Paving The Road to Exascale Computing

brandon@mellanox.com

HPC User Forum

April 16, 2012
Connectivity Solutions for HPC Computing

Mellanox Interconnect Networking Solutions

<table>
<thead>
<tr>
<th>ICs</th>
<th>Adapter Cards</th>
<th>Host/Fabric Software</th>
<th>Switches/Gateways</th>
<th>Cables</th>
</tr>
</thead>
<tbody>
<tr>
<td>![ICs Image]</td>
<td>![Adapter Cards Image]</td>
<td>![Host/Fabric Software Image]</td>
<td>![Switches/Gateways Image]</td>
<td>![Cables Image]</td>
</tr>
</tbody>
</table>
Introducing FDR 56Gb/s InfiniBand Solutions

- 2002 10Gb/s
- 2005 20Gb/s
- 2008 40Gb/s
- 2011 56Gb/s

Highest Performance, Reliability, Scalability, Efficiency
Mellanox Virtual Protocol Interconnect Solutions for PCIe 3.0

- **ConnectX-3**
  - Industry leader
  - PCI Express 3.0
  - Dual-port FDR IB or 40GbE
  - Native RDMA
  - CORE-Direct

- **SwitchX™**
  - Industry leader
  - 36 x FDR IB or 40GE 64 x 10GbE
  - Integrated routers and bridges
  - 4Tbit switching capacity
  - Ultra-low latency
  - Switch systems: from 36-port to 648-port

**Applications**
- Networking
- Storage
- Clustering
- Management

**SW Acceleration Products**
- LOM
- Adapter Card
- Mezzanine Card

**Unified Fabric Manager**

**Switch OS Layer**

**Networking**
- Setup
- System
- Security
- Ports
- Fabric MGMT
- Fabric I/O
- Status
FDR InfiniBand New Features and Capabilities

### Performance / Scalability
- >12GB/s bandwidth, <0.7usec latency
- PCI Express 3.0
- InfiniBand Router and Ethernet Bridging

### Reliability / Efficiency
- Link bit encoding – 64/66
- Forward Error Correction
- Lower power consumption
Double the Bandwidth, Half the Latency
120% Higher Application ROI
FDR InfiniBand in Multiple Deployments

187 Tflops with Only 648 Nodes
#54 on the TOP500 list

And More!

FDR InfiniBand is The New Deal!
FDR 56Gb/s InfiniBand Solutions Portfolio

Modular Switch
- SX6500
  - Up to 648 ports FDR

Edge Switch
- SX60XX
  - 1U 36 port
  - SX6036 – 36 ports managed
  - SX6025 – 36 ports externally managed

Adapters
- Single and dual port FDR adapter card

Management
- Mellanox M-1 E-2-E Cluster Support Services

Cables
- UFM Diagnostics
  - Virtual Protocol Interconnect

Feature License Keys
- Mellanox M-1 E-2-E Cluster Support Services
Accelerating Big Data Analytics

EMC 1000-Node Analytic Platform
Accelerates Industry's Hadoop Development
24 PetaByte of physical storage
Mellanox VPI Solutions (FDR InfiniBand, 40GbE)

• High Bandwidth
• Low Latency
• RDMA
• Mellanox UDA
• Mellanox VPI

Expanding HPC Frontiers!

Hadoop Acceleration 2X Faster Hadoop Job Run-Time
MemCACHED Acceleration Up to 13X More Transactions/Sec
Mellanox ScalableHPC for Accelerating Communications

ScableMPI

Mellanox Co-Design
MellanoX Messaging Accelerations (MXM), Collectives Offloads, GASnet, OpenSHMEM, RDMA
The Future of GPU Computing Deployments

Today - Servers with GPUs

- CPU
- GPU

- CPU
- GPU

- CPU
- GPU

- CPU
- GPU

PCle-equivalent performance
- 56Gb/s bandwidth
- 0.7usec latency

RDMA dwarfs overhead
- Maintains local access model
- Supports memory management

GPU as a Service

Time for matrix-matrix product (4096x4096)

- Remote GPU
  - Over Ethernet: 2.28 sec
  - Over InfiniBand: 0.65 sec

- Local GPU: 0.62 sec

- CPU only (MKL): 1.30 sec
UFM - Unified Fabric Manager

- Monitor & analyze fabric performance
- Real-time fabric-wide health monitoring
- Innovative congestion analysis
- Threshold based alerts
- Scales out to Tens of Thousands of nodes
Thank You
HPC@mellanox.com

PAVING THE ROAD TO EXASCALE

ADVANCING NETWORK PERFORMANCE, EFFICIENCY, AND SCALABILITY.