iSER as Accelerator for Software Defined Storage

Subhojit Roy and Tej Parkash Storage Engineering

IBM
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: 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 40Gb 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 pushing 10 Gigabit Ethernet proliferation.
  - Linux, VMware and Microsoft support iSCSI

---

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Fibre Channel</td>
<td>11.80</td>
<td>12.50</td>
<td>12.60</td>
<td>12.90</td>
<td>13.30</td>
<td>13.70</td>
<td>14.00</td>
<td>14.40</td>
<td>2.7</td>
</tr>
<tr>
<td>iSCSI</td>
<td>3.30</td>
<td>3.50</td>
<td>3.40</td>
<td>3.70</td>
<td>3.90</td>
<td>4.20</td>
<td>4.40</td>
<td>4.70</td>
<td>6.4</td>
</tr>
</tbody>
</table>
Emergence of Ethernet Storage

**Revenue Growth**
- Proliferation of 10Gb iSCSI
- Rapid transition to 40Gb! In 2016 end 40G total revenue will be $\frac{1}{4^{th}}$ 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

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

**iSER: Fiber Channel benefits minus the additional costs**
# iSER vs Other Ethernet Storage Protocols

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

* *iSER is ahead of other Ethernet based technology*
Ever EXPANDING ecosystem for iSER

iSER ecosystem growing with more cloud and enterprise adoption
### iSER for Software Defined Storage

<table>
<thead>
<tr>
<th>Feature</th>
<th>iSER</th>
<th>FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Run on commodity hardware</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Runs on converged networking technology</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>Scalable</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>High Performance</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Driven by new age workloads</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Flash</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Cloud</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Big Data</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>

**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