Services on Advanced Networks

The storage angle

Michael Kagan

Chief Technology Officer

Mellanox Technologies

May 2010
Connectivity Solutions for Efficient Computing

Cloud Computing

High-Performance Computing

Enterprise and Storage

Providing Users Best-in-Class Networking Bandwidth, Application Performance and Response Time
**Leading connectivity solutions provider for data center servers and storage systems**

- Foundation for the world’s most powerful and energy-efficient systems
- >5.8M ports shipped as of March ‘10

**Company headquarters:**

- Yokneam, Israel; Sunnyvale, California
- 375+ employees; worldwide sales & support

**Solid financial position**

- Record Revenue in Q1’10; $36.2M
- Record Revenue in FY’09; $116.0M
- $220.6M cash / no debt
Leading End-to-End Connectivity Solution Provider for Servers and Storage Systems

Server / Compute

Switch / Gateway

Storage Front / Back-End

Virtual Protocol Interconnect

40G IB & FCoIB
10/40GigE & FCoE

40G InfiniBand
10/40GigE
Fibre Channel

Mellanox Interconnect Networking Solutions

ICs | Adapter Cards | Host/Fabric Software | Switches/Gateways | Cables
---|---------------|----------------------|-------------------|-------
![ICs](image1.png) | ![Adapter Cards](image2.png) | ![Host/Fabric Software](image3.png) | ![Switches/Gateways](image4.png) | ![Cables](image5.png)

Leadership in Network Solutions
Networking in Storage Applications: Front End and Back End Connectivity

- **Front End** uses Ethernet (iSCSI), Fibre Channel, InfiniBand, FCoE
- **Back End** uses proprietary, IB, Ethernet, RapidIO, PCIe etc.

*Leadership in Network Solutions*
## Storage Front End Connectivity Solutions Comparison

<table>
<thead>
<tr>
<th>Features</th>
<th>10GigE</th>
<th>10GigE DCB</th>
<th>FC</th>
<th>InfiniBand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth (BW)</td>
<td>10Gb/s</td>
<td>10Gb/s</td>
<td>8Gb/s</td>
<td>40Gb/s</td>
</tr>
<tr>
<td>Raw BW (Unidirectional)</td>
<td>1250MB/s</td>
<td>1250MB/s</td>
<td>800MB/s</td>
<td>4000MB/s</td>
</tr>
<tr>
<td>Reliable Service</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Fabric Consolidation</td>
<td>Yes (iSCSI)</td>
<td>Yes (FCoE)</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Copper Distance</td>
<td>10GBase-CX4 15m</td>
<td>10GBase-CX4 15m</td>
<td>15m</td>
<td>SDR 20m</td>
</tr>
<tr>
<td></td>
<td>10GBase-T 100m</td>
<td>10GBase-T 100m</td>
<td></td>
<td>DDR 10m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>QDR 7m</td>
</tr>
<tr>
<td>Optical Distance</td>
<td>10GBase-SR 300m</td>
<td>10GBase-SR 300m</td>
<td>100m</td>
<td>SDR 300m</td>
</tr>
<tr>
<td></td>
<td>10GBase-LRM 220m</td>
<td>10GBase-LRM 220m</td>
<td></td>
<td>DDR 150m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>QDR 100m</td>
</tr>
</tbody>
</table>

InfiniBand delivers the highest performance, most mature converged I/O solution

FCoE is emerging and high growth converged I/O technology over DCB (lossless) Ethernet

Leadership in Network Solutions
# Back End Connectivity Solutions Comparison

<table>
<thead>
<tr>
<th>Features</th>
<th>GigE</th>
<th>10GigE</th>
<th>PCIe</th>
<th>SRIO*</th>
<th>InfiniBand</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bandwidth (BW)</td>
<td>1Gb/s</td>
<td>10Gb/s</td>
<td>2.5GBaud</td>
<td>9.4Gb/s</td>
<td>40Gb/s</td>
</tr>
<tr>
<td>Reliable Service</td>
<td>No</td>
<td>Yes (with DCB)</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Management</td>
<td>Legacy</td>
<td>Legacy</td>
<td>New</td>
<td>New</td>
<td>Legacy</td>
</tr>
<tr>
<td>FE and BE Consolidation</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Latency</td>
<td>50usec+</td>
<td>10usec+ (1.3usec with RoCE)</td>
<td>&lt;1usec</td>
<td>&lt;1usec</td>
<td>&lt;1usec</td>
</tr>
<tr>
<td>Topologies</td>
<td>Any</td>
<td>Any</td>
<td>Tree</td>
<td>Any</td>
<td>Any</td>
</tr>
</tbody>
</table>

InfiniBand and 10GigE (with RoCE**) delivers best overall features

*SRIO = Serial Rapid I/O

**RoCE = RDMA over Converged Ethernet. Standard for RDMA and low latency over 10GigE
InfiniBand Storage Ecosystem

InfiniBand Backend Clustering and Failover

Native InfiniBand Clustered File Storage

Native InfiniBand Block Storage Systems

InfiniBand - Fibre Channel / Ethernet Gateways

Leadership in Network Solutions
Independent Storage Analyst (The Taneja Group)

- Low latency back-end storage clustering using IB
- Isilon storage solutions example

Equivalent Throughput with Half the # of Boxes
Mellanox in Back End Storage Applications: InfiniBand in IBM SONAS Cloud Storage

Clustered NAS Services Scale Out to 10s of 1000s of Users
Interconnects Map

HPC

Eth is just enough
Commodity Clustering

Government HPC
Raw performance
LargeScale

EDC

“low touch” maintenance
IT as business utility

Cloud computing
IT as a competitive advantage

Leadership in Network Solutions
The Trends

Leadership in Network Solutions
Service or Object?
Virtualized IO services

Leadership in Network Solutions
Remote Direct Memory Access (RDMA)

**Send**

- Network client
  - Write (&data, sector)
  - Command
  - Ready
- Disk server
  - Data
  - Done

**RDMA**

- Network client
  - Write (&data, sector)
  - Command, &data
- Disk server
  - Data
  - Done
  - Done
Advanced Network – Capabilities

- **Built-in Transport**

- **Remote Memory Operations**

- **IO Memory Management**

- **Kernel Bypass**

---

**RDMA**

- Network client
- Disk server

- Write (&data, sector)
- Command, &data
- Done

---

Leadership in Network Solutions
Advanced Network – Performance

- Bandwidth
- CPU utilization

Leadership in Network Solutions
One-side brain – dumb IO device
IO services – tunneling

Leadership in Network Solutions
Mellanox BridgeX BX5020: InfiniBand to Ethernet and Fibre Channel Bridging

Uplinks: 4 X 40Gb/s QSFP (IB) ports
Downlinks: 16 SFP+
- 12 1/10 GigE or
- 16 1/2/4/8 Gb/s FC or
- Combination EN and FC ports

Converges LAN, SAN & IPC traffic on a single 40Gb/s server network
“Split brain” – smart remote device
Inside the Oracle Exadata Storage Server Hardware

- Dual-redundant, hot-swappable power supplies
- 24 GB DRAM
- ILOM
- Disk Controller HBA with 512M battery backed cache
- InfiniBand QDR (40Gb/s) dual port card
- 4 x 96GB Sun Flash PCIe Cards
- 12 x 3.5” Disk Drives
- 2 Quad-Core Intel® Xeon® Processors

Pre-installed Software:
- Oracle Exadata Storage Server Software
- Oracle Enterprise Linux
- Drivers

Leadership in Network Solutions
Mellanox In Front End Storage Applications: InfiniBand Front End in Oracle Exadata

- **Data Warehousing**
  - Table Scan time reduction and scaling with table size
  - 50X DB Query throughput
    - InfiniBand front-end storage integration with SSD

```
10 Hour

Table Scan Time

Typical Warehouse

Exadata

1 TB 10 TB 100 TB Table Size

Performance scales with size

“...everybody is using Ethernet, we are using InfiniBand, 40Gbps InfiniBand”
Larry Ellison Keynote at Oracle Openworld
October 14, 2009 San Francisco
```

More Business Insight, Better Decisions
Intalio|Cloud Appliance H1
The first cloud appliance designed for corporate data-centers

**Compute Blade:** 10 x HP ProLiant BL460c G6
- 1 x QuadCore Intel Xeon X5570 2.93GHz (24MB cache)
- 128GB (8 x 16GB) 1333MHz PC3-10600 Registered DDR3
- No embedded storage

**Database Storage:** 10 x HP StorageWorks CD400a Storage Blade
- 64TB (8 x 8TB) 7200RPM 32MB Cache in RAID-6 configuration
- 200,000 IOPS, 1 µsecond latency
- Used for database storage

**File Storage:** 4 x HP StorageWorks 600 Modular Disk System
- 16 LFF SAS or SMT Drives, 16Gb Gb Connectivity
- 16TB in RAID 64 configuration (striped set with dual distributed parity)
- Used for file storage and SAN disk storage

**Fabric Interconnect:** 2 x H3C 4x QDR InfiniBand Switch Module
- 4x4 port pair, 1.3 microradians latency
- 44 internal 4x DDR switch fabric, 6 external 4x DDR switch fabrics
- 16 internal 4x DDR switch fabric, 3 external 4x DDR switch fabrics

**Configuration Overview:**
- 24 CPUs
- 256 CPU Cores
- 4.0TB Memory
- NVMe 800 Database Storage

**Leadership in Network Solutions**
Elastic Compute and Storage Services

Small Account Configuration

Block Storage
Used for structured content (database data)

Large Account Configuration

File Storage
Used for unstructured content (documents)

Source: Intalio
Efficient IO services

**Major Airline Company**

- Unified I/O: replacement of dedicated FC fabric
- Fewer components, cables to manage

---

Source: Xsigo

**80% TCO Savings**
IBM Systems and Technology Group
Data Sheet

- **Mellanox InfiniScale IV QDR InfiniBand Switches**—High Performance Computing needs high bandwidth with low latency to get the highest server efficiency and application productivity. Mellanox switches deliver the I/O performance that meets these requirements while providing extensive I/O services such as bandwidth, consolidation and unification, and flexibility.

- **Mellanox ConnectX-2 QDR InfiniBand Host Channel Adapters**—Offered in both single port and dual port, these PCIe 2.0 x8 adapters consume up to 10 percent less power than the previous generation and demonstrate up to 30 percent improved performance, with 1us MPI ping latency and 50 M messages/sec.

- **Force 10 Networks Ethernet Switches**—1 Gb and 10 Gb Ethernet Switches with scalable architectures for High Performance Computing that require wire-speed, low latency performance, with an efficient power and cooling design to save on energy costs.

- **Mellanox 10 Gb Ethernet Adapters**—This “ConnectX 2” [dual-port] technology supports PCIe 2.0 x8 (5 GT/s), enabling high availability and high performance. ConnectX EN improves network performance by increasing available bandwidth to the CPU and providing enhanced performance.
Computing as a Service
Storage as a Service
Interconnect Makes the Difference
Unified I/O and Bridging Connectivity

- Converged IPC, LAN and SAN I/O on servers
  - Using InfiniBand or Ethernet

- Ethernet and FC interfaces on servers
  - Bridges encapsulate and de-encapsulate
  - vNICs and vHBAs using EoIB, FCoIB and FCoE