InfiniBand Solutions Overview

Prepared for CERN – JUN 26, 2006

Thad Omura
Vice President of Product Marketing
Mellanox Technologies

Mellanox Technologies, Ltd.

- **Delivering disruptive, high-performance connectivity products demanded by computer and data storage communities**
- **Rapidly growing, fabless semiconductor company**
- **Founded in March 1999**
- **Well-developed international operations**
  - **Business HQ in California**
  - **Engineering HQ in Israel**
  - **Global sales offices and support**
- **$89M funding**
  - **1st tier VCs and corporate investors**
    - Bessemer, Sequoia, USVP, Dell, IBM, Intel, Sun

Leading market provider of InfiniBand silicon solutions

HPC Trends - Overview

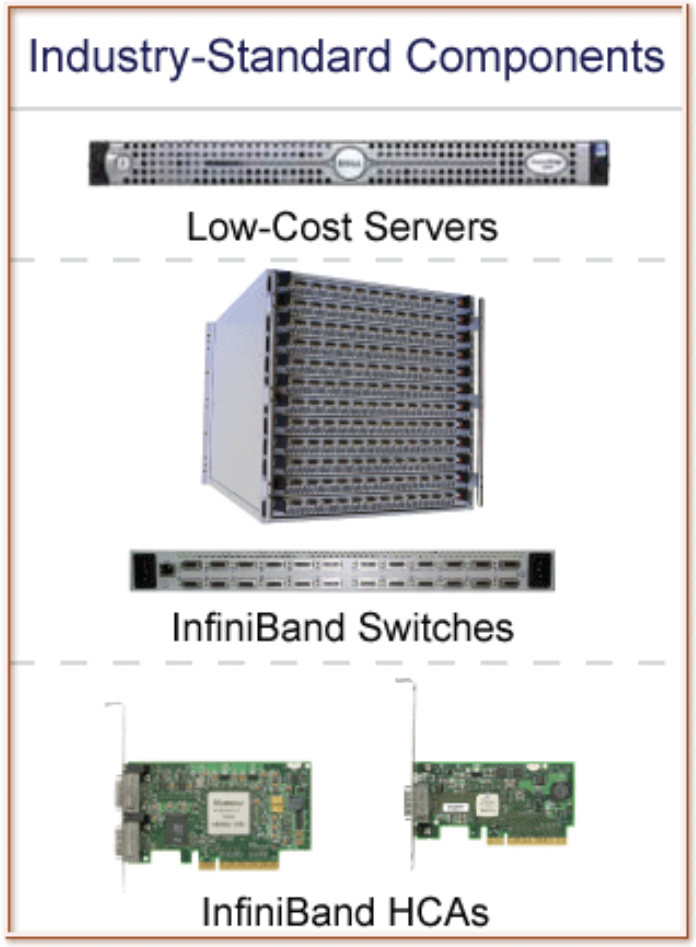
- **Architecture**
  - Clusters dominate ↗
  - 64 bit ↑
  - Multicore ↑
- **Operating systems**
  - Linux dominates →
  - Windows CCS ?
- **Interconnect family**
  - Standard ↗
- **Interconnect**
  - InfiniBand ↗
  - Myrinet ↘
  - Quadrics ↘
  - GigE →

- **Two highest ranked industry-standard clusters use InfiniBand**
  - #4 NASA, 52TFlops
  - #5 Sandia National Laboratories, 38Tflops

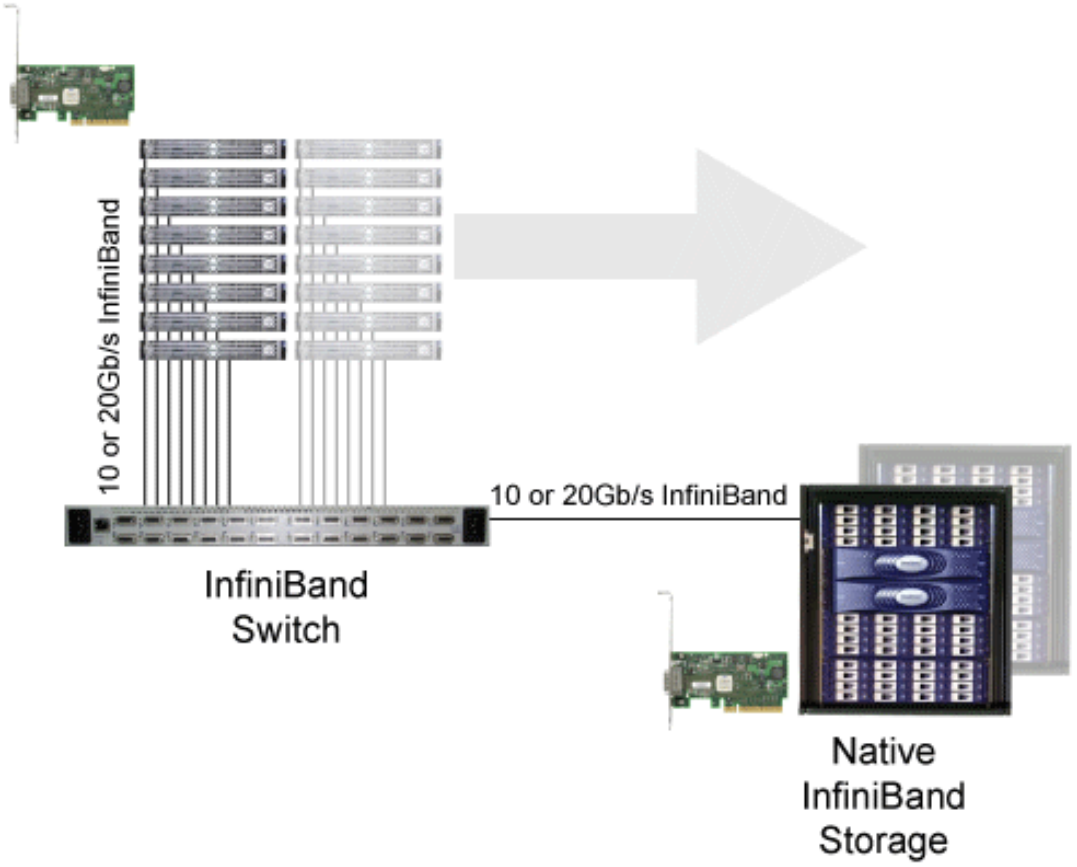
- **World's highest reported x86 efficiency**
  - #130 Galactic Computing
  - InfiniBand MemFree
  - 84.35% Linpack efficiency

蚬壳星盈 GALACTIC COMPUTING

InfiniBand – The Optimal Cluster and Connectivity Solution



InfiniBand enables industry-standard servers and storage to scale affordably



HPC Trends –Compute Nodes

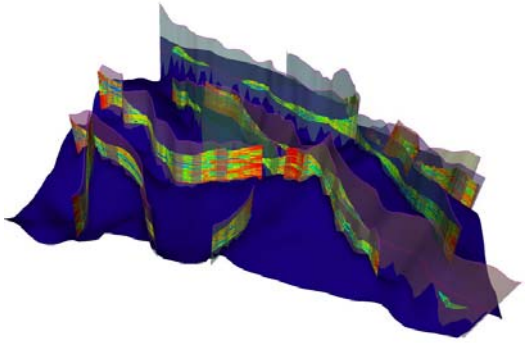
- **Bigger clusters**
  - **Petaflop**
  - **Scalability**
  - **Congestion control**
    - HW congestion control
- **Multicore**
  - **Efficiency**
  - **Low CPU processing overhead**
    - Overlapping I/O communication with CPU computation cycles
  - **Uni socket, dual socket, quad socket**
    - Ensure the same world-class performance regardless of the amount of the server CPU cores

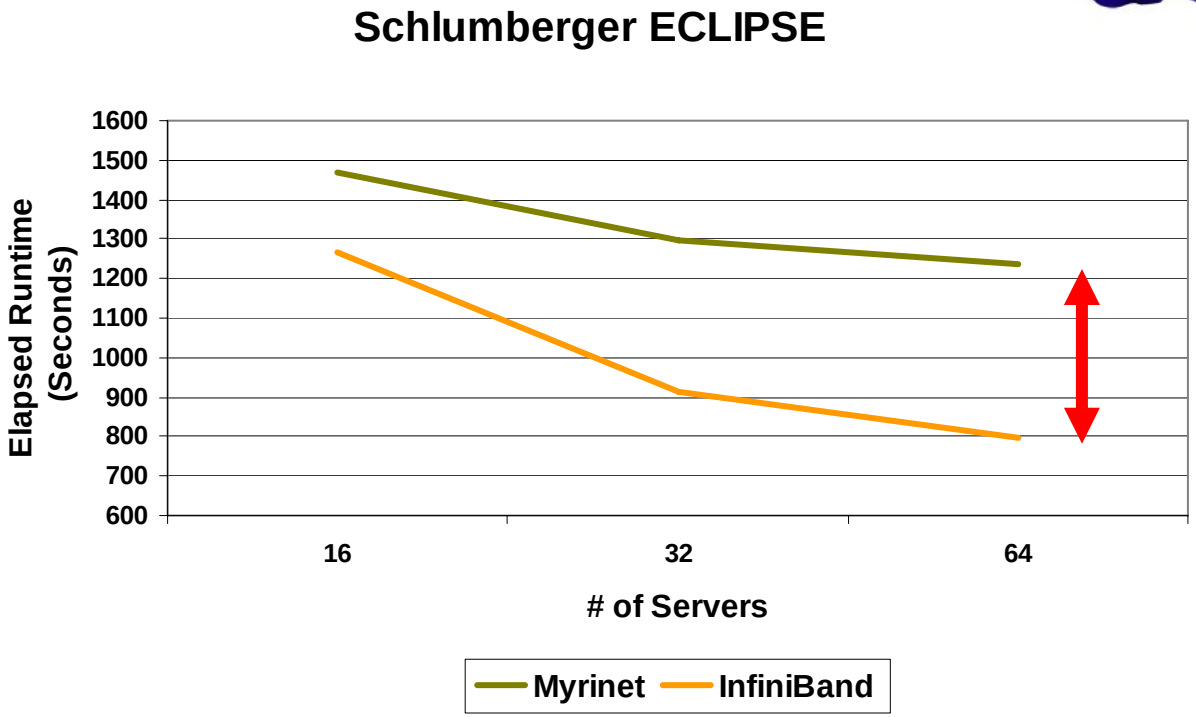
Virginia Tech, 2003
- 12TFlops

Tokyo Institute of Technology, 2006
- 38.18TFlops
- Sun Fire x64 servers with 10,480 Opteron processor cores
- 1300 Mellanox InfiniBand DDR MemFree HCAs
- Native InfiniBand storage with 1 PB of hard disk storage

HPC Trends – Scalable Applications Example: Schlumberger Eclipse

- **ECLIPSE million cell model**
  - HP DL145 2.6Ghz servers, single CPU
  - OS: SUSE 9



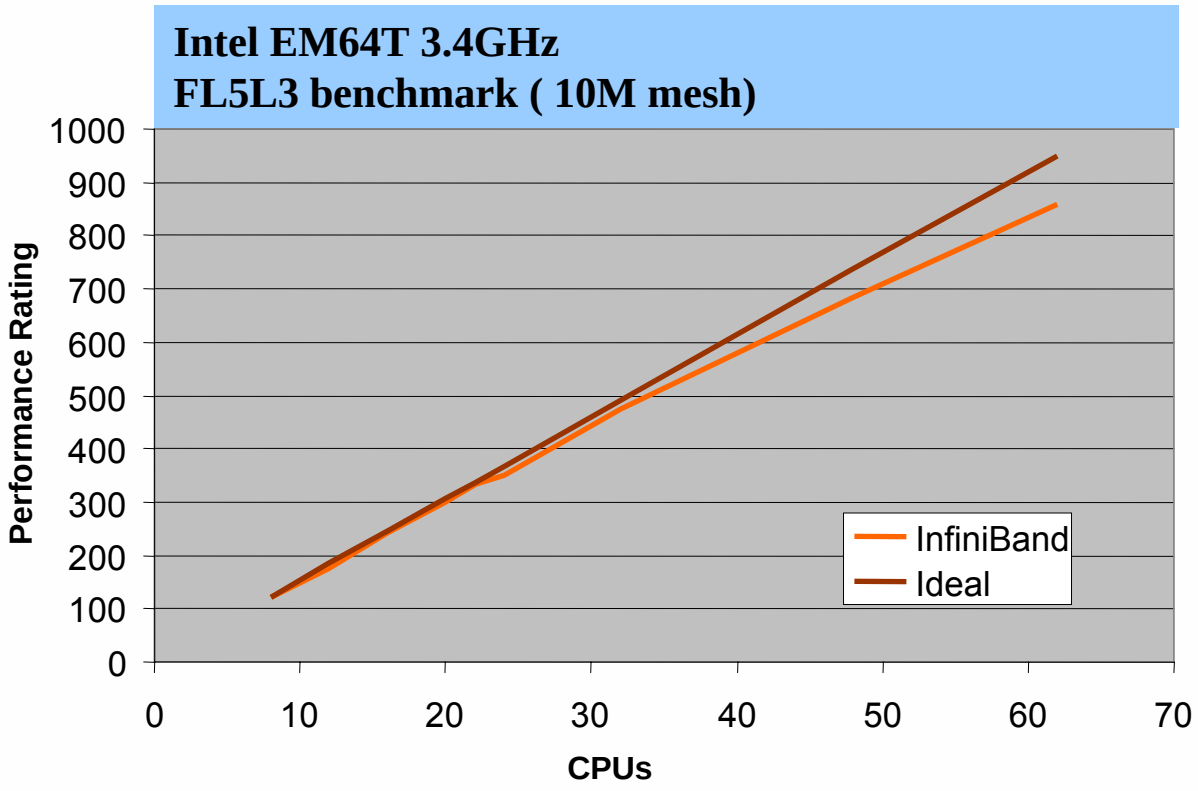


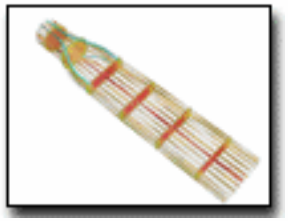
Infiniband is 55% more efficient !

Data provided by

Schlumberger

HPC Trends – Scalable Applications Example: Fluent

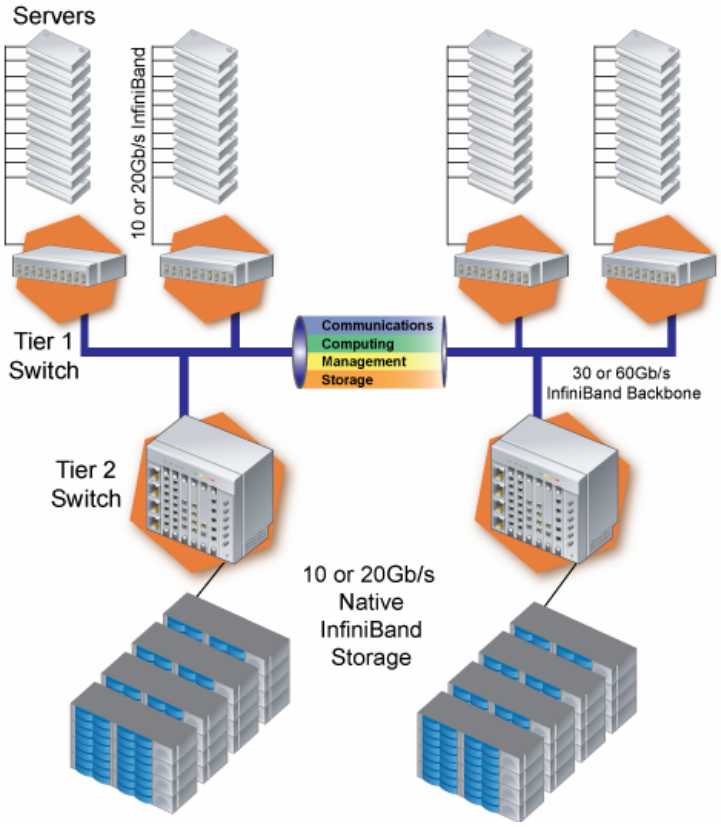


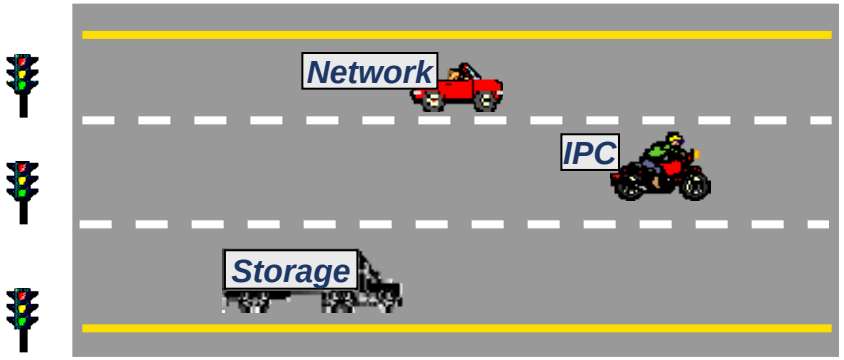


Great Scaling from small to large clusters!

HPC Trends – I/O Consolidation

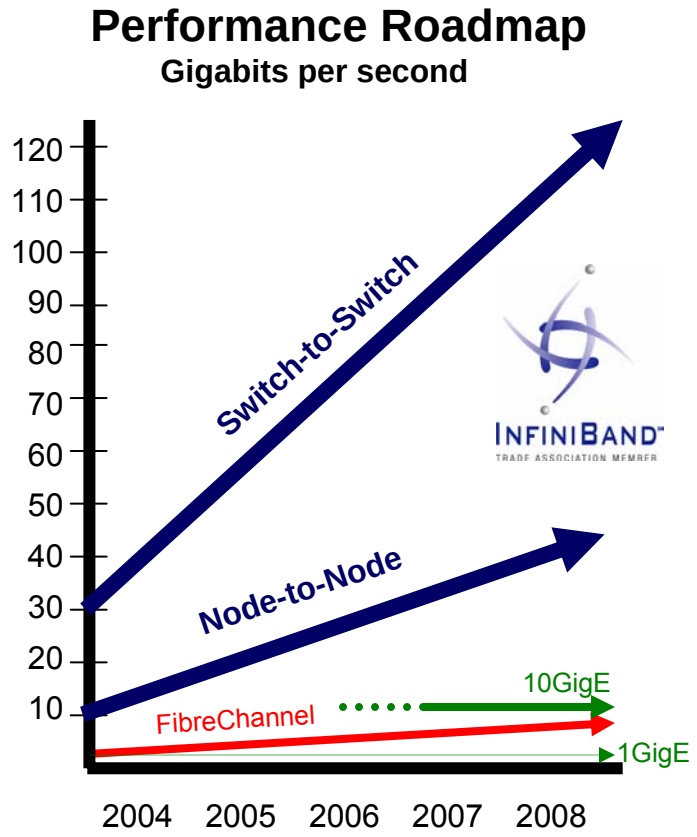
- Communications, computing, management and storage onto a single link
- Simplified Management
- Scalability
- Total cost of ownership
- Quality of Service
- Channel I/O Architecture





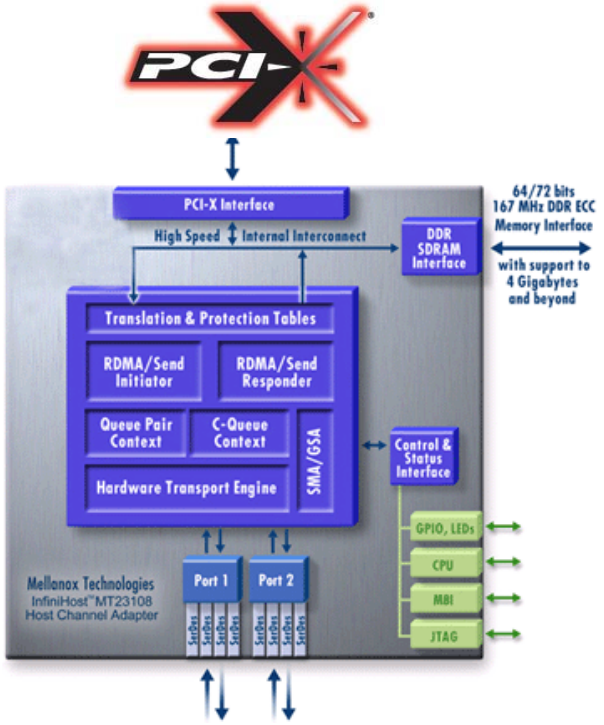
InfiniBand - The Complete Fabric

- Industry-standard
- Price/Performance
  - 10/20/40Gb/s per node
  - 30/60/120Gb/s switch-to-switch
  - 2.6us -> 1us application latency
- Offload
  - RDMA and Transport
- Reliable Fabric
- I/O consolidation
- Scalable to tens of thousands of nodes
- Low power
- Roadmap



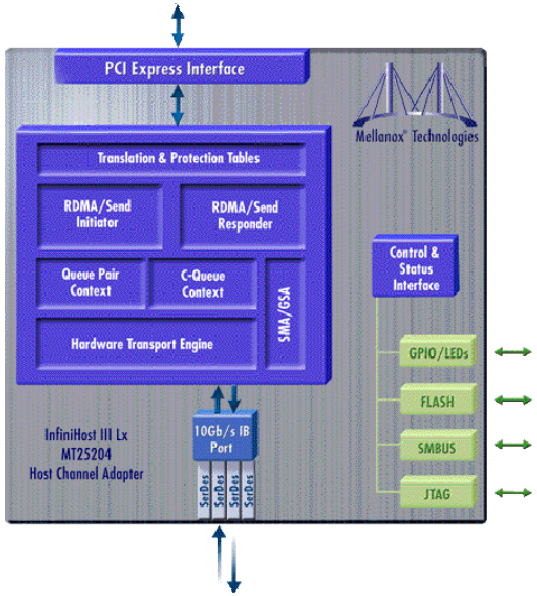
HCA Architecture

InfiniHost



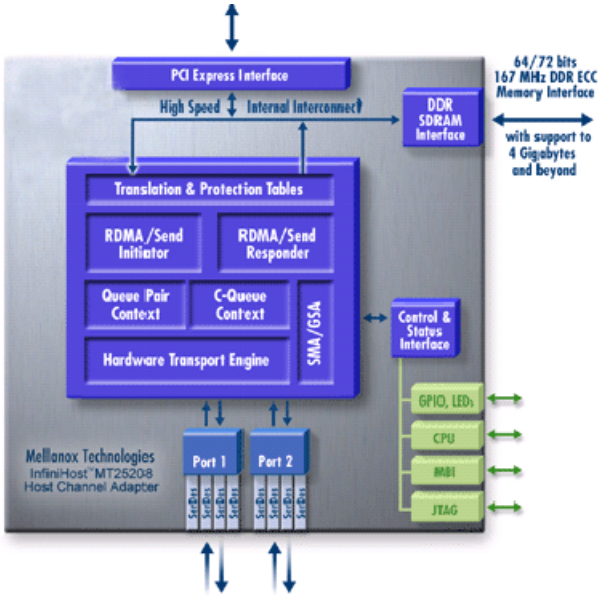
InfiniHost III Lx

PCI EXPRESS™ x8



InfiniHost III Ex

PCI EXPRESS™ x8



- RDMA and hardware transport
- Low Latency
- Wire-Speed Capabilities
- Native 64-bit Support

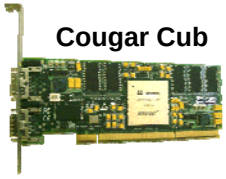
- 3rd generation HCA architecture
- Backward software compatible
- InfiniHost III Lx and Ex have both SDR and DDR InfiniBand versions

HCA Cards

InfiniHost

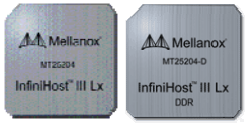
Dual 4X IB

Cougar Cub PCI-X



MHXL-CFXXXT
128/256MB Memory Down

InfiniHost III Lx SDR/DDR



Single 4X IB

Tiger PCI EXPRESS x4

MHES14-XT
MemFree, Media adapter
GA DEC 2005

Cheetah PCI EXPRESS x8

MHES18-XT
MemFree, Media Adapter

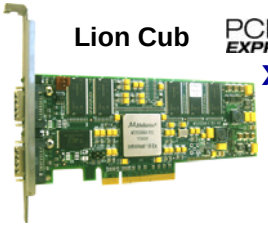
Cheetah DDR PCI EXPRESS x8

MHGS18-XT
MemFree, Media Adapter

InfiniHost III Ex SDR/DDR

Dual 4X IB

Lion Cub PCI EXPRESS x8



MHEL-CFXXXT
128/256MB Memory Down

Lion Mini PCI EXPRESS x8

MHEA28-XT
MemFree, Media Adapter

Lion Cub DDR PCI EXPRESS x8

MHGA28-1T
128MB, Media Adapter

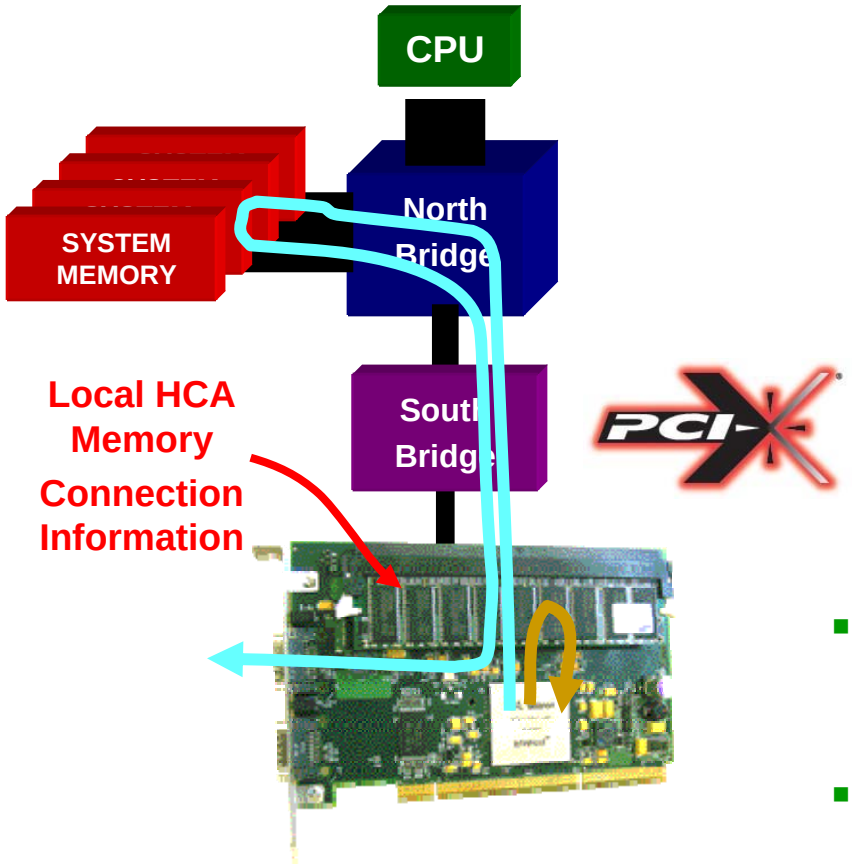
Lion Mini DDR PCI EXPRESS x8

MHGA28-XT
MemFree, Media Adapter

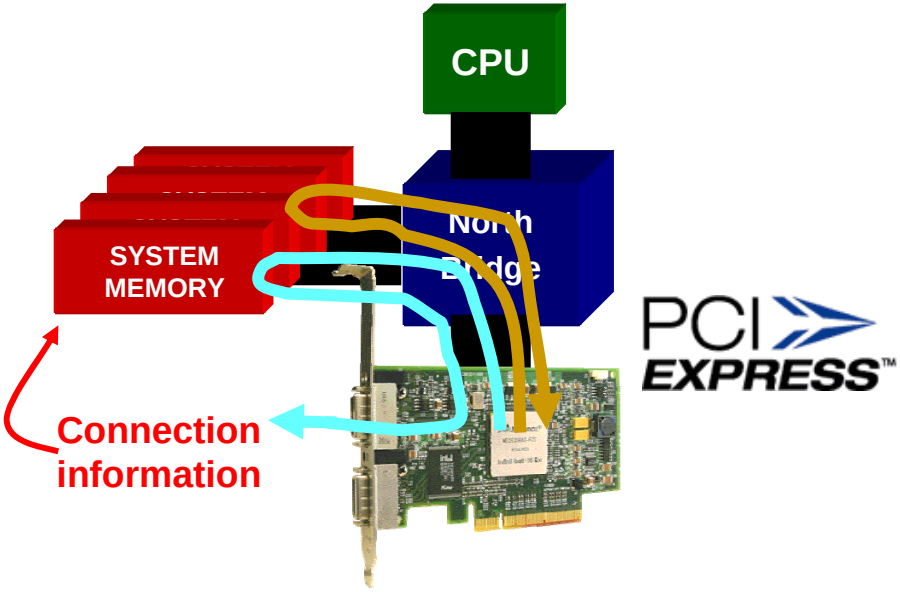
SDR **DDR**

PCI Express Enables MemFree

InfiniBand HCA with Local Memory



InfiniBand MemFree HCA

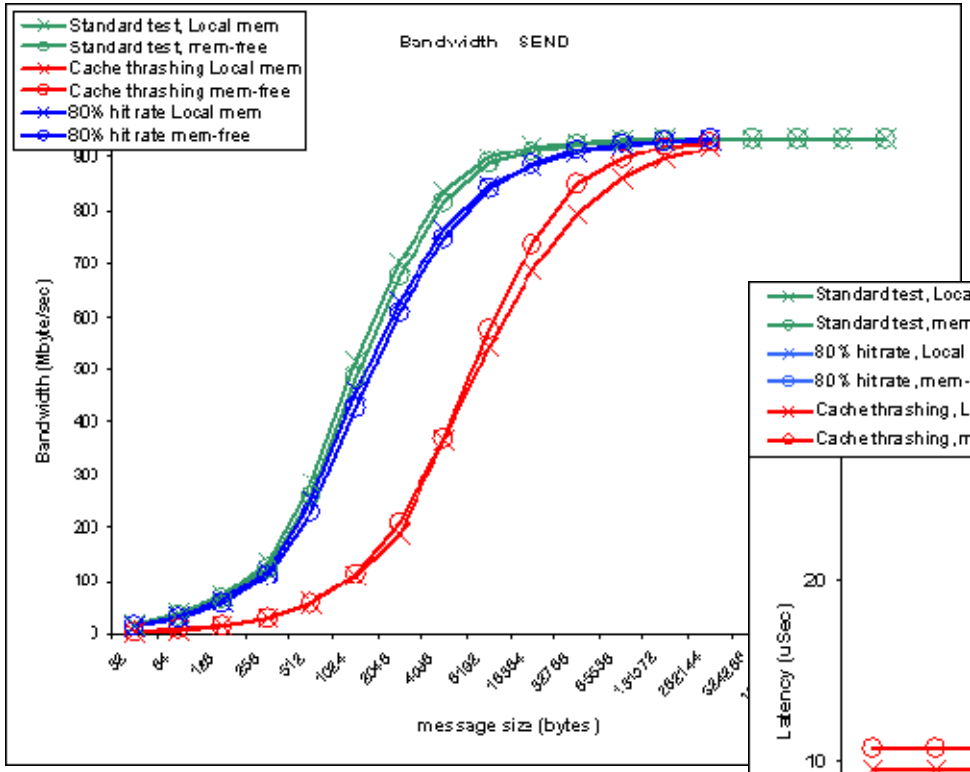


- PCI Express architecture provides lower latency access from IO device to memory and thus enables MemFree
- Increased IO Bandwidth makes context cache replacement painless
- Increased reliability

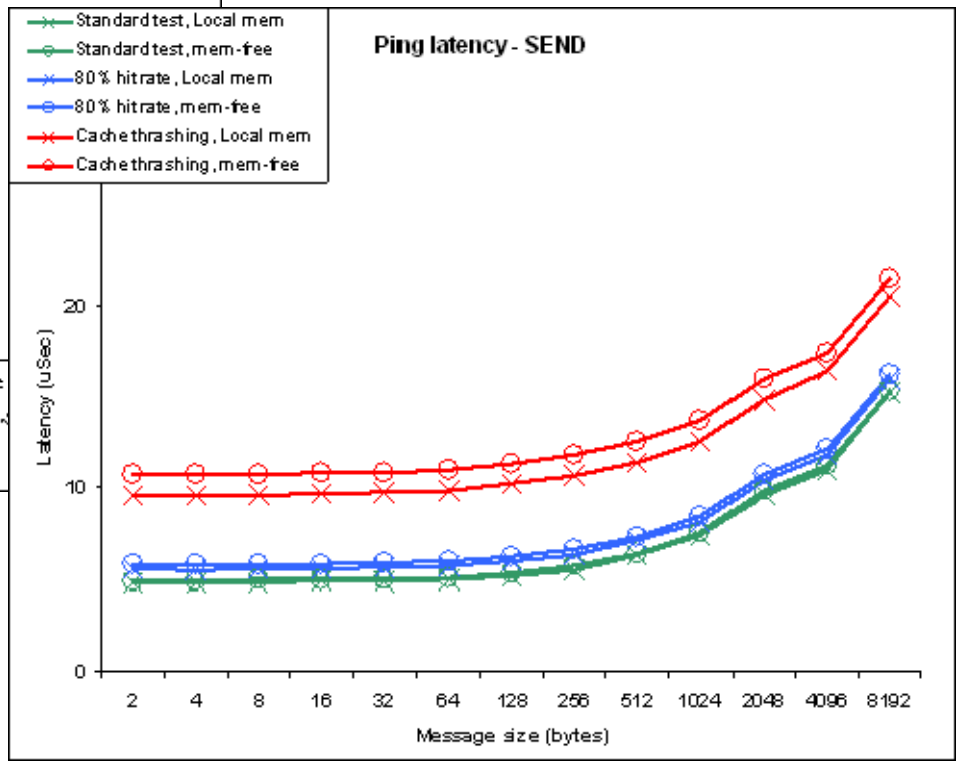
Data

Connection

MemFree Performance



Negligible to NO performance impact!

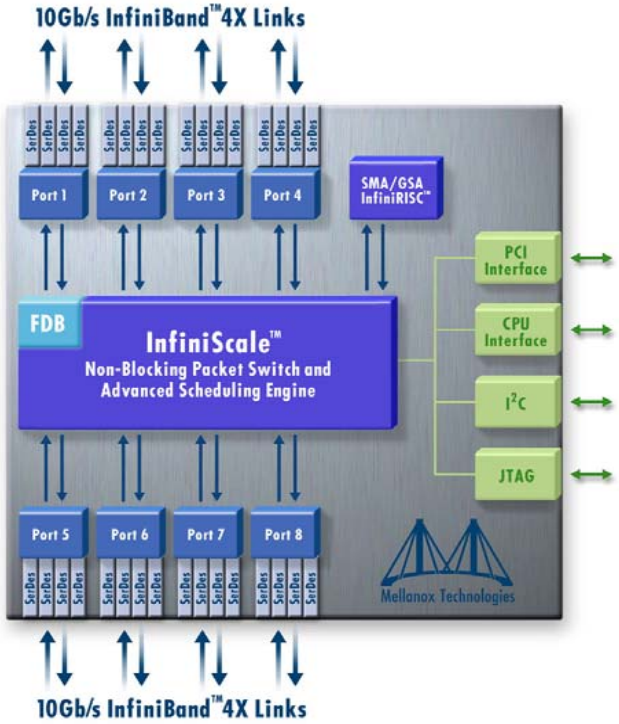


Significant Cost, Power and Footprint Reduction for Landed on Motherboard (LOM) and Blade Servers!

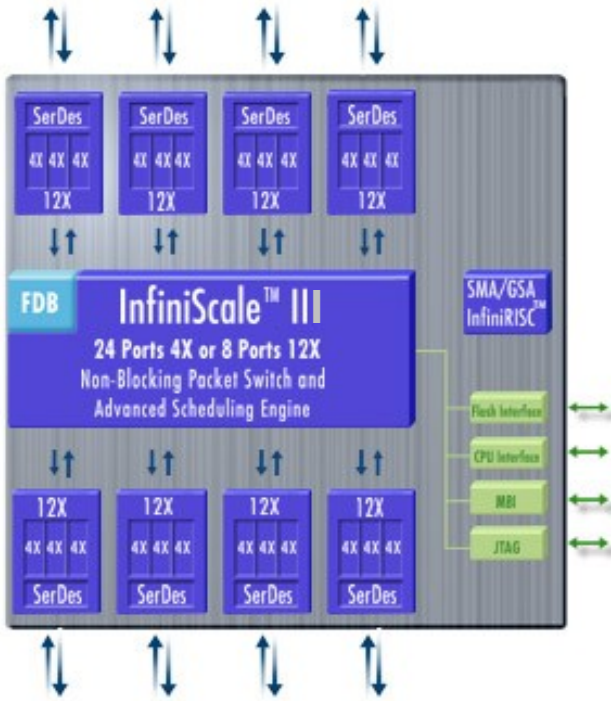
MemFree Whitepaper available for further information

Switch Architecture

InfiniScale



InfiniScale III



- Full Wire Speed, Non-blocking
- Integrated SerDes

- InfiniScale III has SDR and DDR
- 200ns SDR, 140ns DDR ball-to-ball switching latencies

Broad Industry Adoption

Mellanox TECHNOLOGIES

InfiniBand Landed on Motherboard

InfiniBand Blade Servers

Servers

Switches & Infrastructure

InfiniBand Backend Clustering and Failover

Native InfiniBand Visualization Storage System

Native InfiniBand Block Storage Systems

Native InfiniBand Solid State Storage

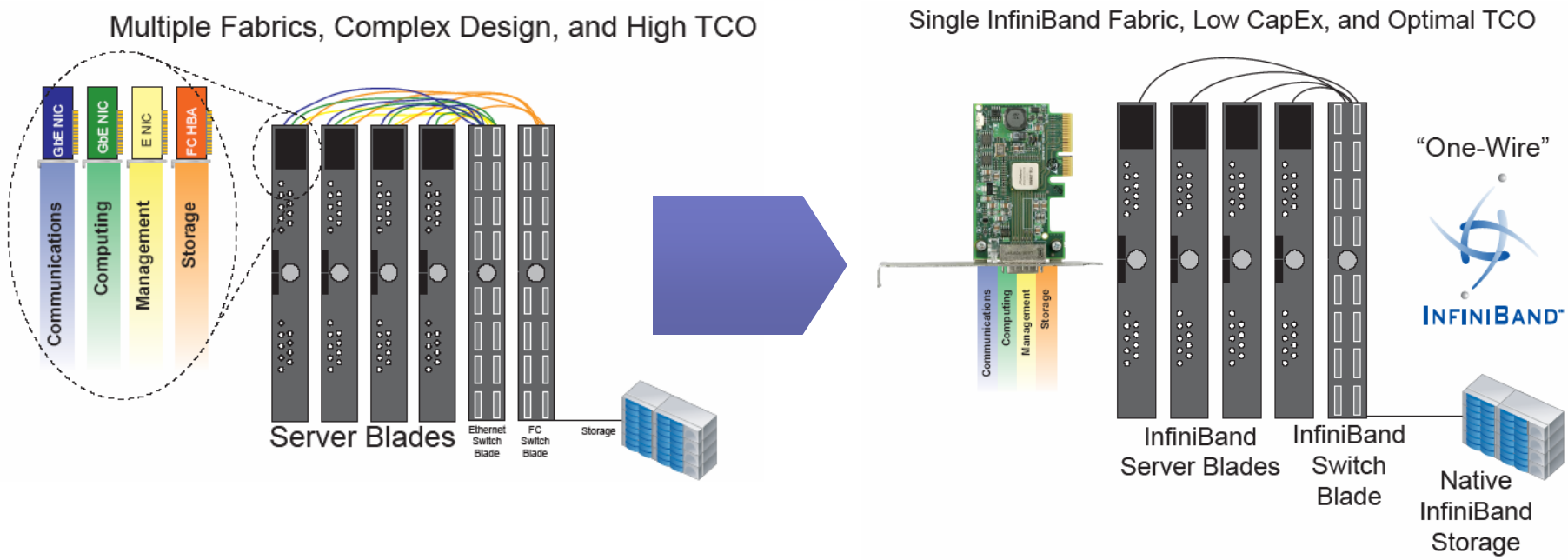
Native InfiniBand Clustered File Storage System

Storage

Embedded

*Partial Lists

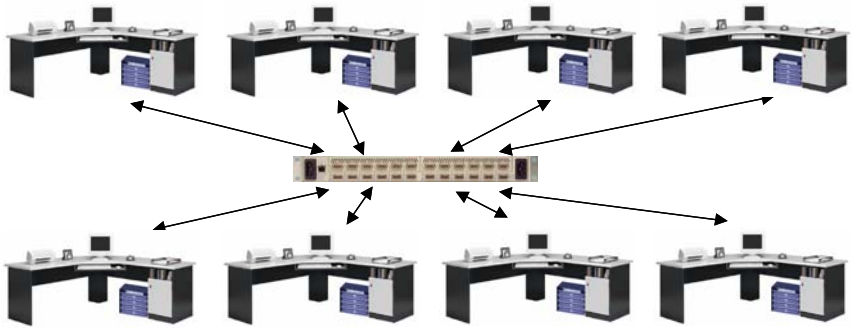
InfiniBand is Ideal for Blade Servers



- **Single fabric backplane through InfiniBand I/O consolidation capabilities**
- **The ONLY 10Gb/s and now 20Gb/s interconnect shipping in blade servers today!**

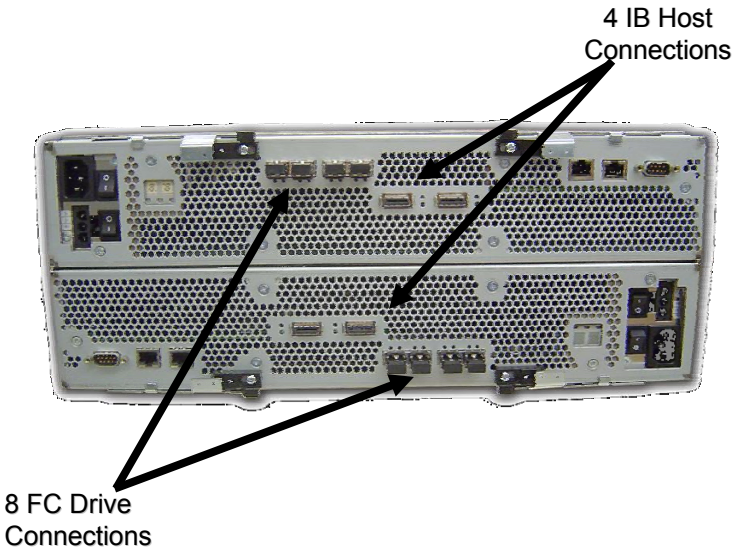
Personal Supercomputers

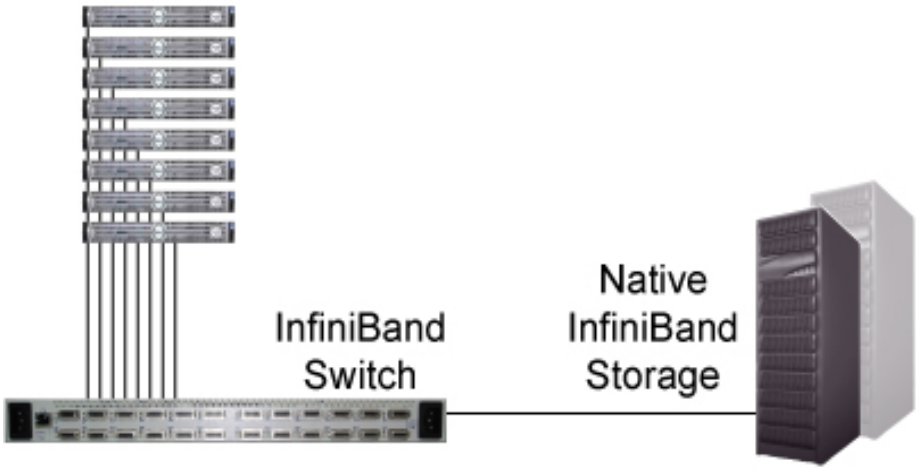
- Easy to use, Turnkey cluster
- Fits into “cubicle” environment
- Standard power, quiet operation
- Ability to scale efficiently



Block Storage Controller
LSI Logic: 6498

- Dual-active 6498 controllers
- Four 10 Gb/s IB Host Channels
- Eight 4 Gb/s FC drive channels
- 2 GB of dedicated data cache
- Custom XOR engine generates RAID parity with no performance penalty
- 1300MB/s Sustained throughput
- Up to 90 TB of capacity and fully-featured functionality
- www.lsilogic.com





Xiranet XAS1000-IB

- **Scalable InfiniBand / iSCSI Storage System**
- **Based on Serial Attached SCSI (SAS)**
- **Modular Architecture:**
  - **Controller; up to 6 internal SCSI disks**
  - **Up to 4 SAS links per Controller**
  - **1+ SAS JBODs can be attached**
  - **Capacity of more than 100 TB supported**
  - **Fail Over und Clustering**
- **SAS JBOD**
  - **Mixed SAS and SATA disk support**
  - **Up to 5 SAS JBODs per chain (60 drives)**
  - **2U, up to 12 drives, hot-plug, redundant architecture**
- **www.xiranet.com**

XAS1000-IB

Expanding Industry Commitment

Processors	AMD 64 Opteron, Intel Xeon inside, Intel Itanium 2 inside, PowerPC, SPARC
Operating Systems	OpenFabrics Alliance, Novell SUSE, redhat, HP HPUX, IBM AIX, Microsoft Windows, Solaris, Mac OS X v10.4 "Tiger"
Virtualization	Cisco Systems, vmware, Xen, Microsoft Virtual Server 2005
Enterprise Applications	Ready for IBM DB2 software for Linux, Oracle 10g Database, Microsoft SQL Server 2005, IBM WebSphere
High Performance Computing Applications	Schlumberger ECLIPSE, FLUENT, LSTC Livermore Software Technology Corp. LS-DYNA, WOMBAT Financial Software, Platform Platform Rocks, accelrys, SCALI Scaling the Linux Datacenter
Storage	System Fabric Works, CFS Cluster File Systems, Inc., TERRASCALE Technologies, IBRIX, NetApp The evolution of storage., HP, IBM eserver General Parallel File System, FalconStor Software, ISILON Systems, LSI LOGIC

* Partial Lists

Copper Cables

- Gore, Amphenol, Leoni, Tyco, FCI, Molex, Meritec (many others!)
  - Standard length up to 12 meters
  - 24/26AWG, ~4-5” bend radius
- DDR support up to 10m over 24AWG
- 30 Gauge Cables (Gore, Leoni and others)
  - Thin cables (same as Category 5/6)
  - 2” bend radius

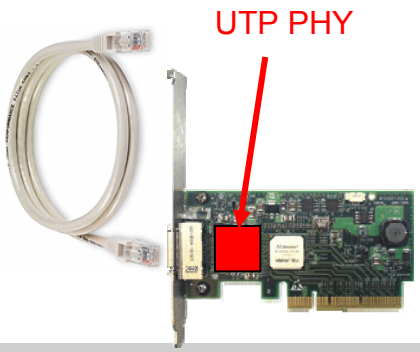
4X Cable

12X Cable

- UTP Solutions in Testing
  - Proprietary and non-proprietary
  - Deployment pending production

KeyEye Communications

Teranetics



Parallel Fiber Modules

- **Media Adapter Module (Emcore, Fujitsu)**
  - Plugs into existing IB copper connector
  - 2.5Gb/s per channel (10Gb/s total)
  - ~1W, ~no latency penalty (< 5ns)
  - 300m over MTP/MPO
    - 12 parallel MMF Fiber Ribbon Cable
  - 5Gb/s (DDR) in testing now
- **POP4 and SNAP12 "Fiber Down" solution available**
  - 300m over MTP/MPO

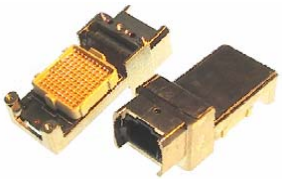
MTP/MPO Cable (12 Fiber Ribbons)

QTR3400

FCU-050M008-012 Media Adapter

POP4

SNAP12



Supercomputing Is Mainstream

Enterprise Data Centers	High Performance Computing	Embedded
Database	Computer Aided Eng	Data Acquisition
Financial	Entertainment	Life Sciences
Business Analytics	Geosciences/Weather	Military

InfiniBand Solutions Summary

- **Well established cluster connectivity solution**
  - **Price, Performance, Offload, Scalability, Reliability, Consolidation, Low Power, Roadmap and more**
- **Wide variety of computing form factors**
  - **Standalone servers, Landed on Motherboard, Blade Servers, Personal Supercomputers**
- **Wide variety of switching and infrastructure solutions**
  - **8-port 10Gb/s to 288-port 20Gb/s switches**
  - **InfiniBand-over-WAN**
- **Emergence of Native InfiniBand storage delivers true I/O consolidation**
- **Unified software ecosystem with OpenFabrics.org**
- **InfiniBand will continue to proliferate as the optimal cluster interconnect**