Paving the Road to Exascale

Interconnecting Your Future

February 19, 2016 | PC Cluster Workshop

Mellanox TECHNOLOGIES

Connect. Accelerate. Outperform.™

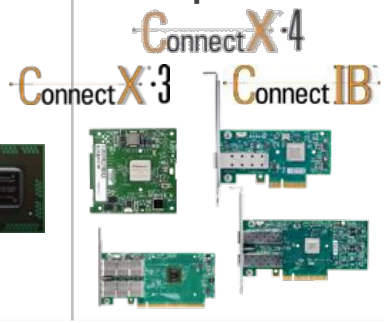
本資料に関するお問い合わせ先

メラノックステクノロジーズジャパン株式会社

japan_sales@mellanox.com

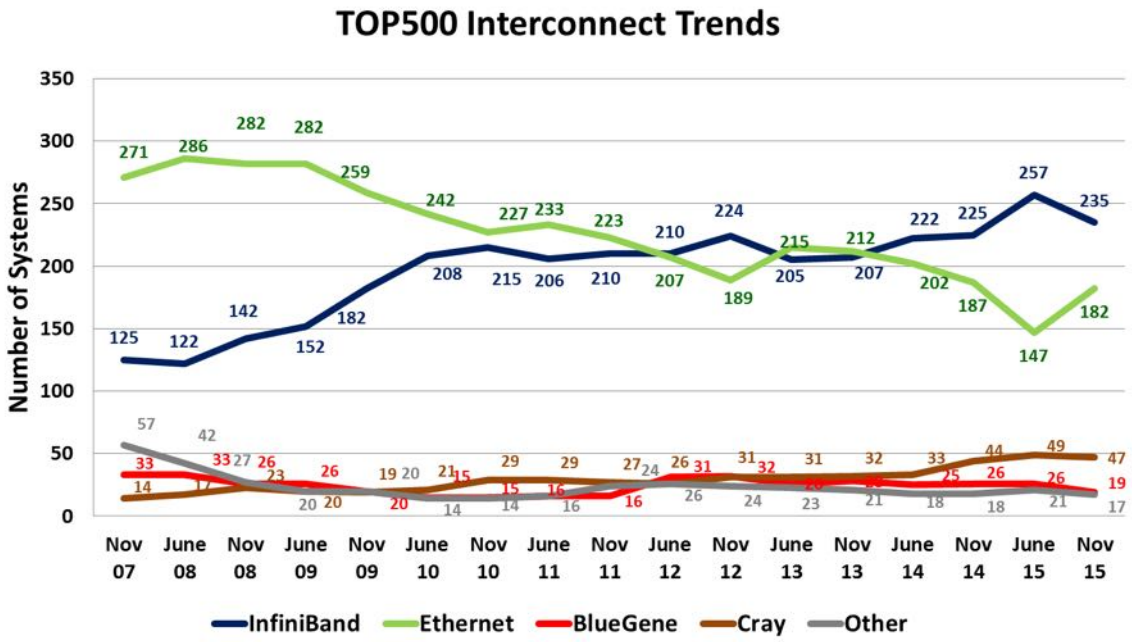
Leading Supplier of End-to-End Interconnect Solutions

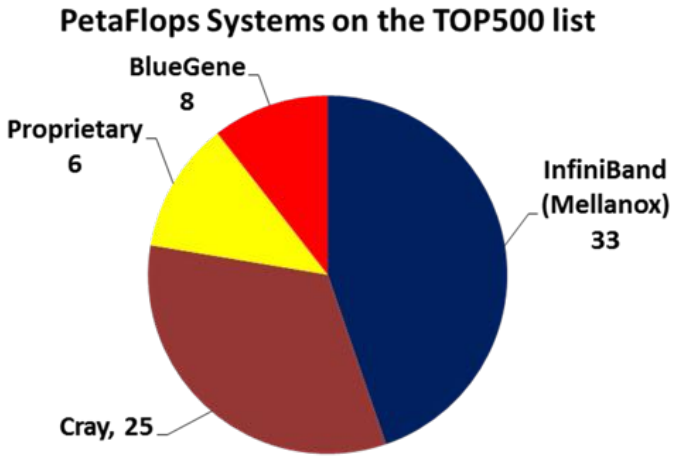
Comprehensive End-to-End InfiniBand and Ethernet Portfolio

ICs	Adapter Cards	Switches/Gateways	Software and Services	Metro / WAN	Cables/Modules
					

At the Speeds of 10, 25, 40, 50, 56 and 100 Gigabit per Second

TOP500 Interconnect Trends





"LENOX" EDR InfiniBand system
Lenovo HPC innovation center

Shanghai Supercomputer Center
Magic Cube II supercomputer

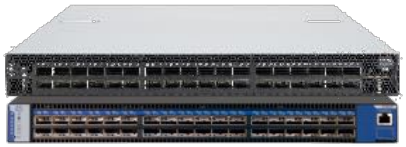
InfiniBand Switch Portfolio

Modular Switches




648 port | 324 port | 216 port | 108 port

Edge Switches



36 ports externally managed

18 port externally managed

8-12 ports externally managed

36 ports managed

18 port managed

12 port managed

Long Distance

metroX™

Bridge – VPI

SwitchX®-2

Management

High-Performance Designed 100Gb/s Interconnect Solutions

Adapters

ConnectX-4

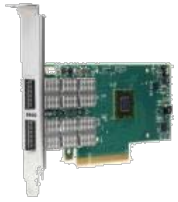
100Gb/s Adapter, 0.7us latency

150 million messages per second

(10 / 25 / 40 / 50 / 56 / 100Gb/s)

ConnectX-4 Lx

ConnectX-4 Lx Programmable



Switch

Switch-IB 2

36 EDR (100Gb/s) Ports, <90ns Latency

Throughput of 7.2Tb/s

7.02 Billion msg/sec (195M msg/sec/port)

Switch

Spectrum

32 100GbE Ports, 64 25/50GbE Ports

(10 / 25 / 40 / 50 / 100GbE)

Throughput of 6.4Tb/s

Interconnect

LinkX

Transceivers

Active Optical and Copper Cables

(10 / 25 / 40 / 50 / 56 / 100Gb/s)

VCSELs, Silicon Photonics and Copper

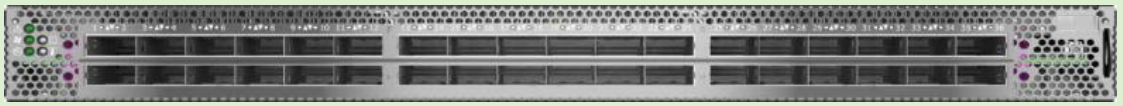
Enter the Word of Scalable Performance – 100Gb/s Switch

Best Performance

- 90ns switch latency
- Throughput of 7.2 Tb/s in 1U
- 195M messages per second
- 136W ATIS reported power

Enhanced Capabilities

- InfiniBand Router
- Adaptive Routing (AR)
- Fault Routing (FR)



Switch IB™

High Resiliency

- High efficiency power supplies
- AC / DC / BBU power supplies option
- Class 4 (3.5W) supported on all ports

x86 Powerful CPU

- Improved software upgrade time
- Up to 2048 nodes in cluster
- Run Virtual Machine (VM)

Benefits of InfiniBand Router

- Enable scaling an HPC cluster over 48K LIDs
- Enable sharing a common storage network by multiple disconnected subnets
  - Limit congestion spread to source subnet
- Allow running HPC/MPI jobs efficiently on the joint network
  - Maintaining large bisectional bandwidth between the subnets
  - Low latency penalty for crossing subnets
- Isolation of SM responsibilities
- Simple administration and out-of-the-box experience

Enter the World of Scalable Performance – 100Gb/s Adapter

ConnectX-4: Highest Performance Adapter in the Market

InfiniBand: SDR / DDR / QDR / FDR / EDR

Ethernet: 10 / 25 / 40 / 50 / 56 / 100GbE

100Gb/s, <0.7us latency

150 million messages per second

OpenPOWER CAPI Technology

CORE-Direct Technology

GPUDirect RDMA

Dynamically Connected Transport (DCT)

Ethernet Offloads (HDS, RSS, TSS, LRO, LSOv2)

Shattering The World of Interconnect Performance!

ConnectX-4 EDR 100G InfiniBand	
Uni-Directional Throughput	100 Gb/s
Bi-Directional Throughput	195 Gb/s
Latency	0.61 us
Message Rate	149.5 Million/sec

InfiniBand Adapters Performance Comparison

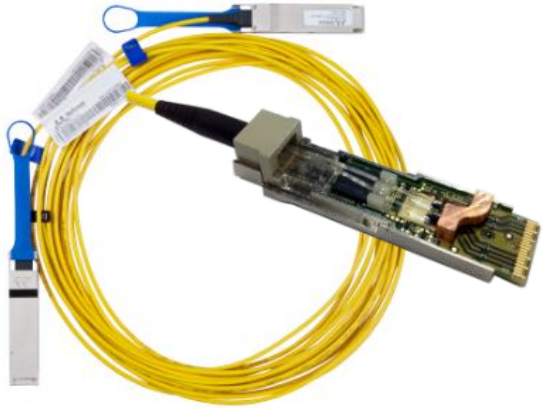
Mellanox Adapters Single Port Performance	ConnectX-4 PCI Express 3.0 x16 EDR 100Gb/s	Connect-IB PCI Express 3.0 x16 FDR 56Gb/s	ConnectX-3 Pro PCI Express 3.0 x8 FDR 56Gb/s
Uni-Directional Throughput	100 Gb/s	54.24 Gb/s	51.1 Gb/s
Bi-Directional Throughput	195 Gb/s	107.64 Gb/s	98.4 Gb/s
Latency	0.61 us	0.63 us	0.64 us
Message Rate	149.5 Million/sec	105 Million/sec	35.9 Million/sec

Mellanox QSFP 100Gb/s Cables

Complete Solution of 100Gb/s Copper and Fiber Cables

Copper Cables

VCSEL AOCs

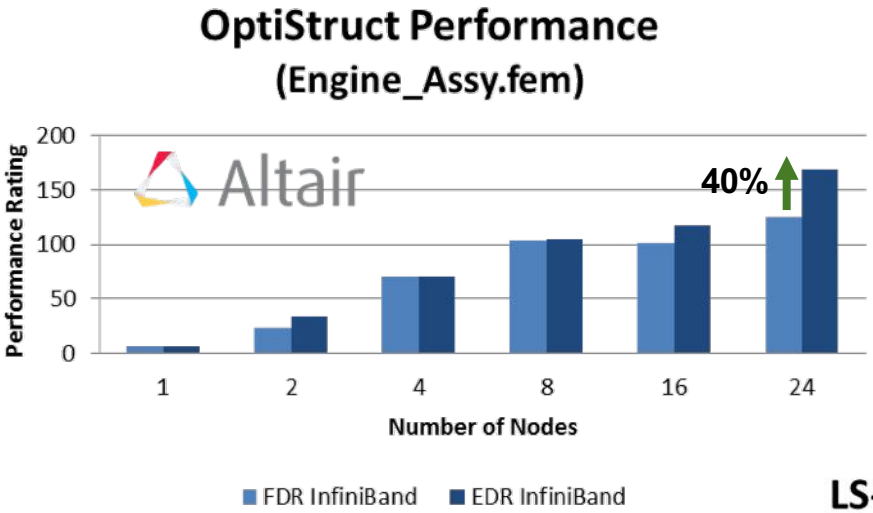


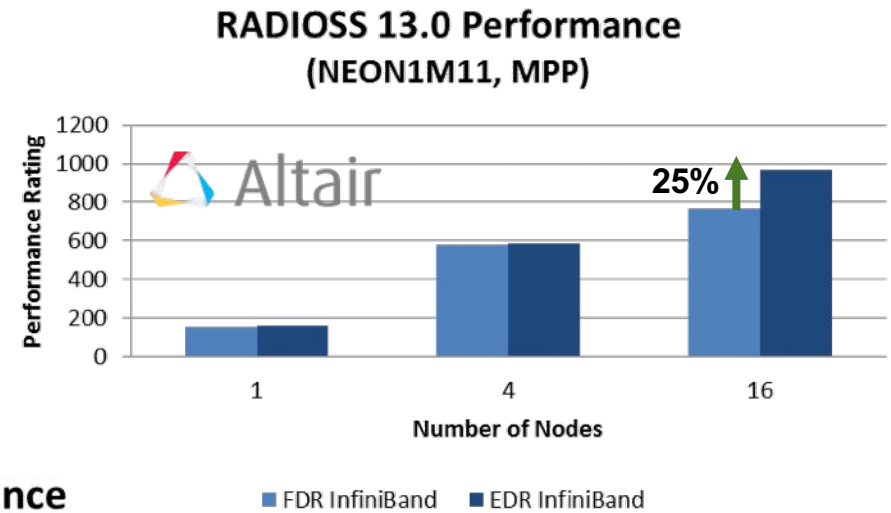
Silicon Photonics AOCs

LinkX™

Making 100Gb/s Deployments as Easy as 10Gb/s

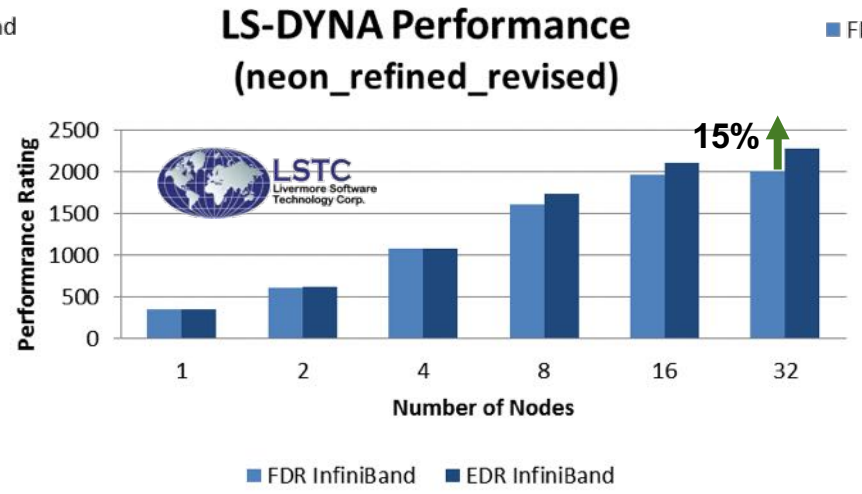
EDR InfiniBand Performance Leadership





ConnectX-4

SwitchIB



HPC-X™

For all graphs: higher is better

Introducing Switch-IB 2 World's First Smart Switch

Introducing Switch-IB 2 World's First Smart Switch

Switch-IB™ 2

World's First Smart Switch

Build for Scalable Compute and Storage Infrastructures

10X Higher Performance with The New Switch SHArP Technology

- The world fastest switch with <90 nanosecond latency
- 36-ports, 100Gb/s per port, 7.2Tb/s throughput, 7.02 Billion messages/sec
- Adaptive Routing, Congestion Control, support for multiple topologies

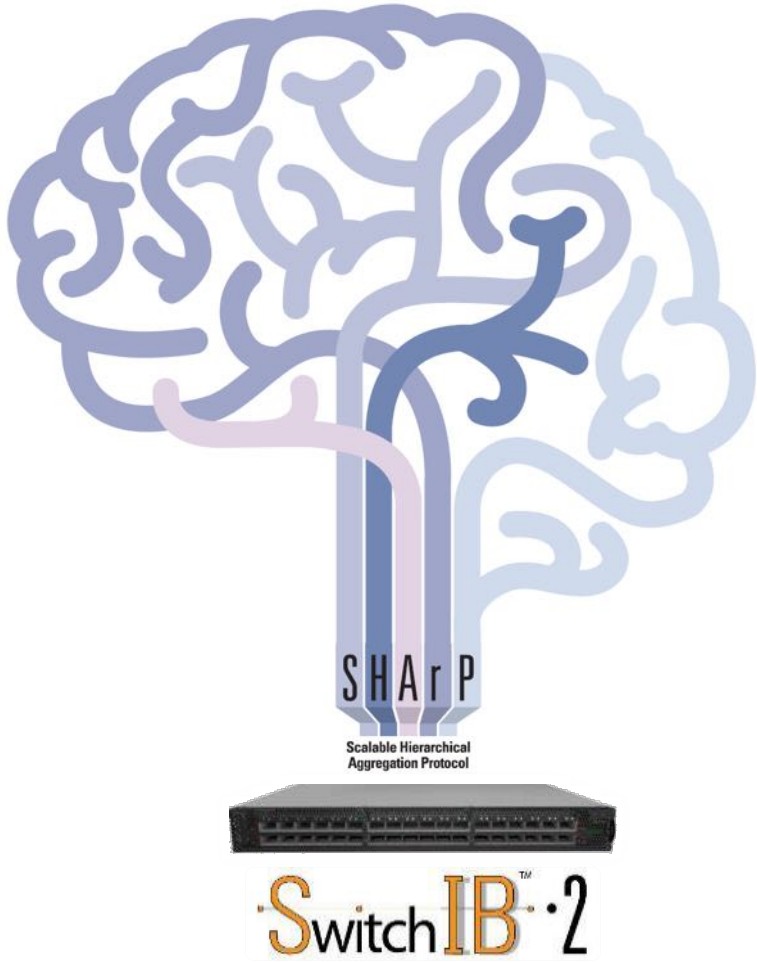
SHArP
Scalable Hierarchical Aggregation Protocol

SHArP (Scalable Hierarchical Aggregation Protocol) Technology

SHArP Enables Switch-IB 2 to Manage and Execute MPI Operations in the Network

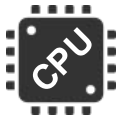
Switch-IB 2 Enables the Switch Network to Operate as a Co-Processor

Delivering 10X Performance Improvement for MPI and SHMEM/PAGS Applications



The Intelligence is Moving to the Interconnect

Communication Frameworks (MPI, SHMEM/PGAS)



Applications

ConnectX-4

Transport
RDMA
SR-IOV
Collectives
Peer-Direct
GPUDirect
More…

SwitchIB-2

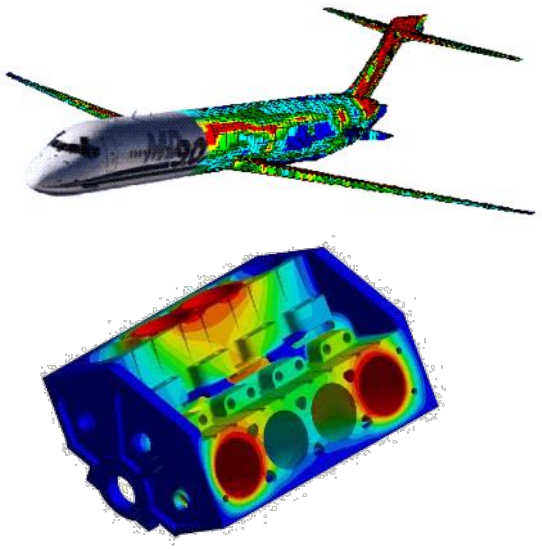
MPI / SHMEM Offloads

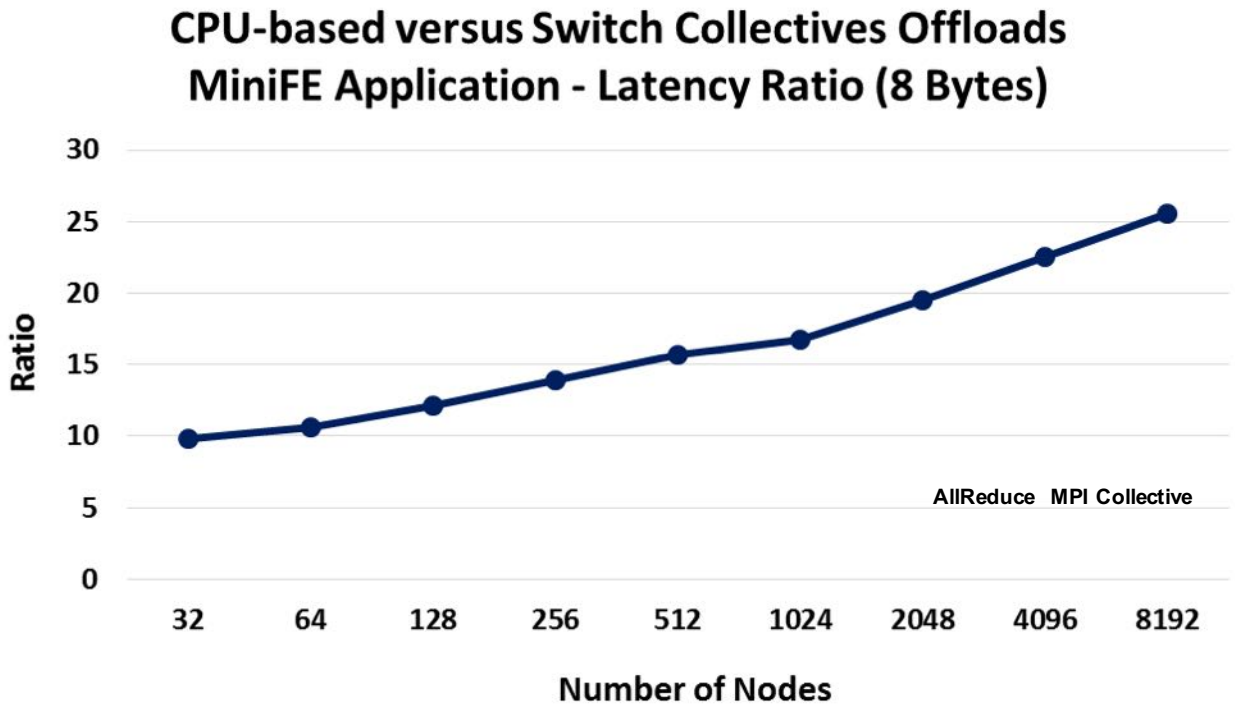
The Only Approach to Deliver 10X Performance Improvements

SHArP Performance Advantage – MiniFE

MiniFE is a Finite Element mini-application

Implements kernels that represents implicit finite-element applications





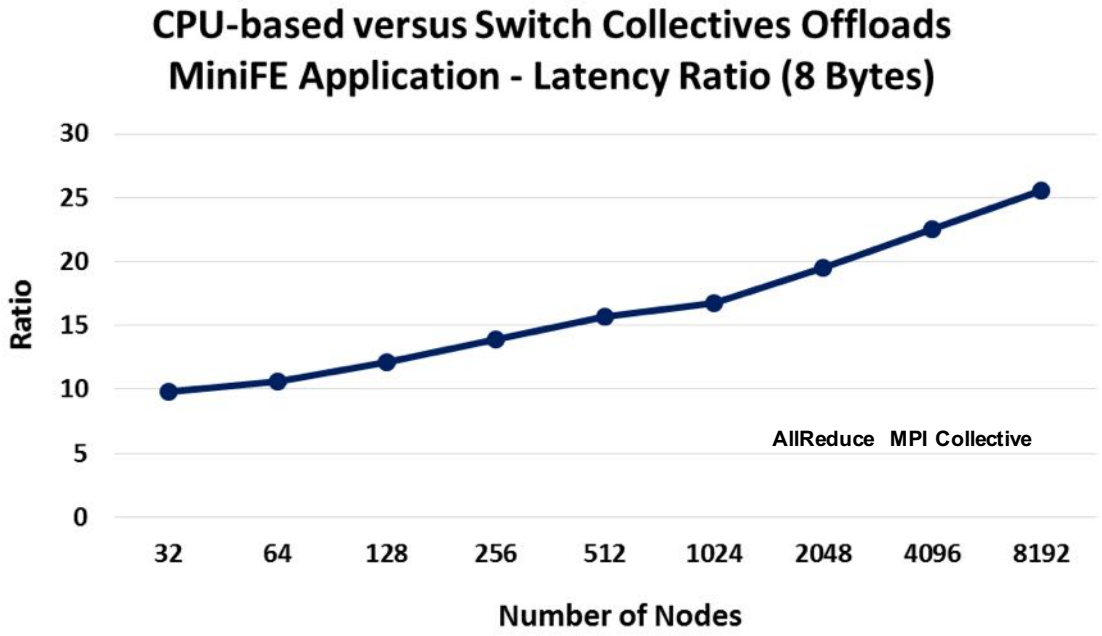
10X to 25X Performance Improvement

SHArP Performance Advantage – MiniFE Details

MiniFE is a Finite Element mini-application

Implements kernels that represents implicit finite-element applications

Number of Nodes	CPU-based Latency (usec)	SHArP-based Latency (usec)	Ratio
32	41.7	4.24	9.9
64	49.08	4.63	10.6
128	57.67	4.76	12.1
256	67.76	4.87	13.9
512	79.62	5.09	15.6
1024	93.55	5.58	16.8
2048	109.92	5.63	19.5
4096	129.16	5.73	22.5
8192	151.76	5.94	25.5



10X to 25X Performance Improvement

The Ever Growing Demand for Higher Performance

Performance Development

Terascale

Petascale

1st

"Roadrunner"

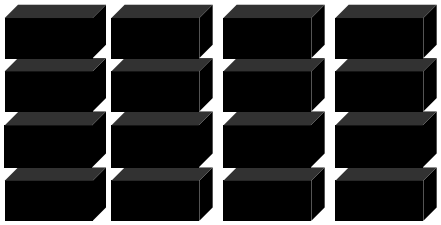
Exascale

OAK RIDGE National Laboratory
"Summit" System

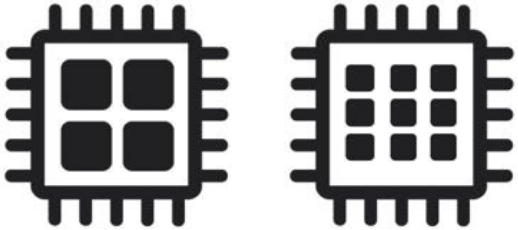
Lawrence Livermore National Laboratory
"Sierra" System

2000 2005 2010 2015 2020

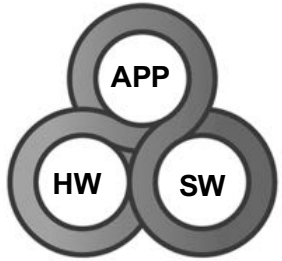
The Interconnect is the Enabling Technology



SMP to Clusters



Single-Core to Many-Core

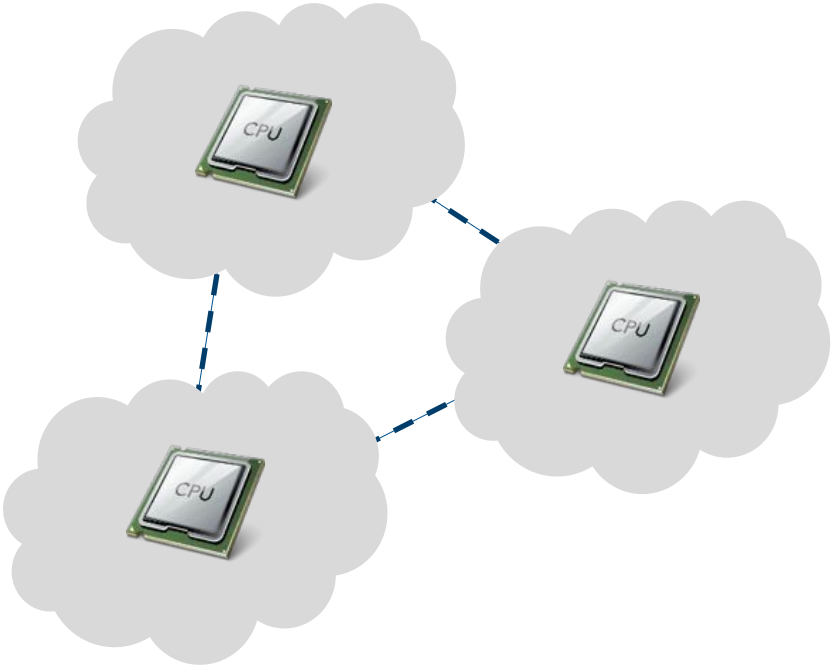


Application
Software
Hardware

Co-Design

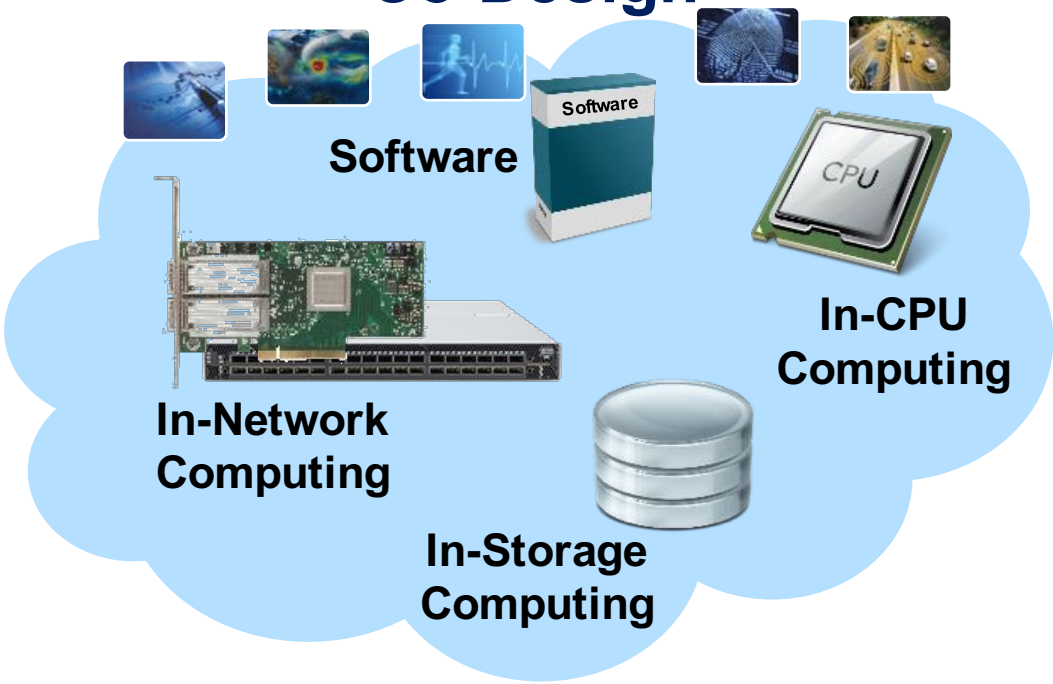
Co-Design Architecture to Enable Exascale Performance

CPU-Centric



Limited to Main CPU Usage
Results in Performance Limitation

Co-Design



Creating Synergies
Enables Higher Performance and Scale

Mellanox InfiniBand Solutions Deliver Highest ROI for Any Scale

Smart Network For Smarter Systems
RDMA, Acceleration Engines, Programmability

Higher Performance

100 Gb/s Link Speed → 200 Gb/s Link Speed

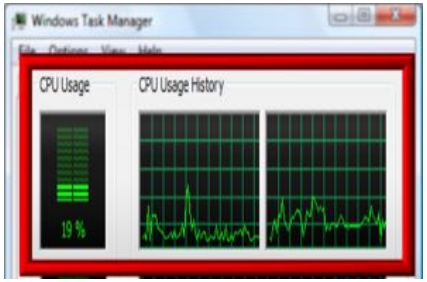
Mellanox delivers the HIGHEST return on investment for ANY scale deployment!

Protect Your Future

Power Consumption	Message Rate	Switch Latency	Scalability
Per Switch Port			CPU efficiency
25% Lower	44% Higher	20% Lower	2X Higher

Mellanox Solutions

Offload Technology



Proven Solutions

Eco-System

Standard

Speed-Up Your Present, Protect Your Future

Mellanox Delivers Best Interconnect

Higher Performance

- 100Gb/s throughput at 0% CPU utilization
- Adapter: 150 Million messages/sec on today's systems, 44% higher
- Switch: 7.02 Billion messages/sec (195 Million per port)
- 20% lower switch latency, with deterministic latency!

Lower TCO

- 25% lower power consumption per switch port
- Standard-based solutions, large eco-system support
- Backward and future compatibility – protect investments
- Offloading Architecture (RDMA, GPUDirect etc.) delivers highest system efficiency

Higher Reliability

- 1,000X higher reliability - Mellanox delivers Bit Error Rate of 10^{-15} versus 10^{-12}
- Superior signal integrity
- Support for Multiple data integrity mechanisms (FEC[1], LLR[2], COD[3] and more)

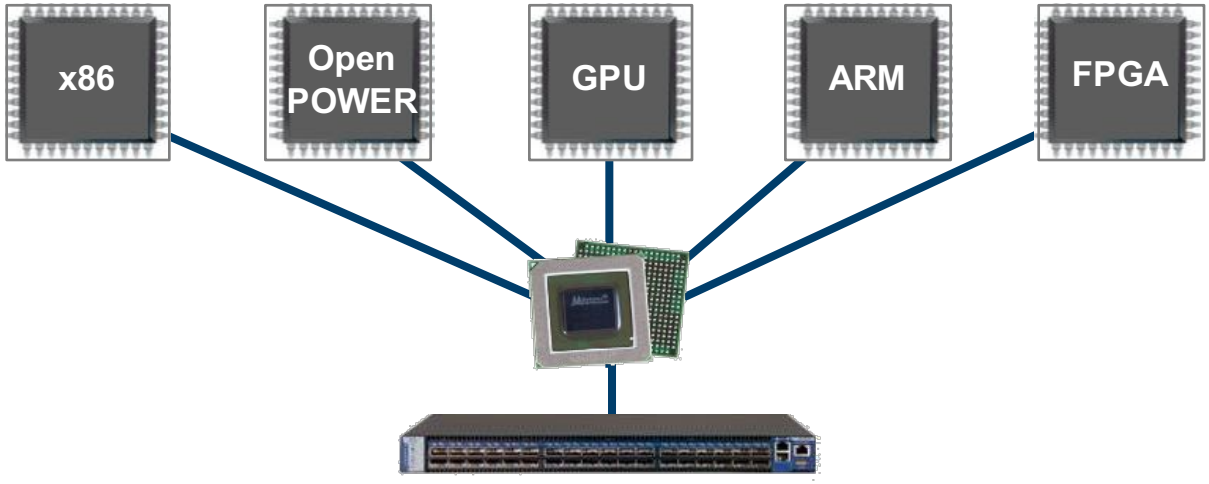
1 - Forward Error Correction; 2 - Link Level Retransmission; 3 - Correction on Demand

End-to-End Interconnect Solutions for All Platforms

Highest Performance and Scalability for

x86, Power, GPU, ARM and FPGA-based Compute and Storage Platforms

10, 20, 25, 40, 50, 56 and 100Gb/s Speeds



Smart Interconnect to Unleash The Power of All Compute Architectures

Thank You

Mellanox TECHNOLOGIES

Connect. Accelerate. Outperform.™