12th ANNUAL WORKSHOP 2016

ISER AS ACCELERATOR FOR SOFTWARE DEFINED STORAGE

Subhojit Roy, Tej Parkash and Pravin Shinde, Storage Engineering

IBM

[APRIL 6th, 2016]

AGENDA

- **Network storage virtualization**
- **Current state of Fiber Channel**
- **iSCSI seeing significant adoption**
- **Emergence of Ethernet Storage**
- **iSER: Exploit RDMA for iSCSI**
- **iSER vs Fibre Channel**
- **iSER vs other Ethernet Storage protocols**
- **Ecosystem around iSER**
- **SDS with RDMA and iSER**
- **Key considerations and challenges**

WHAT DO WE DO?

- **Network Storage Virtualization – IBM Spectrum Virtualize**
  - *SAN Volume Controller* (*SVC*) and *Storwize* platforms
  - Block Storage Target for servers
  - Block Storage Initiator for storage
  - SCSI
  - Attach to diverse hosts: Linux, Windows, VMWare etc.
  - Virtualize storage from vendors: IBM, Hitachi, EMC etc.
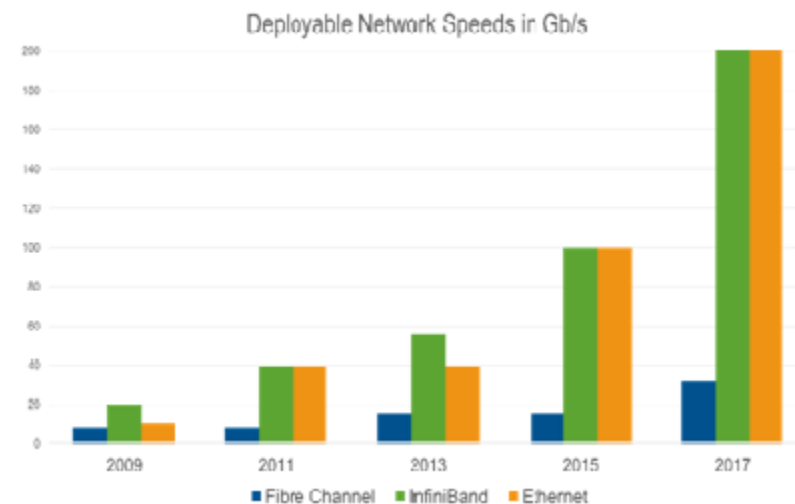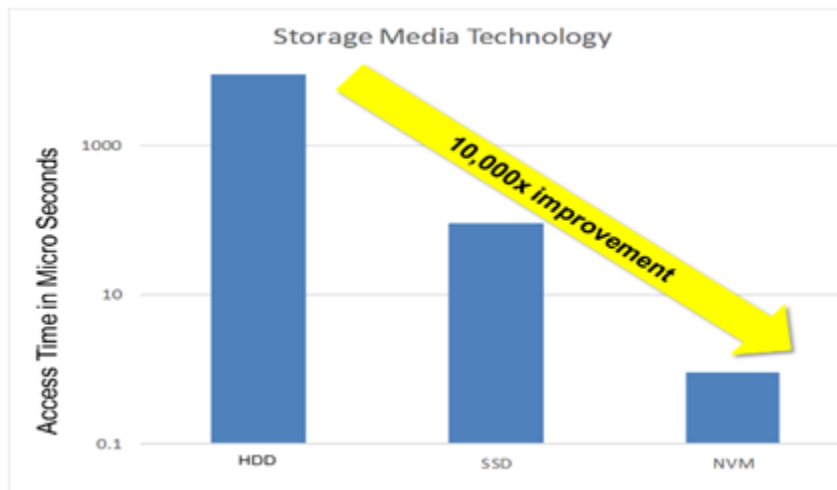- **Workloads - Enterprise, Cloud…**
  - Traditionally connected over Fiber Channel (structured data)
  - iSCSI (Ethernet) gaining momentum (cloud)

3
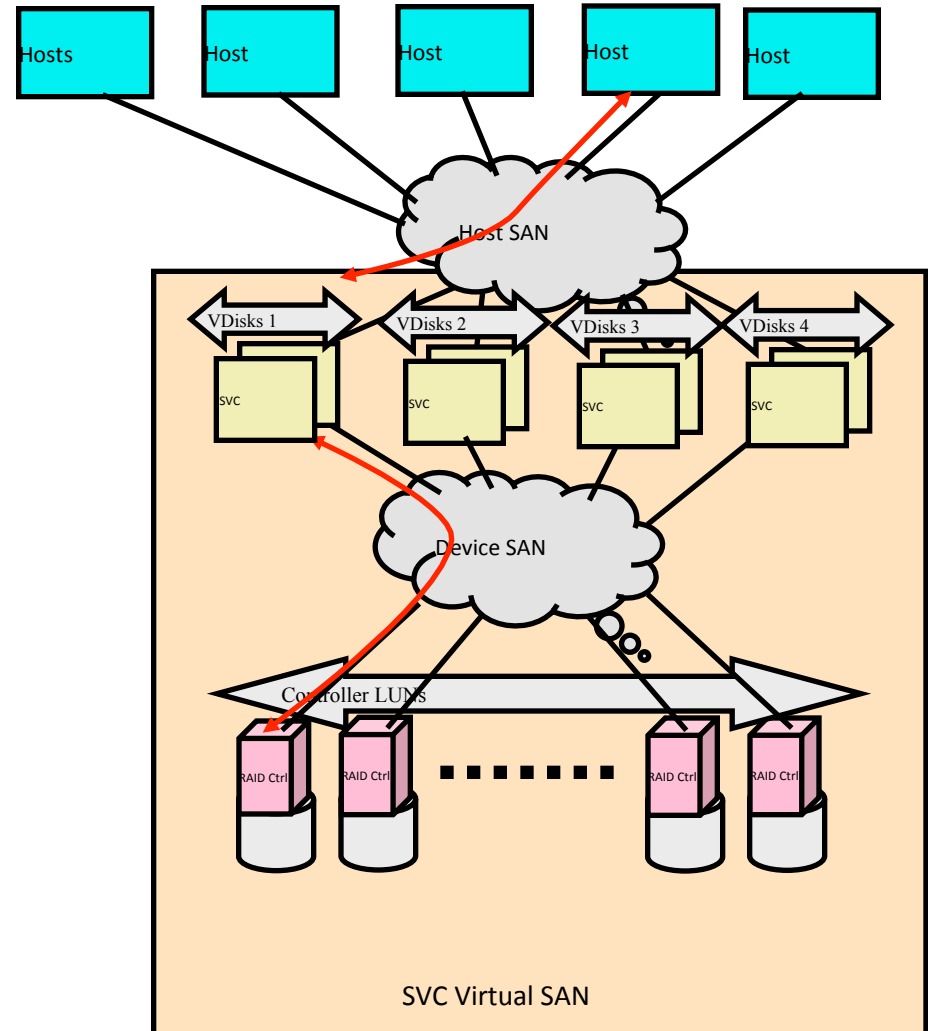
WHAT'S HAPPENING TO FIBER CHANNEL?

- **Fibre Channel block storage access is fine but…..**
  - Flash Storage is driving the need for next generation network speeds to fully utilize its capabilities
  - Clients prefer Ethernet speeds and converged infrastructure for Cloud economy
  - Fiber Channel is behind in the speed war - 32Gb is expected in 2017 while 40G Ethernet already has $200M revenue today
  - Gartner predicts declining FC port counts at 2% to 5% annually and flattening sales

Storage Media Technology

10,000x improvement

Deployable Network Speeds in Gb/s

■ Fibre Channel ■ InfiniBand ■ Ethernet

ISCSI ADOPTION IS SIGNIFICANT

- iSCSI has become the fastest growing interconnect method for network storage systems and growing at 6.4% CAGR between 2013 to 2018 compare to fibre channel which is increasing only by 2.7% CAGR

- Key to iSCSI growth are
  - Lower cost for storage network infrastructure
  - DCBx introduces enterprise capabilities
  - Server visualization, 10 Gigabit Ethernet proliferation, cloud…
  - Linux, VMWare and Microsoft support iSCSI

Installation ($ billion)	2011	2012	2013	2014	2015	2016	2017	2018	CAGR % (13-18)
Fibre Channel	11.80	12.50	12.60	12.90	13.30	13.70	14.00	14.40	2.7
iSCSI	3.30	3.50	3.40	3.70	3.90	4.20	4.40	4.70	6.4

OpenFabrics Alliance Workshop 2016

EMERGENCE OF ETHERNET STORAGE

- **Performance**
  - Proliferation of 10Gb iSCSI
  - Rapid transition to 40Gb! In 2016 end 40G total revenue will be 1/4th of 10Gb
  - DCBx enabled Ethernet fabric enables QOS & reliable data transfer necessary for storage

- **25G Standards**
  - Promises minor increment in cost to move from 10Gb to 25Gb
  - Lower power consumption, network consolidation, scales to 50/100Gb easily
  - Hyperscale data center architectures like Google and Facebook are lured by the promise of higher bandwidths and lower costs

- **Server and Storage network convergence**
  - Ethernet supports converged infrastructure for cloud vendors that use block, file, object and distributed scale out storage
  - Wikibon predicts server SAN (compute and storage over converged network) will grow 44.2% CAGR

EMERGENCE OF ETHERNET STORAGE CONTD.

- **Multitenancy support**
  - QoS enabled by DCBx networking standards
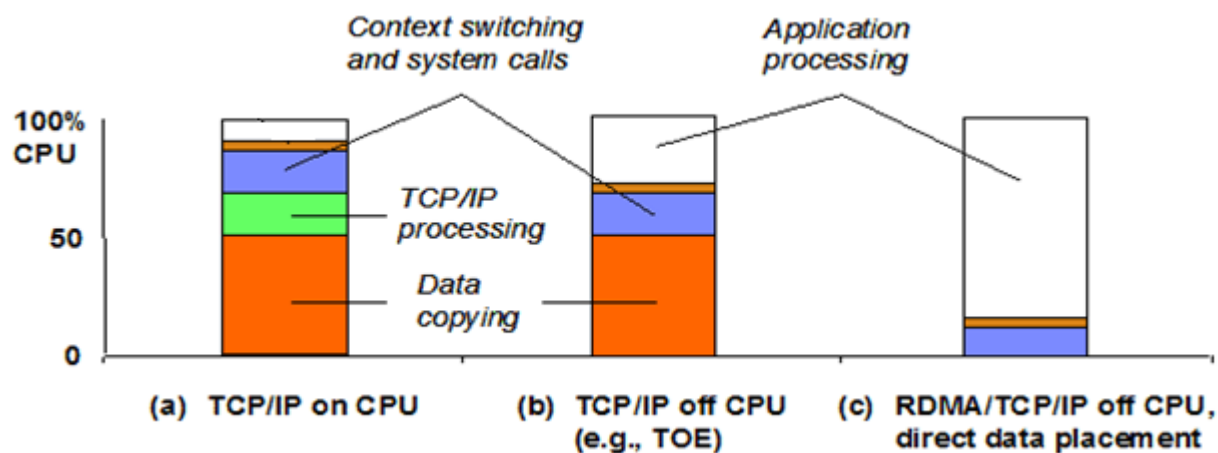 - IPSec provides for strong authentication & data confidentiality

- **Ecosystem evolution**
  - Cloud adoption drives Ethernet ecosystem adoption due to economic benefits
  - LAN on Motherboard (LOM) makes Ethernet adoption simpler & less expensive
  - Major switch vendors adopting higher bandwidths DCBx standards and quickly

WHY RDMA OVER ETHERNET

- **Application Performance**
  - Low CPU utilization leaves space for more applications per server
  - Allows bandwidth utilization to scale higher to i.e. 25/40/50/100 Gb speeds


Context switching and system calls

Application processing

100% CPU

TCP/IP processing

50

Data copying

0

(a) TCP/IP on CPU   (b) TCP/IP off CPU (e.g., TOE)   (c) RDMA/TCP/IP off CPU, direct data placement

- **RDMA drives down latencies**
  - Fully Zero copy (Reads and Writes)
  - Kernel bypass
  - Very low latencies

- **RDMA is mature technology**

ISER: CONFLUENCE OF ISCSI & RDMA

- iSER is iSCSI with a RDMA data path

- Requires no changes to SAM-2/3 and uses iSCSI RFC with minimal changes to realize iSER

- Network protocol independence: iWARP, RoCE, Infiniband
  - Common OFED stack

- Leverages existing knowledge of iSCSI administration & ecosystem on servers and storage

ISCSI and ISER Architecture

© 2005 Copyright Voltaire Inc.

DoSTOR.COM
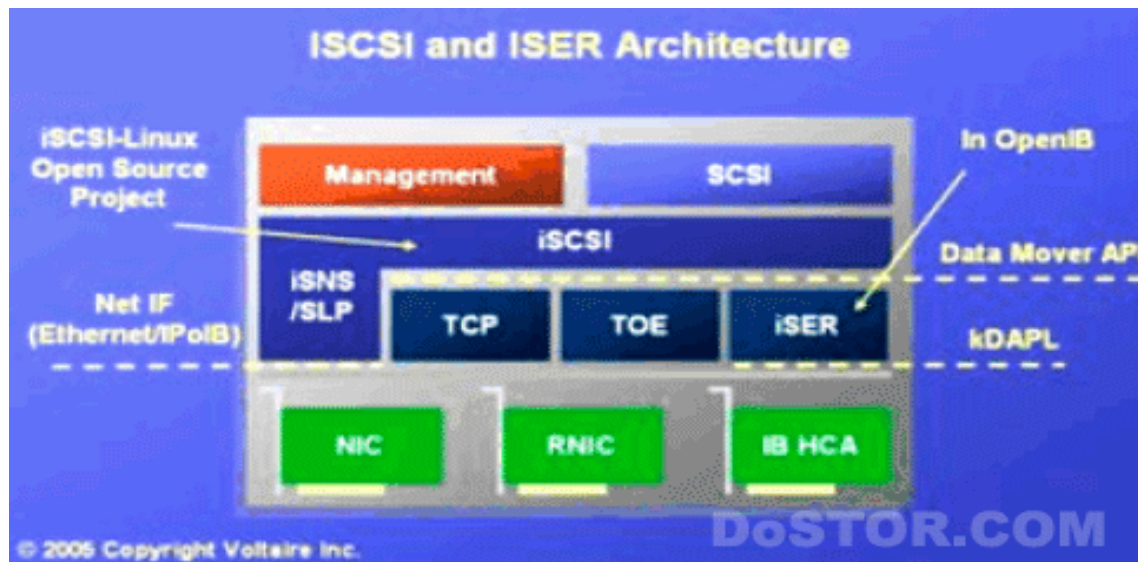
ISER VS FIBRE CHANNEL

| Feature/Protocol	iSER	Fibre Channel
Read Latency	15-25us	25-35us
Bandwidths	10/25/40/50/100 Gb	8/16/32(?) Gb
CPU Utilization	Low	Low
Security	Authentication, Confidentiality, Integrity	Integrity
Ownership cost	Low	Medium - High
Market	Growing rapidly and evolving	Mature and stable
Workloads	Cloud, Analytics, Enterprise	Enterprise

iSER: Fiber Channel benefits minus the additional costs

OpenFabrics Alliance Workshop 2016

ISER VS OTHER ETHERNET STORAGE PROTCOLS

	iSER	SRP	FCoE
Management	iSCSI based	NA	FC Based
RDMA	Yes	Yes	No
Physical Networks	Ethernet and Infiniband	Infiniband	Ethernet Only
OS	Linux/VMware/BSD	Linux	Linux/VMware/BSD
Security	Authentication, Confidentiality (IPSec), Integrity	Unknown (??)	Integrity only
Scalability	High (runs on DCBx enabled switches)	Unknown (??)	Low (until BB6 takes hold)
Routability	Yes	No	No
Ecosystem	Rapidly evolving	Not growing	Low movement on BB6

iSER is ahead of other Ethernet based technology

11

EVER EXPANDING ECOSYSTEM FOR ISER
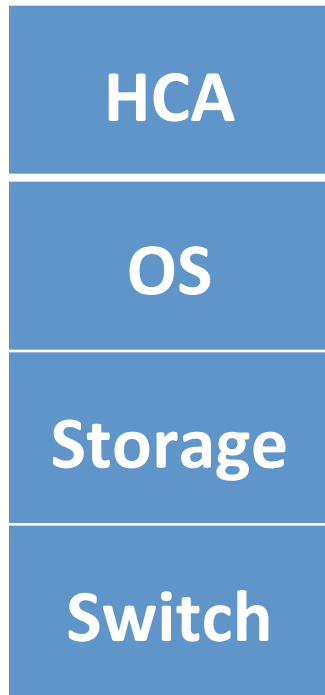
iSER ecosystem growing with more cloud and enterprise adoption

ISER FOR SOFTWARE DEFINED STORAGE

	iSER	FC
Run on commodity hardware	✔	✘
Runs on converged networking technology	✔	✘
Scalable	✔	✔
High Performance	✔	✔
Driven by state of the art Storage Technology		✔
Flash	✔	
Big Data	✔	✘
Cloud	✔	✘

Techology independence & vendor independence makes iSER compelling for SDS

KEY CONSIDERATIONS & CHALLENGES

- **User space architecture**
  - To drive very low latencies & CPU utilization
  - Storage Virtualization functions are in user space

- **Speed of Memory Registration**
  - Data transfers to/from Scattered physical memory
  - Need to use Fast Memory Registration technique (FMR, FRWR, UMR)
  - Fast memory registration available mainly through kernel ib verbs

- **Vendor & Technology independence for Software Defined**
  - Must work with iWARP, RoCE (v1 and v2) and Infiniband
  - Vendor independence: **Mellanox**, **Chelsio**, **Emulex** & **Qlogic**
  - Only Mellanox supports UMR to simplify registration of discontiguous physical pages

12th ANNUAL WORKSHOP 2016

THANK YOU

Subhojit Roy, subhojit.roy@in.ibm.com