iSCSI Extensions over RDMA (iSER): A New High-Speed Storage Protocol for Flash Memory

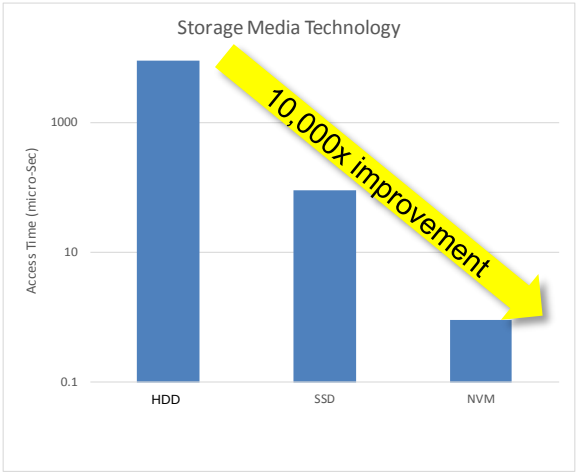
Rob Davis

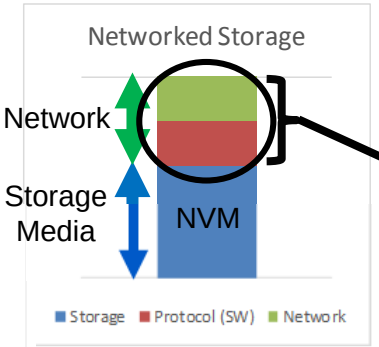
robd@mellanox.com

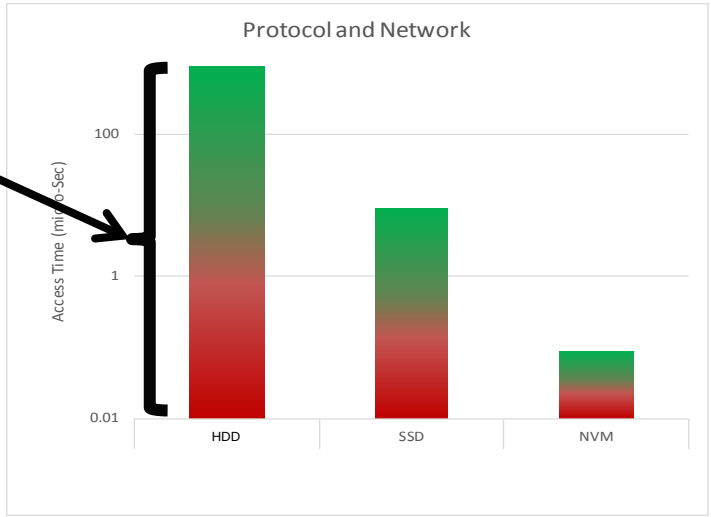
The FASTEST Storage Protocol: iSER
The FASTEST Storage: Flash

- What it is: iSCSI With RDMA Transport
  - Runs over Ethernet or InfiniBand at speeds up to 100Gb/s
  - Works with all applications that support SCSI/iSCSI
- Benefits
  - Lets Flash Shine: Highest bandwidth, Highest IOPs, Lowest Latency
  - iSCSI storage features, management and tools (security, HA, discovery...)
  - Faster than iSCSI, FC, FCoE; Easier to manage than SRP
- Ideal for Flash Storage Applications
  - Latency-sensitive workloads; Small, random I/O
    - Databases, Virtualization, VDI
  - Bandwidth-sensitive workloads; Large, sequential I/O
    - Post production, oil/gas

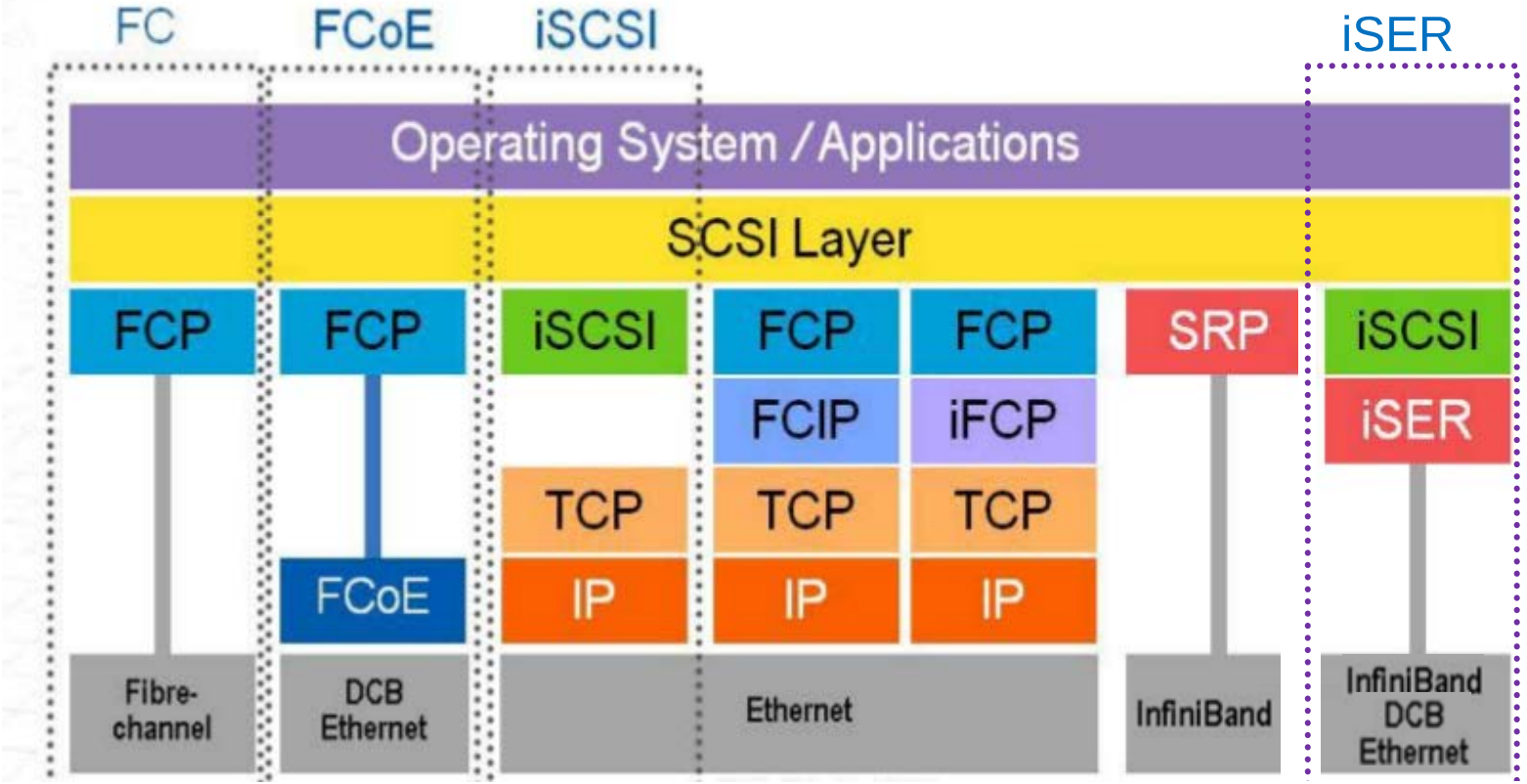
Why iSER? Flash Memory Creates IO Bottleneck



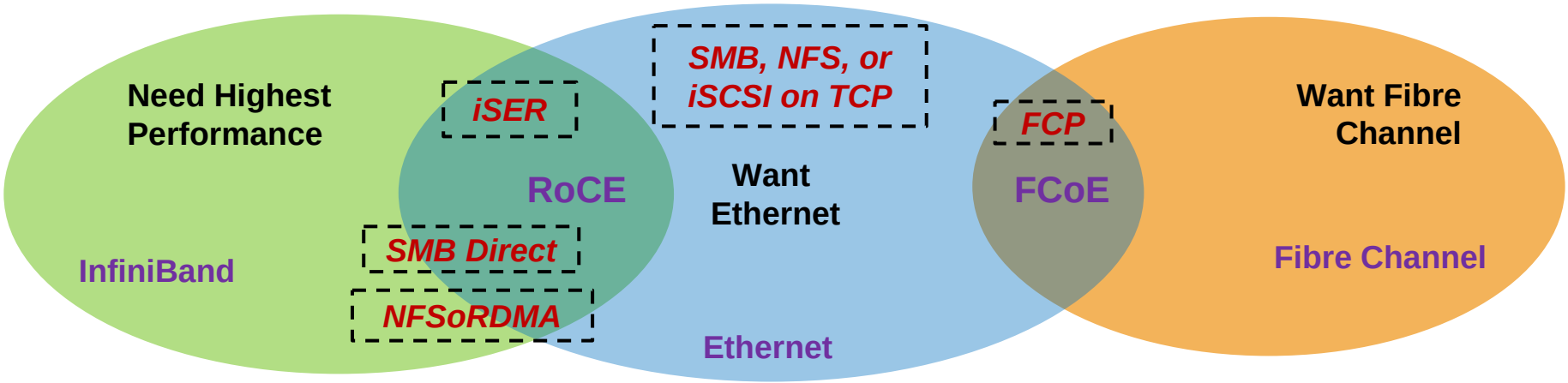




iSER Has No TCP/IP Stack

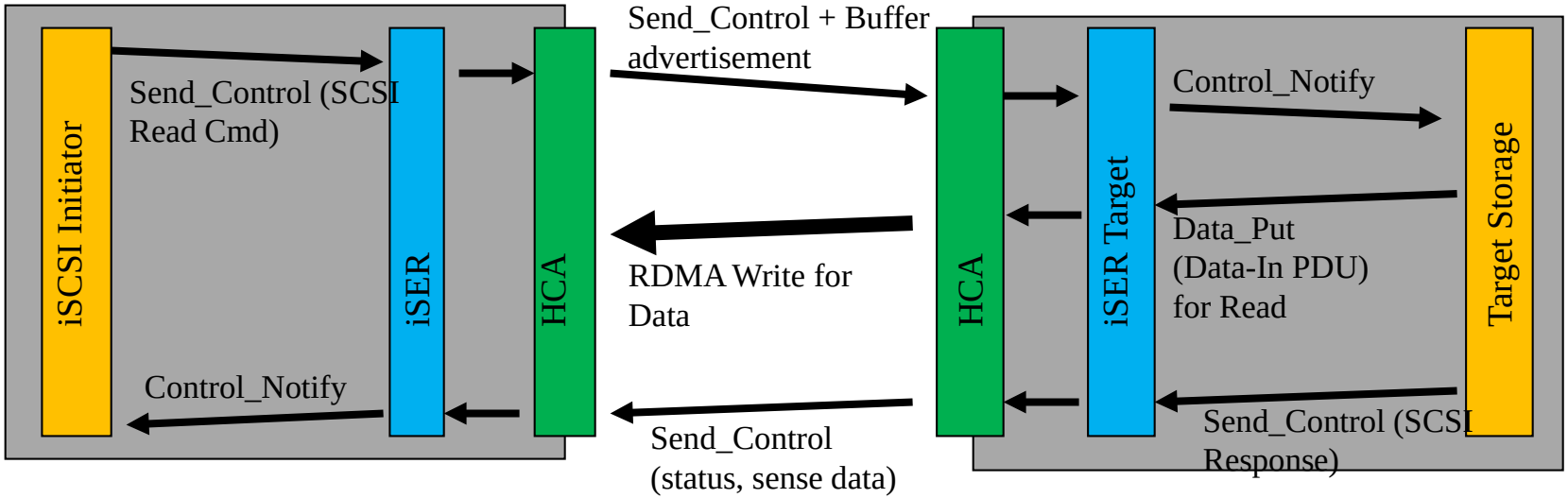


Protocol / Transport Comparison



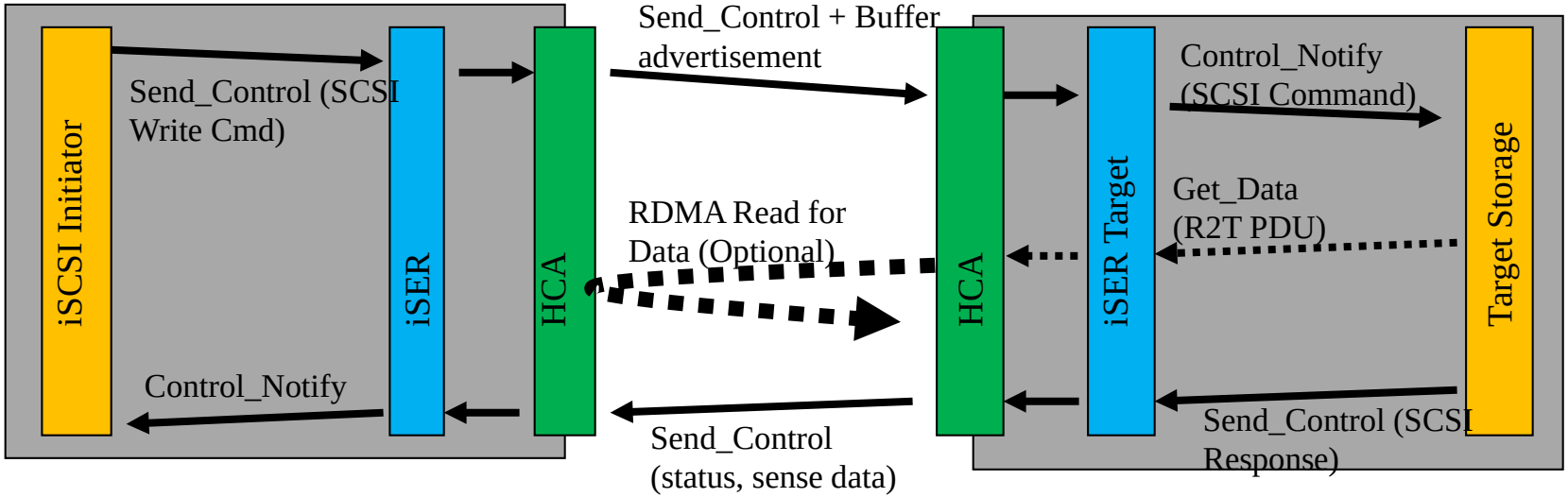
Transport	InfiniBand	Ethernet RoCE	Ethernet TCP	FCoE	FC
Speed	40/56/100 Gb/s	10/25/40/50/100 Gb/s	10/25/40/50/100 Gb/s	10/40 Gb/s	8/16/32 Gb/s
RDMA	Yes	Yes	No	No	No
Routable	Yes	Yes	Yes	No	No

iSER Protocol Overview (Read)



- **SCSI Reads**
  - **Initiator Send Command PDU (Protocol data unit) to Target**
  - **Target return data using RDMA Write**
  - **Target send Response PDU back when completed transaction**
  - **Initiator receives Response and complete SCSI operation**

iSER Protocol Overview (Write)

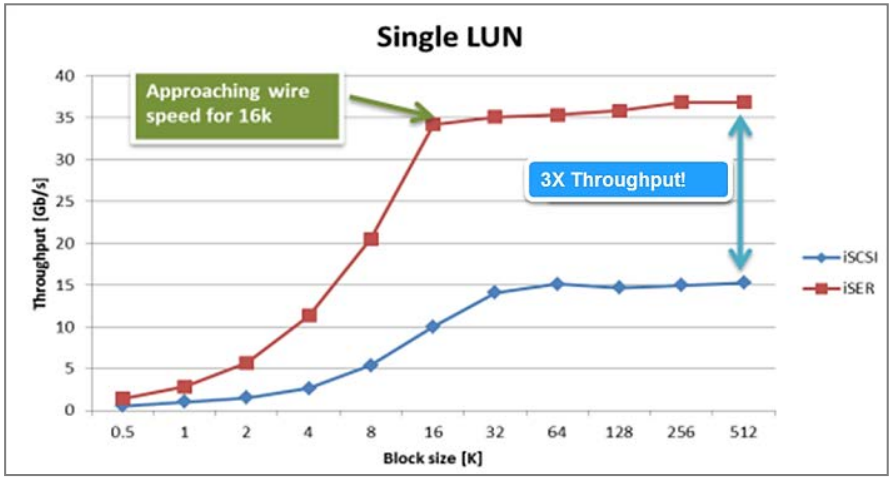


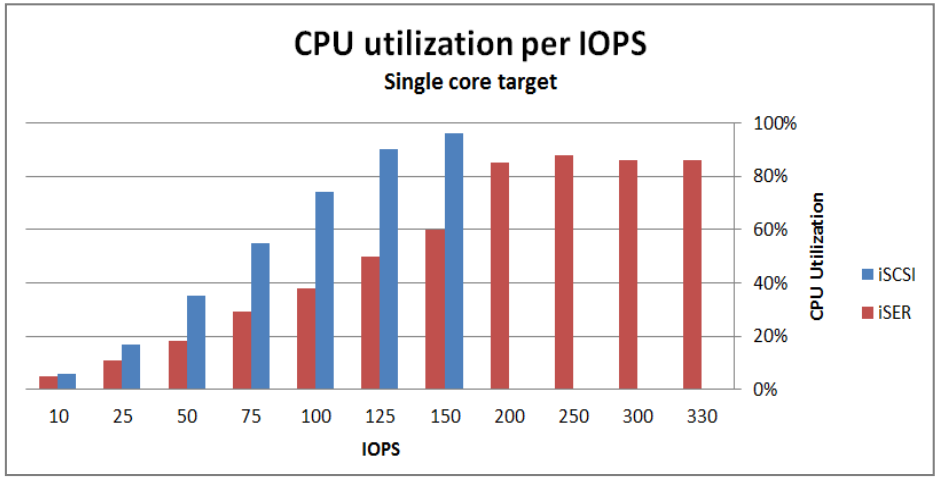
- **SCSI Writes**
  - **Send Command PDU (optionally with Immediate Data to improve latency)**
  - **Map R2T to RDMA Read operation (retrieve data)**
  - **Target send Response PDU back when completed transaction**

Requirements to Deploy iSER

- Application(s) that can use SCSI/iSCSI
  - All applications that use SCSI-based block storage work with iSER
- OS or Hypervisor that Supports an iSER initiator
  - Today: Linux & VMware ESXi, Oracle Solaris
  - Expected soon: Windows, FreeBSD
- iSER Storage Target(unless Hyper Converged)
  - NetApp, HP SL4500, Oracle ZFS, Violin Memory, Zadara, Saratoga Speed
  - Create in Linux using LIO, TGT, or SCST target
- Network that supports RDMA
  - Adapters support InfiniBand, iWARP or RoCE
  - Switches support InfiniBand or Ethernet DCBx with ECN

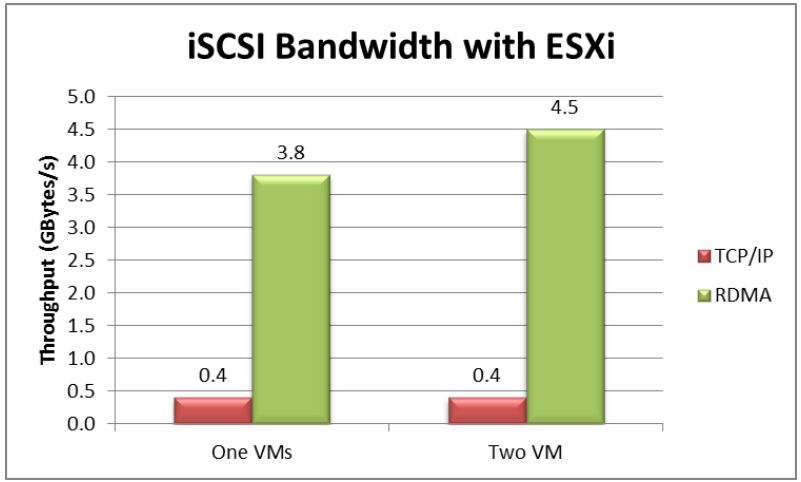
iSER Ethernet Performance



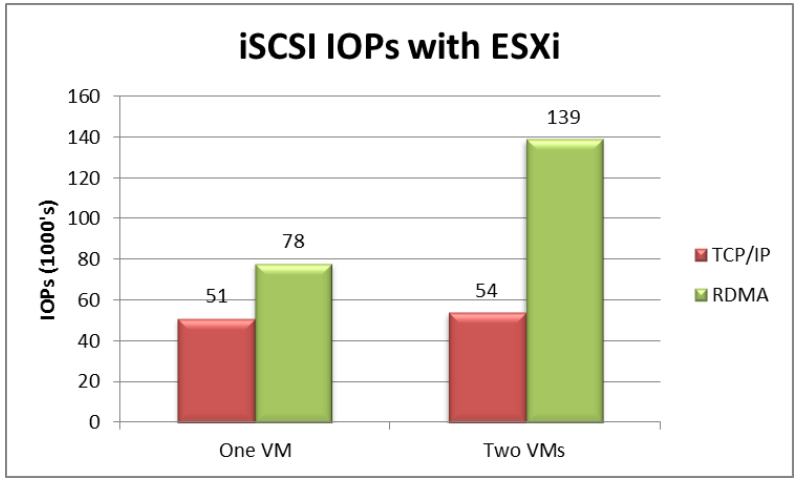


Higher Bandwidth and IOPS with Less CPU Utilization than iSCSI

VMWare: iSER over ESXi



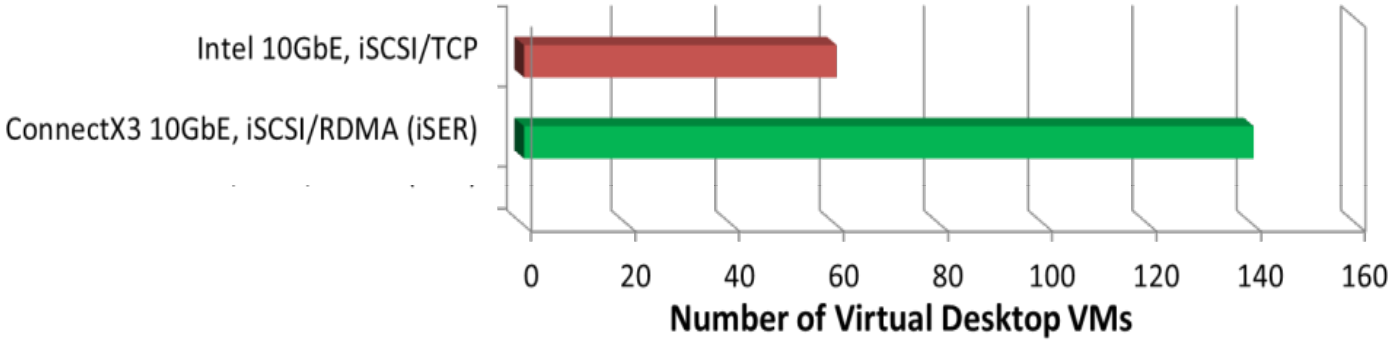
iSER/RDMA has 10X Bandwidth Advantage vs TCP/IP and 2.5X IOPs

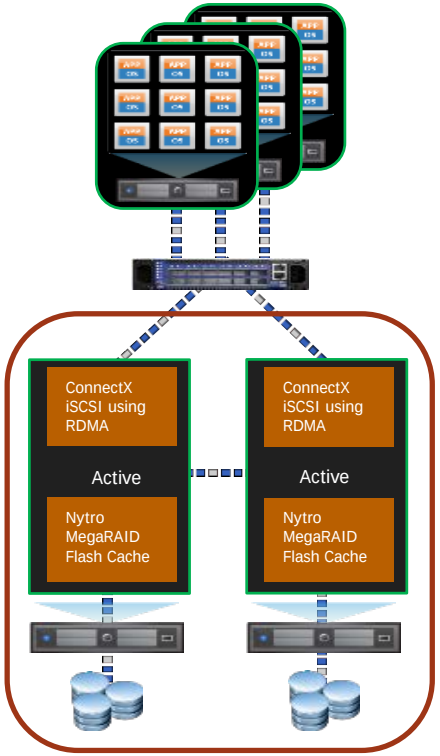


Test Setup: ESXi 5.0, 2 VMs, 2 LUNS per VM

VDI – Real World iSER Application

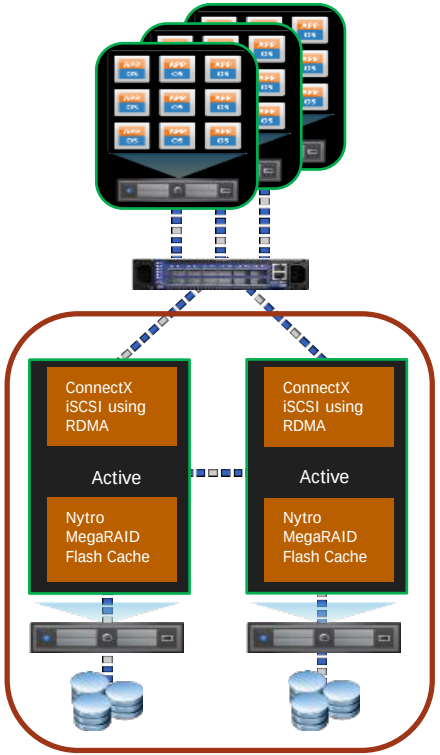
- iSER eliminates storage bottlenecks in VDI deployments
  - iSER accelerates the access to cache over RDMA
  - 140 Virtual desktops with iSER/RoCE vs. 60 virtual desktops over TCP/IP





ROI Comparison Shows iSER Value

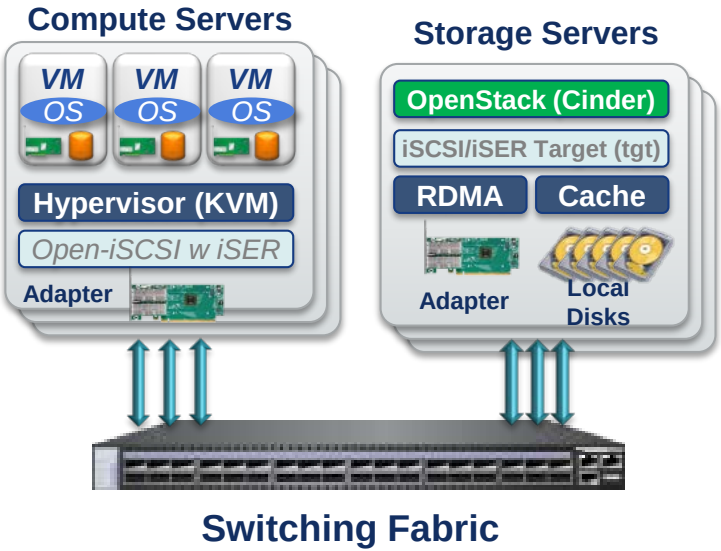
Interconnect	# Virtual Desktop per Server	# Servers	# Switches	CapEx	CapEx per Virtual Desktop
10GbE with TCP/IP	60	84	2	$3,418,600	$684
10GbE with RoCE	140	36	1	$1,855,900	$371

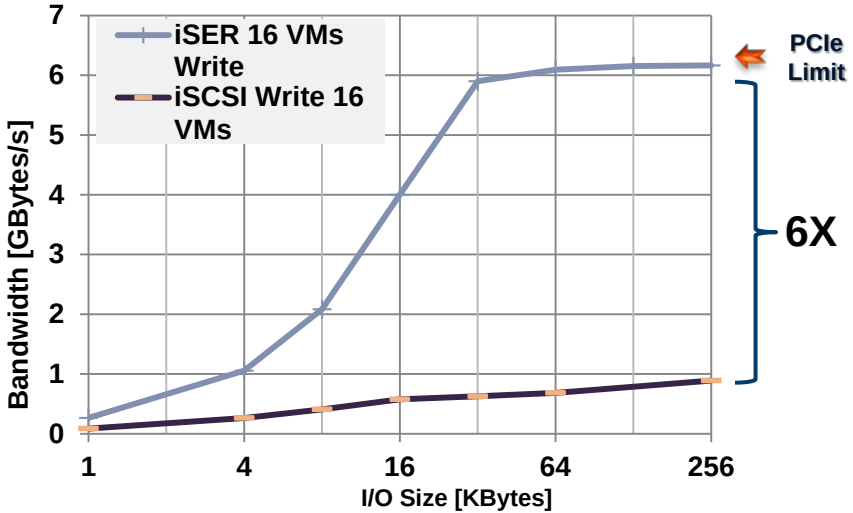


iSER Delivers $1.5M CapEx Savings For VDI Deployments

http://www.mellanox.com/related-docs/whitepapers/SB_Virtual_Desktop_Infrastructure_Storage_Acceleration_Final.pdf

iSER OpenStack Support





- Built-in OpenStack components and management
  - No additional software required
  - RDMA is already inbox and ready for OpenStack users

Flash Memory SUMMIT

[Press Release] New Zadara Storage Solution Using iSER RDMA Boosts Application Performance and Reduces Costs

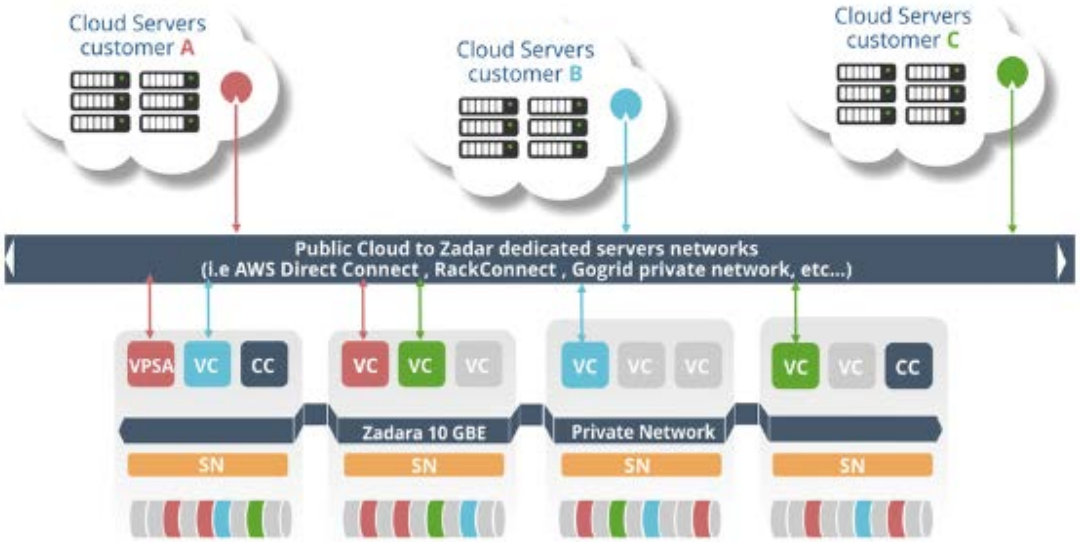
Posted on: August 25th, 2014 | by Press Release | No Comments

Virtual Private Storage Array Adds Mellanox 40 Gigabit Ethernet Technology for iSER RDMA to Deliver Lower Latency and Higher Transactional Rate for Databases

VMworld, San Francisco, CA – August 25, 2014 –Zadara™ Storage today announced a new high performance storage as a service (STaaS) solution offering for private clouds – the Zadara iSER VPSA™ – using a first of its kind Ethernet transport mechanism for exceptional performance at reduced costs. The latest generation of its award-winning Virtual Private Storage Array™ services iSER VPSA takes advantage of iSCSI Extensions for RDMA (iSER) using 40 Gigabit Ethernet to cut latency and boost application performance. Zadara Storage is the first vendor to offer a storage array supporting this next generation Ethernet-based transport. While all block-based applications running on Zadara iSER VPSA will see significant improvements, latency-sensitive databases will see a marked improvement in their transaction per second (TPS) count.

ZADARA STORAGE

CloudFabric Diagram



NetApp

E5500

- Models:
  - E5500/5600: Hybrid HDD/SSD
  - EF550/560: All-flash
- Performance:
  - 530K IOPS
  - 12 GB/s

Saratoga-Speed

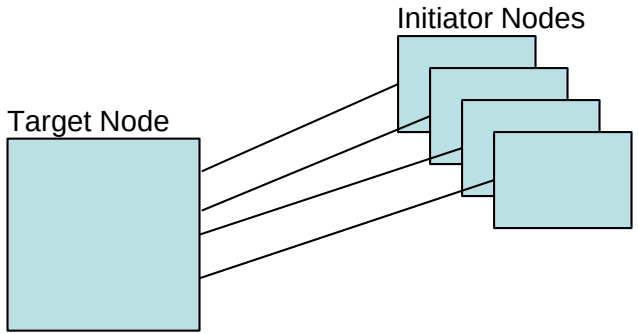
- Model:
  - AltamontXP: All-flash
  - 40Gb/s Ethernet

I/O Type	Read-Write Mix	4KB Blocks		32KB Blocks	
		Bandwidth (KB/s)	IOPS	Bandwidth (KB/s)	IOPS
Sequential	100% Read	8,618,900	2,154,725	10,981,200	343,163
Sequential	100% Write	7,758,300	1,939,575	10.654.000	332,938
Sequential	80% Read, 20% Write	2,866,885	716,721	9,400,997	293,781
Random	100% Read	902,709	225,677	6,057,400	189,294
Random	100% Write	822,935	205,734	5,609,700	175,309
Random	80% Read, 20% Write	868,848	217,212	5,592,925	174,779

SARATOGA SPEED

Samsung

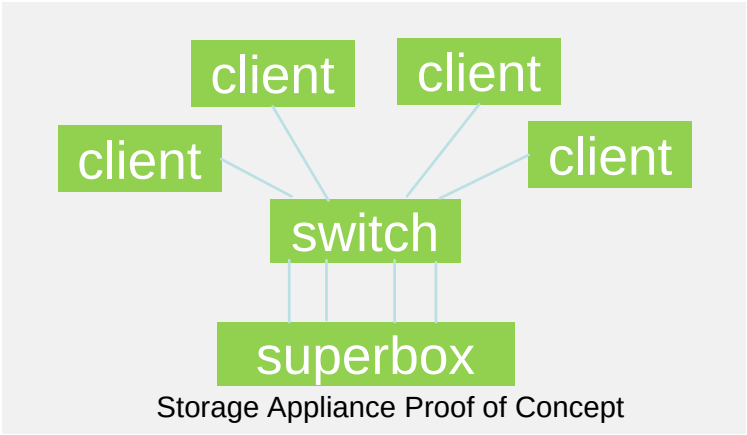
- Target node
  - Dual-socket x86 server
  - 4x40GbE NICs
  - iSER LIO target
  - 20xPM953 NVMe drives
- Initiators
  - Dual-socket x86 server
  - 1x40GbE NIC
- Performance
  - 2.1M – 4K Random Read
  - 17.2GB/s – 128K Seq Read



Micron

+ 15GB/s Max

+ 2.7MIOPS @ 512B
+ 1.8MIOPS @ 4K
+ 1.2MIOPS @ 8K
+ 850KIOPS @ 16K
+ 480KIOPS @ 32K



Storage Appliance Proof of Concept

+ "Datasheet" 4K
  + 122us read, 43us write
+ 1MIOPS @ 8K 50/50
  + 471us average

+ 2 dual port 40GbE NICs

Conclusions

- iSER gives users the full performance benefits of flash based solutions across a Ethernet or InfiniBand network
- RDMA technology like RoCE enables iSER performance by bypassing the TCP/IP network stack
- A growing number of storage solutions providers support iSER in their Flash based products

Thankyou!

Rob Davis

robd@mellanox.com