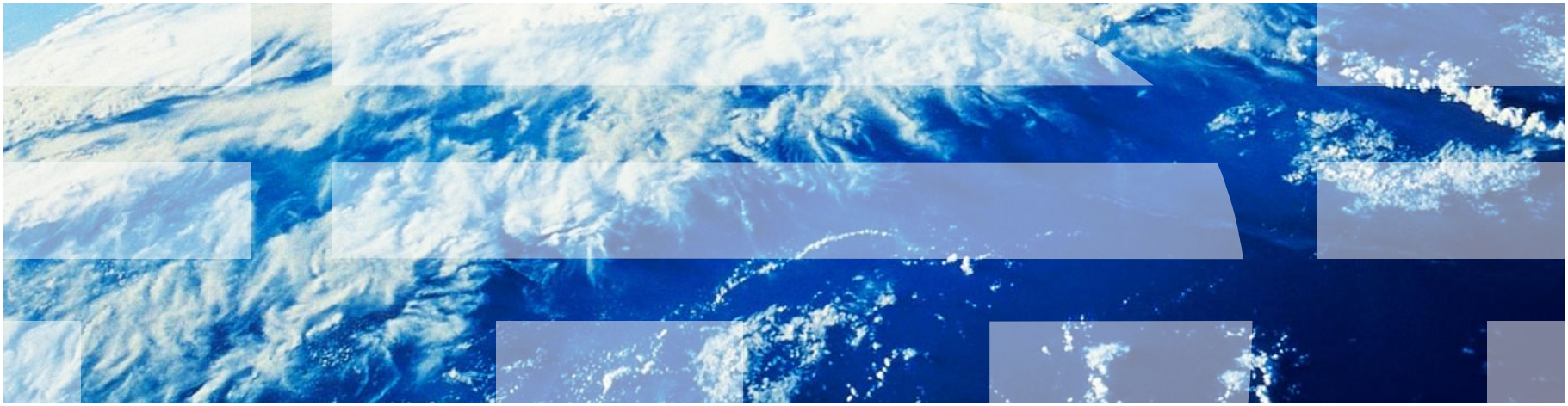
# iSER as Accelerator for Software Defined Storage

**Subhojit Roy and Tej Parkash Storage Engineering**



# Agenda

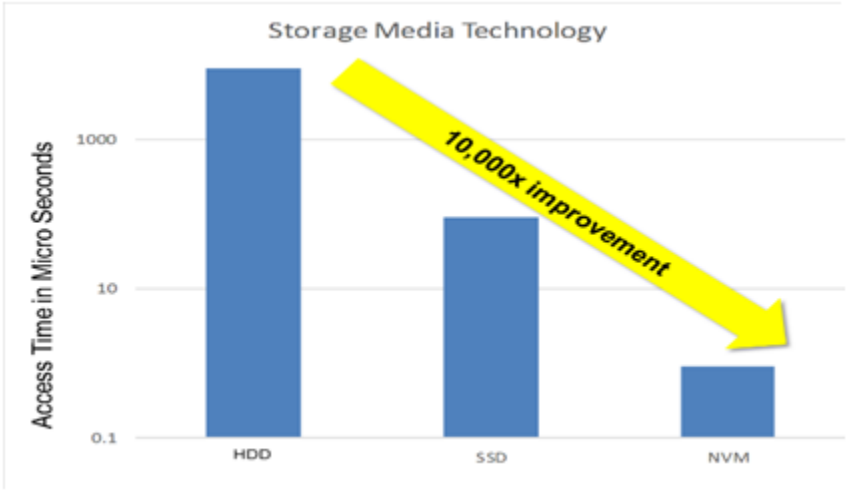
- Key Requirement for Software Defined Storage (SDS)
- Current state of Fiber Channel
- RDMA over Ethernet
- Emergence of iSCSI  and iSER (iSCSI Extension for RDMA)
- iSER vs others protocols
- IBM Spectrum Virtualize
- Considerations and challenges in iSER adoption

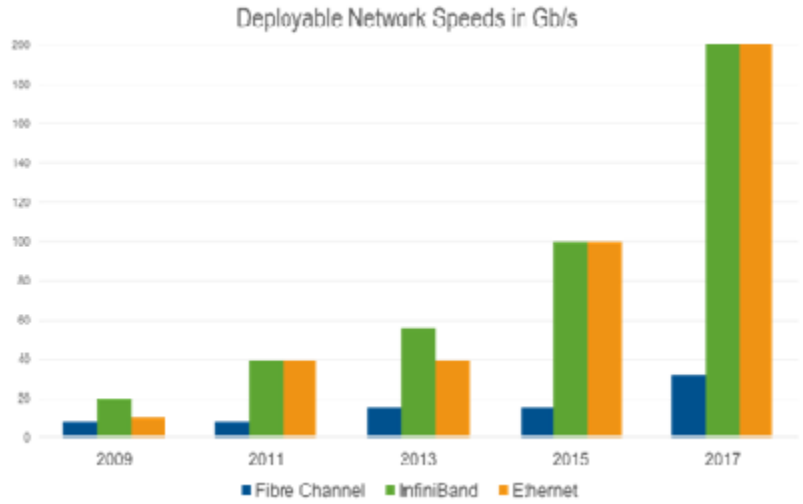
# SDS and its key requirements

- **SDS**: **Virtualized storage with a service management interface, considering software and hardware independence. Criteria are:**
  - **Standard interfaces:** APIs for management, provisioning and maintenance of storage devices and services
  - **Virtualized data path:** Block, File and/or Object interfaces that support applications written to these interfaces
  - **Commodity hardware**: Software should run on off-the-shelves hardware
  - **Scalability**: Ability to scale storage infrastructure without disruption to specified availability or performance
  - **Support for new age workload**
  - **Converged networking:** Same network could carry both compute and storage data

# 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



# 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
  - Cloud data centers pushing10 Gigabit Ethernet proliferation.
  - 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 |

# Emergence of Ethernet Storage

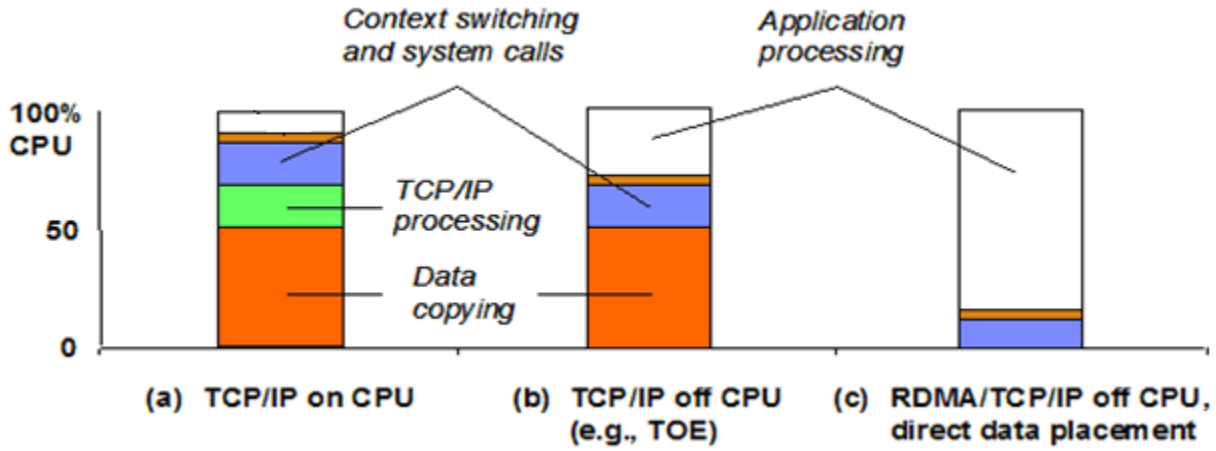
- **Revenue Growth**
  - Proliferation of 10Gb iSCSI
  - Rapid transition to 40Gb! In 2016 end 40G total revenue will be 1/4$^{th}$ of 10Gb
  - DCBx enabled Ethernet fabric enables QOS & reliable data transfer necessary for storage
- **25G Standards**
  - Promises minor increment in cost to move from 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

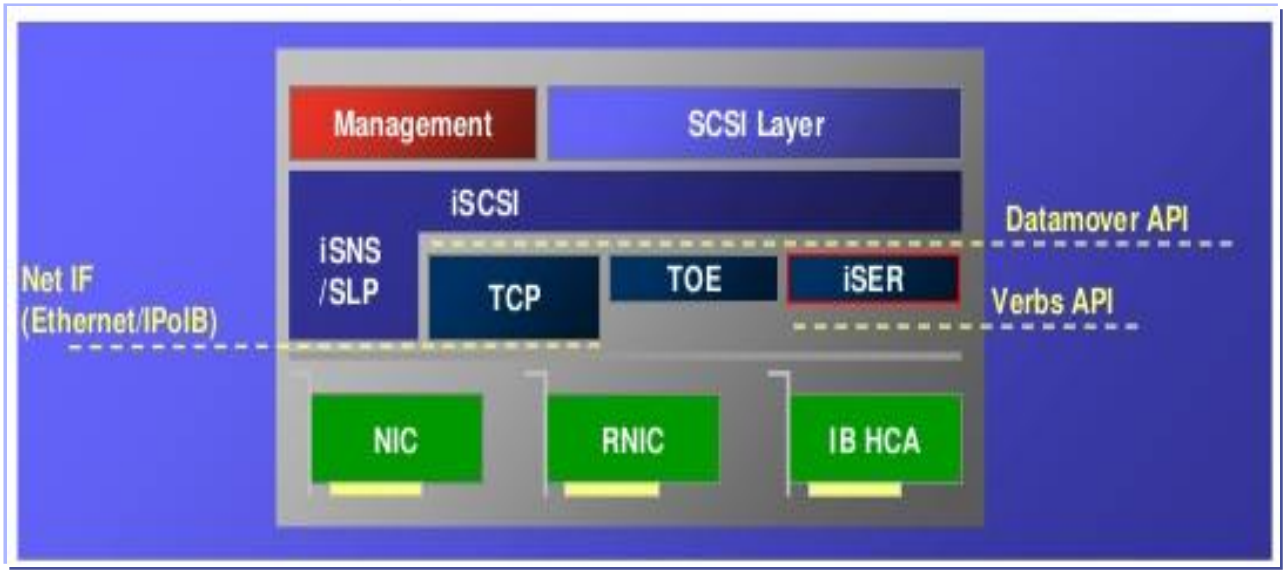
- **Application Performance**
  - Low CPU utilization leaves space for more applications per server
  - Allows bandwidth utilization to scale higher to i.e. 25/40/50/100 Gb speeds




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

# ISER: Confluence of iSCSI & RDMA modify

- 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



# 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***

# 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 | Slow movement on BB6 |

***iSER is ahead of other Ethernet based technology***

# Ever EXPANDING ecosystem for iSER

| | |
|---|---|
| HCA | QLOGIC, EMULEX, Mellanox TECHNOLOGIES, intel, Chelsio Communications Accelerate |
| OS | Linux, vmware, ORACLE LINUX, FreeBSD |
| Storage | IBM, ZADARA STORAGE, hp invent, NetApp |
| Switch | BROCADE, Mellanox TECHNOLOGIES, ARISTA, CISCO |

***iSER ecosystem growing with more cloud and enterprise adoption***

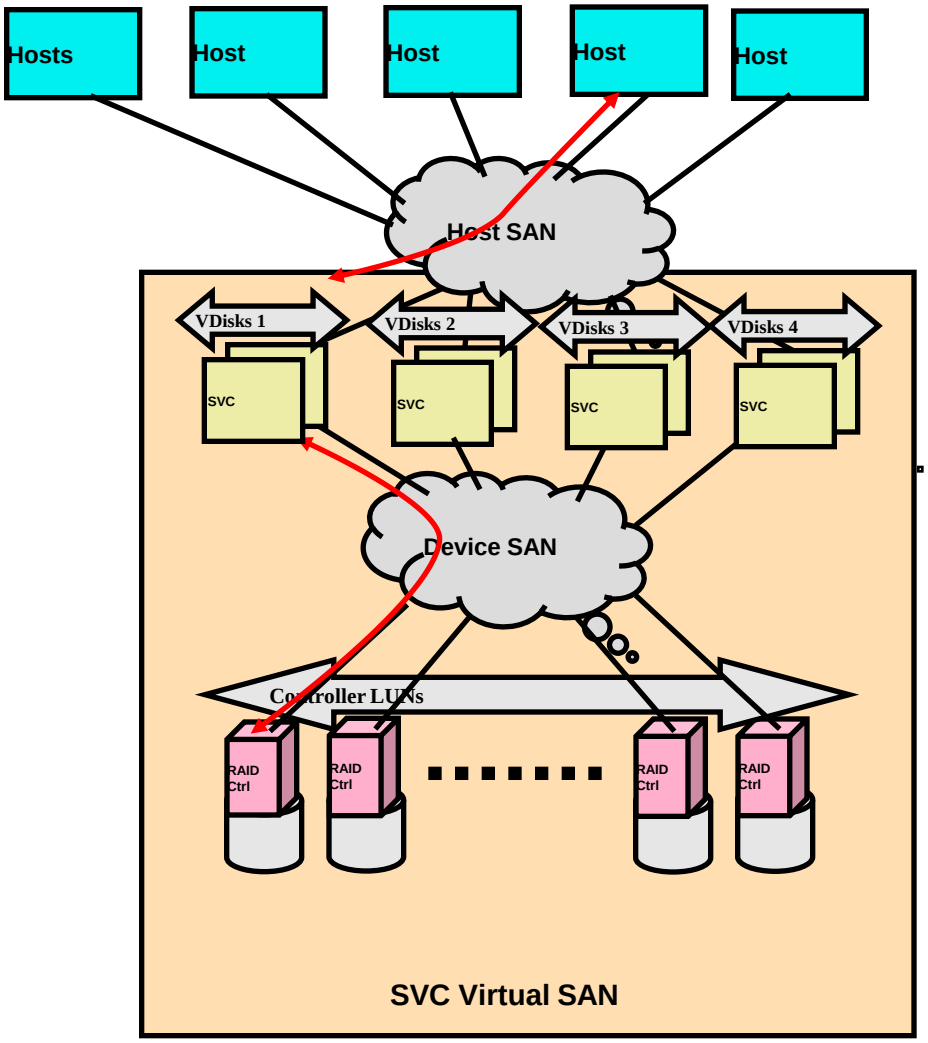
# iSER for Software Defined Storage

| | iSER | FC |
|---|---|---|
| *Run on commodity hardware* | ✓ | ✗ |
| *Runs on converged networking technology* | ✓ | ✗ |
| *Scalable* | ✓ | ✓ |
| *High Performance* | ✓ | ✓ |
| *Driven by new age workloads*<br>*Flash*<br>*Cloud*<br>*Big Data* | ✓<br>✓<br>✓ | ✓<br>✗<br>✗ |

***iSER qualifies for more SDS criteria***

# 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)



# CONSIDERATIONS

- We are both Initiator and target
- Storage Virtualization stack is in user space
- Fast memory registration available mainly through kernel ib verbs
- Match or exceeds Fiber Channel (FC) latencies & CPU utilization
- Use vendor independent Fast memory registration technique (OFED)
- Must work with iWARP, RoCE (v1 and v2) and Infiniband

# CHALLENGES

- Reduce latency of Memory Registration for initiator
- Data transfers from Scattered physical memory

# THANK YOU