InfiniBand Update

Addressing new I/O challenges in HPC, Cloud, and Web 2.0 infrastructures

Brian Sparks

IBTA Marketing Working Group Co-Chair

IBTA & OFA Update

- IBTA today has over 50 members; OFA has over 35
    - All of the key enterprise IT and HPC vendors in the RDMA community represented

- InfiniBand current speed now at FDR (Fourteen Data Rate) 56Gb/s
    - Roadmap to EDR (Enhanced Data Rate) 100Gb/s in 2013

- Traction of RDMA technologies in HPC with increased enterprise, Web 2.0 and cloud market adoption

- IBTA Webinars: 4-part series addressing the role of RDMA in the enterprise data center, highlighting the rising importance of I/O technology in the on-going transformation of modern data centers

- Product testing with UNH-IOL: Interoperability List is growing steadily and now includes more than 430 products

- Annual OFA workshop planned for March 25-28, 2012 in Monterey

HPC and Data Centers Demand RDMA Networks

Scale-Out Clustering   Data Warehousing   Financial Services   Cloud Computing   Web 2.0

- Increasing networking I/O requirements
  - Driven by multi-core CPUs, GPUs, virtualization
  - LAN, SAN, IPC I/O convergence

- Enabling green data center efficiency and scalability
  - Maximize the use of each server; deterministic performance; meet SLAs
  - Elastic storage scalability to meet exponential user and data volume growth
  - Decrease energy consumption, cable sprawl, and space

- High bandwidth and low-latency I/O is critical to ROI
  - Accelerate data throughput, application response and job execution time, increase utilization

InfiniBand & OFED Connectivity Benefits for IT*

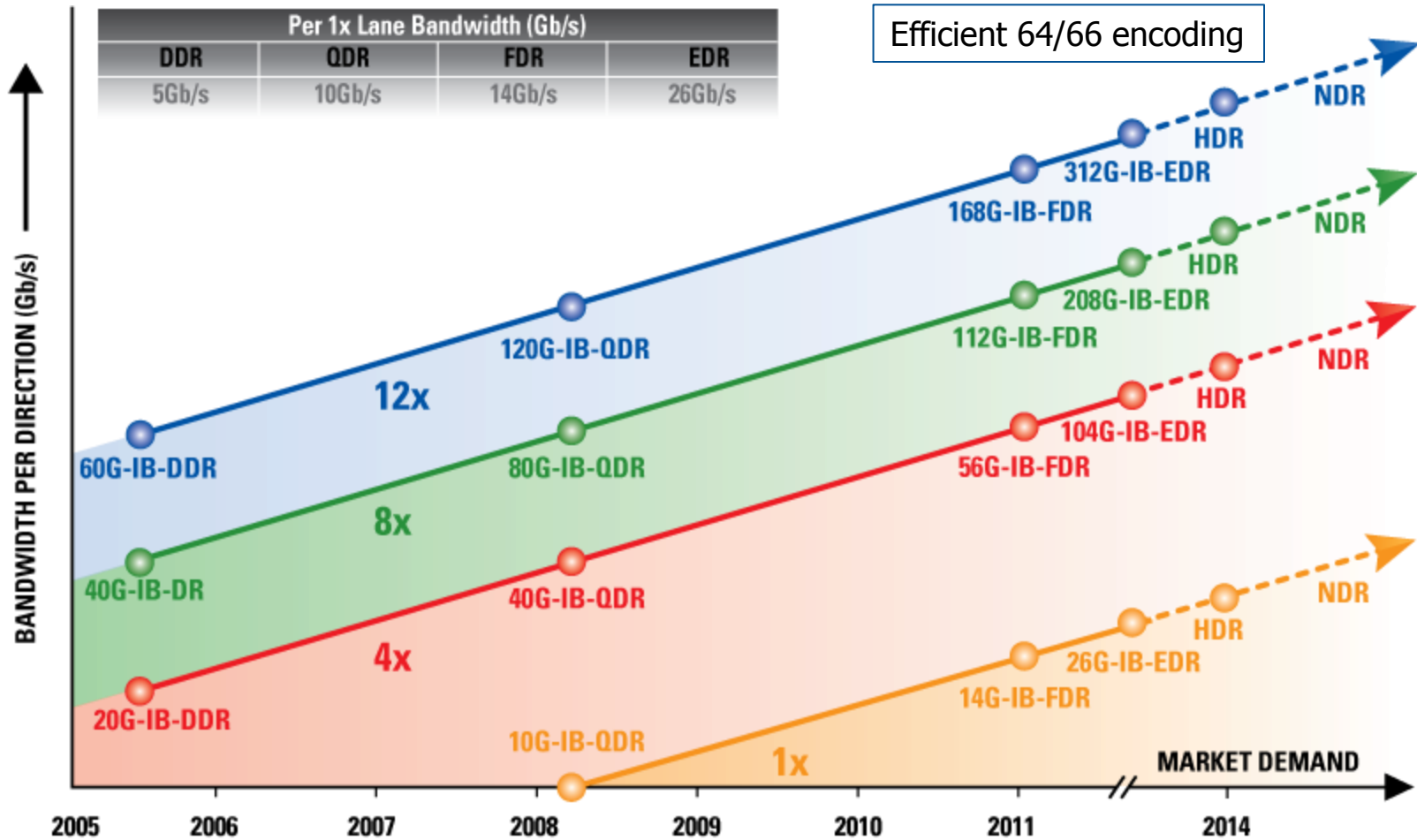* Based on end user testimonies

| Infrastructure Reduction | 60% | Energy Cost Reduction | 65% | Performance Increase | 10X |

InfiniBand Roadmap

InfiniBand in the TOP500

- **InfiniBand is the dominant I/O technology for high-performance computing**
  - 210 clusters, 42% of the list, InfiniBand demonstrates CAGR of 15% 2008-2011
  - The first appearance of FDR 56Gb/s InfiniBand – 2 systems on the list

- **InfiniBand is the interconnect of choice for large scale systems**
  - InfiniBand connects the best power efficient system in the TOP10 (Petaflop systems) – 1.7x better versus TOP10 average
  - InfiniBand connects the most efficient system on the TOP500 – nearly 96% system efficiency
  - The InfiniBand connected CPUs grew 24% YoY, The overall InfiniBand based systems' performance grew 44% YoY

- **InfiniBand connects the most powerful clusters - TOP10, TOP20**
  - 5 of the TOP10 (#4, #5, #7, #9, #10), 8 of the TOP20 (#14, #15, #18)
  - 50% (5 systems out of 10) of the world sustained Petaflop systems

- **The most used I/O solution in the TOP100, TOP200, TOP300, TOP400**
  - Connects 55% (55 systems) of the TOP100 while Ethernet only 2% (2 system)
  - Connects 59.5% (119 systems) of the TOP200 while Ethernet only 14.5% (29 systems)
  - Connects 50.7% (152 systems) of the TOP300 while Ethernet only 29% (88 systems)
  - Connects more systems in the TOP400 versus Ethernet or any other technology – 46.8% (187 systems)

- **InfiniBand is the interconnect of choice for GPU-based systems**
  - 92% of the GPU based systems are connected with InfiniBand
  - 37 GPU based systems on the TOP500 list, 34 are InfiniBand connected, 3 proprietary

InfiniBand's Unsurpassed System Efficiency

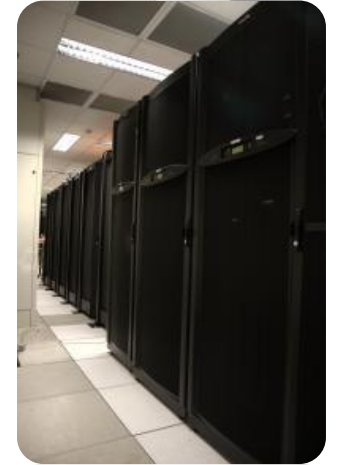## World Leading Compute Systems Efficiency Comparison
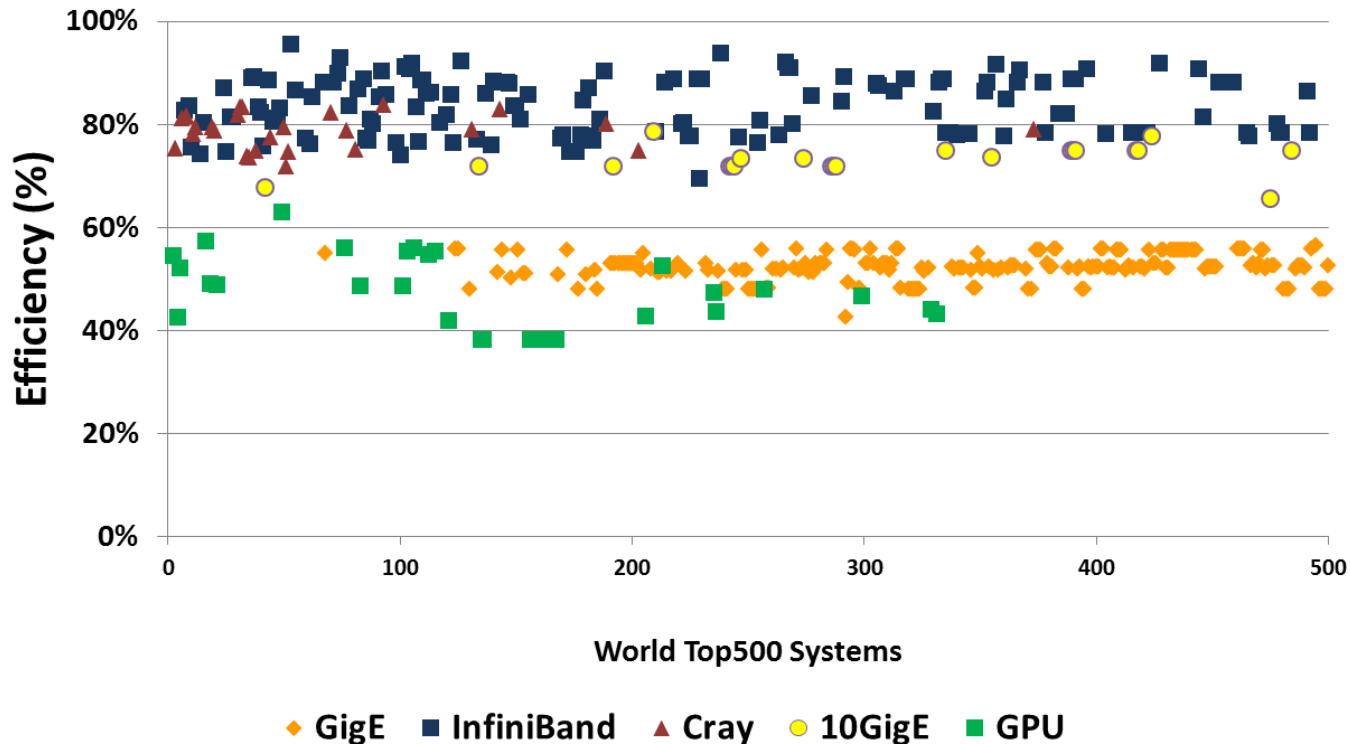
- TOP500 systems listed according to their efficiency
- InfiniBand is the key element responsible for the highest system efficiency
- Up to 96% efficiency

FDR InfiniBand with PCIe 3.0 In Multiple Deployments

Nearly 110Tflops with Only 648 Nodes
A Top 100 Supercomputer on the TOP500 list

And More!

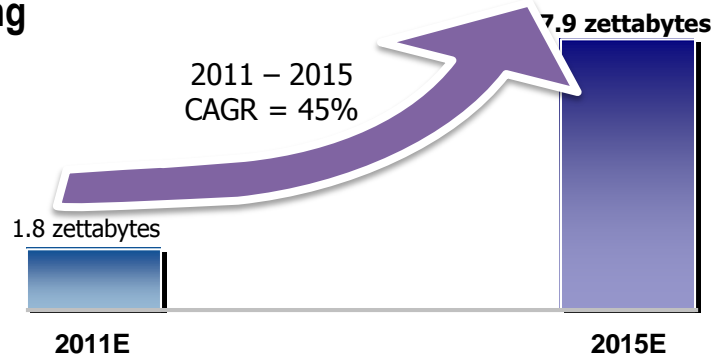# The "Big Data" Opportunity

High-Performance Computing

Web 2.0 and Cloud Computing

Enterprise Data Center

The Digital Universe[1]

2011 − 2015
CAGR = 45%

7.9 zettabytes

1.8 zettabytes

2011E

2015E

Proliferation of Data will be a Catalyst for Growth of RDMA

Accelerating Big Data Analytics

- EMC 1000-Node Analytic Platform

- Accelerates Industry's Hadoop Development

- 24 PetaByte of physical storage

  - *(half of every written word since inception of mankind)*

- FDR InfiniBand
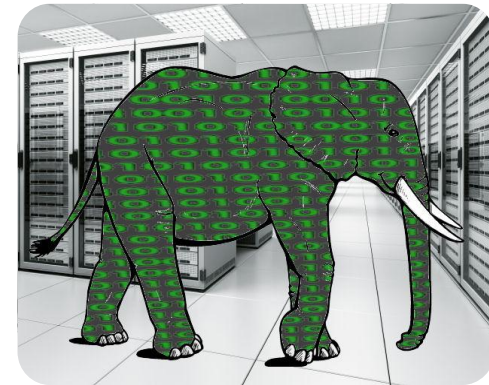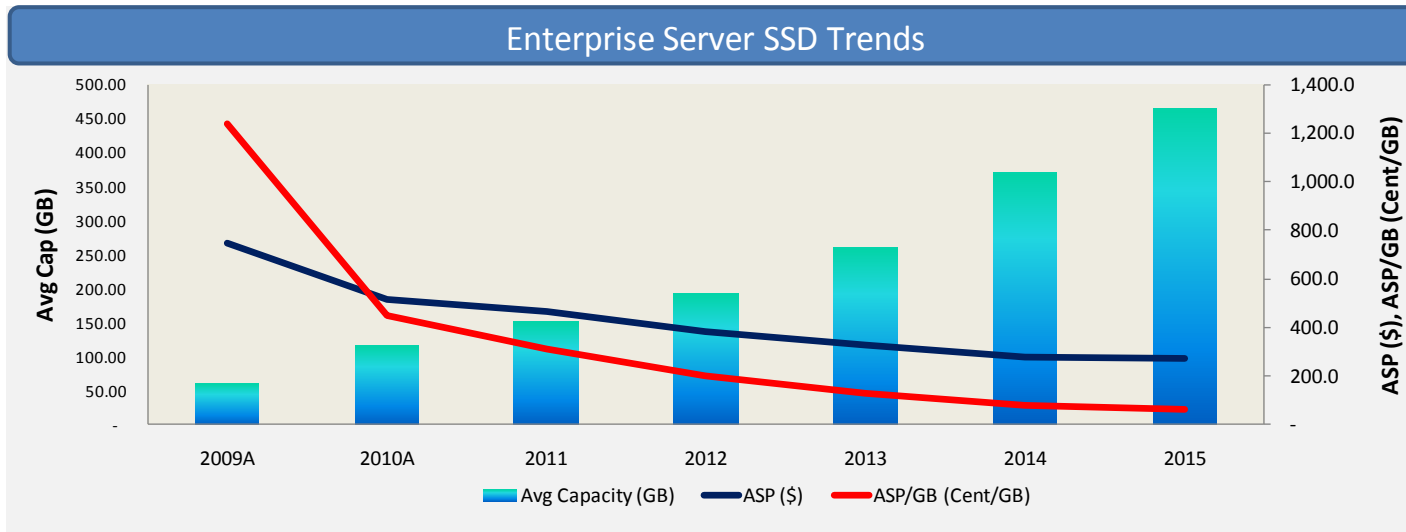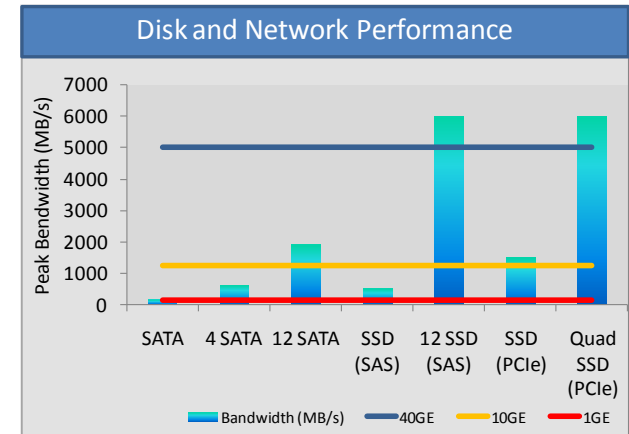
Hadoop Acceleration	*2X Faster Hadoop Job Run-Time*
MemCached Acceleration	*Up to 13X More Transactions/Sec*

I/O Bottlenecks: from Disk to Network

- # Trends
  - – Use multiple SATA drives
  - – SSD is around the corner

- # Matching network speed
  - – 10GbE and beyond

Disk and Network Performance

Peak Bandwidth (MB/s)

SATA | 4 SATA | 12 SATA | SSD (SAS) | 12 SSD (SAS) | SSD (PCIe) | Quad SSD (PCIe)

Bandwidth (MB/s) — 40GE — 10GE — 1GE

Enterprise Server SSD Trends

Avg Cap (GB) — ASP ($), ASP/GB (Cent/GB)

2009A | 2010A | 2011 | 2012 | 2013 | 2014 | 2015

Avg Capacity (GB) — ASP ($) — ASP/GB (Cent/GB)

Solid State Drives Demand InfiniBand I/O Solutions

- Storage applications demand higher storage throughput
- Economics / green datacenters driver I/O consolidation
- SSDs provide the lowest latency and energy efficient solution

Fibre Channel + SSD ~ 45usec
Fibre Channel Eliminates SSD Storage Advantage

RDMA + SSD ~ 26usec
InfiniBand/Ethernet Deliver the SSD Promise!

Component	Latency (usec)
InfiniBand	0.7
Ethernet RDMA (RoCE)	1.3
Ethernet TCP	6
Fibre Channel	20
SSD Read	25
Disk Read	6000

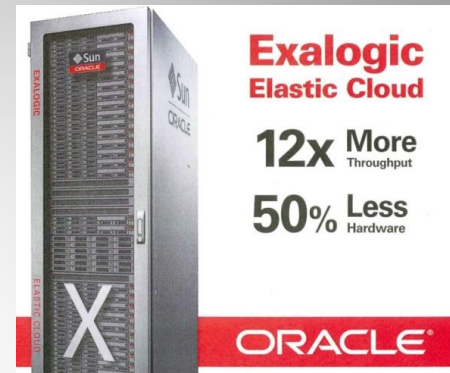Database Server   FDR InfiniBand Or 40GbE   SSD   Disk
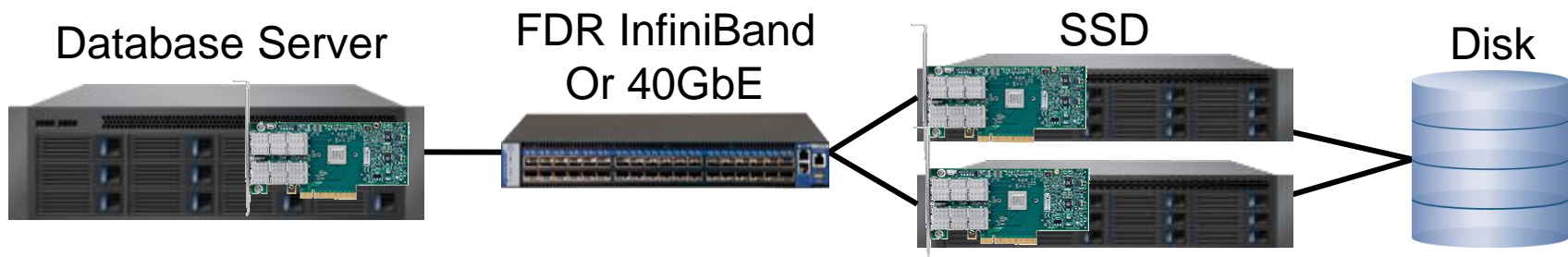
RDMA Delivers the SSD Promise!

RDMA Enabled Enterprise and Cloud

Oracle Exadata

The World's Fastest
Database Machine

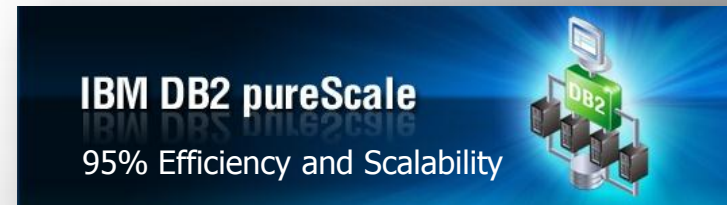*Runs Oracle 10X Faster*

Exalogic Elastic Cloud
12x More Throughput
50% Less Hardware
ORACLE

IBM XIV Storage System

Power to serve even more
applications from a single system

4 Times the Throughput

IBM DB2 pureScale
95% Efficiency and Scalability

RDMA provides leading performance for virtualized server and storage environments, online transaction processing, analytics and cloud computing

Enabling Cost-Effective Accelerated Web 2.0 Services

- **Bing maps**
  - High-performance 40Gb/s InfiniBand-accelerated scalable microsite to support image processing for Bing Maps
    - 10x performance improvement
    - Scaled-out solution less than half the cost of 10GbE

- **StubHub**
  - InfiniBand enables StubHub to dynamically manage and balance loads and bandwidth between application and Web servers
    - $1.5M capex savings on infrastructure alone
    - 50% to 75% reduction in deployment time
    - 75% less space required

SC11 Booth Demo

Transcontinental 40Gpbs Network Collaboration Between Orange Silicon Valley, IBTA and OFA members, and ESnet

Showcasing three breakthrough applications for RDMA over the WAN:

- The ability to move multiple simultaneous video and visualization data streams using newly available 100Gb links

- Low latency and bandwidths approaching theoretical maximums allows a user to access remote data as though it were local

- Fair allocation of bandwidth between multiple streams allows efficient simultaneous sharing of the 40 gigabits of bandwidth for large file transfers, transaction and image data

Thank you Orange Silicon Valley for providing access to the 40Gb/s link, and thanks to all of our OFA and IBTA sponsors for their contributions to make this demo happen!

www.infinibandta.org

www.openfabrics.org

InfiniBand & OFED: Growing Adoption

InfiniBand

- IBTA continues to evolve the InfiniBand standard to meet emerging customer requirements
    - Solves server and storage I/O bottlenecks
- InfiniBand provides end users with superior application response time and enhanced cost/performance benefits
- FDR-based products available today; EDR in early 2013
- Extensive cross-vendor interoperability testing

OFED

- OFED: open-source, RDMA-enabled, transport-independent software stacks for Linux and Windows operating systems
- Growing adoption by vendors and users worldwide: over 44% of Top500, 20% growth year-over-year
- OFED benefits: 10X application performance, 50% hardware costs, wire-speed data transfers, microsecond latencies, 90% server utilization