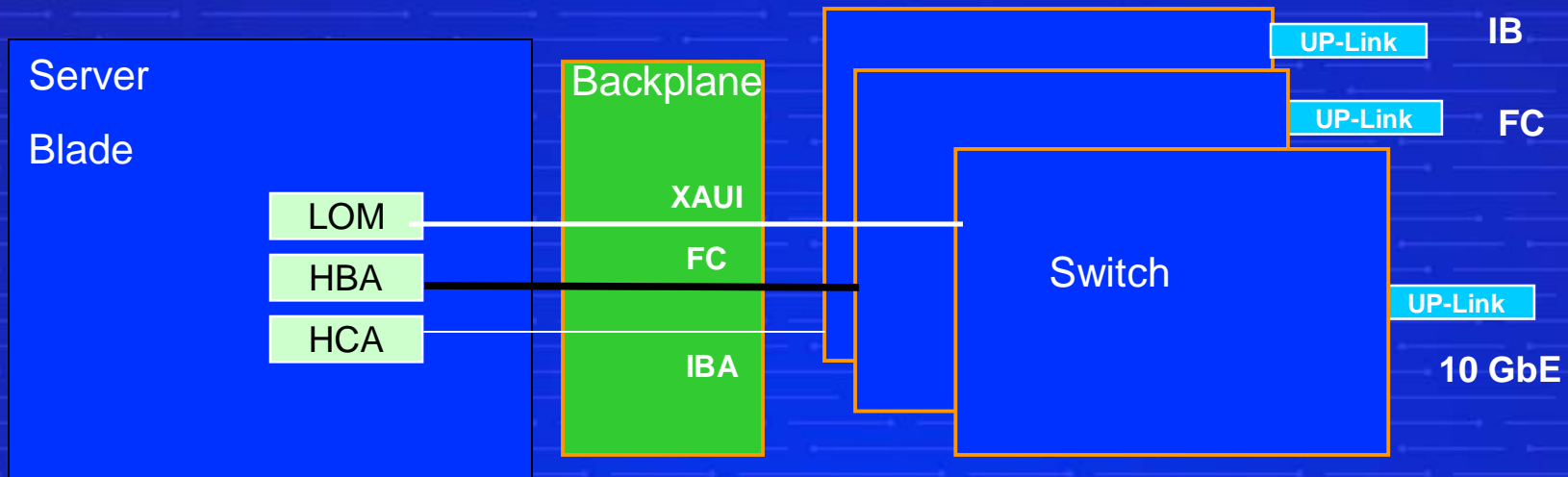
Outline

- Key Messages
- Blade Market Trends and Emerging usage models
- Application characterizations for emerging usage models
- Emerging next generation (NG) architectures
- Possible Ethernet enhancements
- Conclusion



Emerging NG Architecture – 3 Fabrics

Burdens Base Chassis cost



- One fabric each w/ redundancy for LAN, Storage and IPC
 - + Easier to meet varying characteristics of FE, BE, MT applications
 - + Traditional approach
 - - Burdens base chassis cost
 - - Backplane & Chassis design is more complicated
 - - Requires vendors to design & maintain three fabric types

Emerging NG Architecture – InfiniBand® as converged fabric

Significantly higher system cost



- Use IB as the convergent fabric & provide translators for LAN, SAN
 - + IB can tunnel LAN, SAN & IPC traffic
 - + IB good for IPC and storage traffic
 - Revolutionary usage model – much harder to sell to IT
 - Overall system cost could be ~ 2-3 X cost of system based on Ethernet fabric

Outline

- Key Messages
- Blade Market Trends and Emerging usage models
- Application characterizations for emerging usage models
- Emerging next generation architectures
- Possible Ethernet enhancements
- Conclusion



Possible Enhancements

- Improved congestion handling
 - Rate control schemes
 - Decrease probability of packet drops or selective packet dropping
 - Critical for storage and IPC
- Better traffic management
 - Service levels or traffic prioritization



Path Forward

“Increased bandwidth alone with Ethernet is not good enough, need improvements in congestion management, latency, reliability, etc.” -- Advanced TCA Telecom Customers

- Perform this effort in Backplane Ethernet
 - Providing 10G serial may not be a strong enough solution
 - Simple schemes exist in the market for improvements

Outline

- Key Messages
- Blade Market Trends and Emerging usage models
- Application characterizations for emerging usage models
- Emerging next generation architectures
- Possible Ethernet enhancements
- Conclusion



Conclusion

- Ethernet blade architectures will need to have Fabric capabilities
- Congestion handling can be improved using rate control
- Traffic management can be enhanced through prioritization
- BESG should consider layer 2 enhancements for backplane Ethernet

