SDC
STORAGE DEVELOPER CONFERENCE
SNIA ■ SANTA CLARA, 2015

Next Generation Low Latency SAN's

Rupin Mohan
Chief Technologist,
Storage Networking
Hewlett-Packard

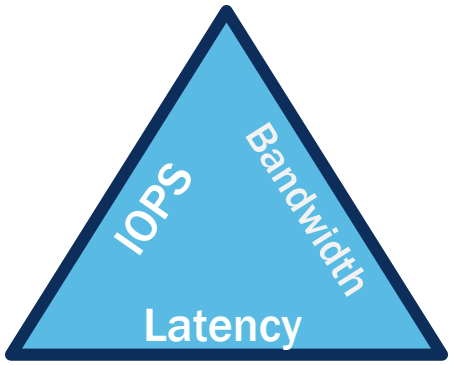
Sep 20th, 2015
Santa Clara, CA

Agenda

- Introduction
- How do we measure performance?
- Storage Protocol Comparison
- Current state of the union
- What are the new options?
  - A look into the future
- Summary / Key takeaways

How do we measure performance?

- **IOPS** – I/O's per second – a measure of the total I/O operations (reads and writes) issued by applications
- **Bandwidth** – a measure of the data transfer rate, or I/O throughput, MB/s
- **Latency** – time taken to complete an I/O request, also known as response time. Usually measured in milliseconds (1/1000 of a second). Going forward, in microseconds (1/1000 of a millisecond)

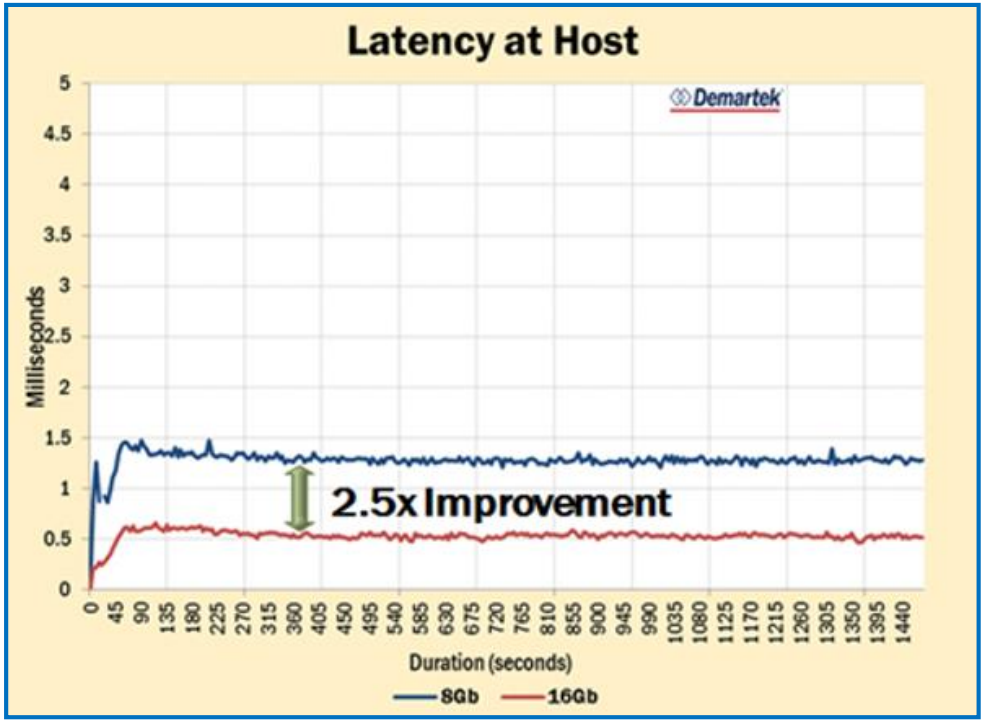


The application/user experience

Storage Protocol Comparison

Attribute	IB	FC	FCoE (DCB Ethernet)	iSCSI
Bandwidth (Gbps)	56	8/16/32	10/25/40/100	10/25/40/100
Adapter Latency	25 us	50 us	200 us	Wide range(w offload / w/o offload)
Switch Latency per hop	100-200ns	700 ns	200ns to1 us	200ns to1 us
Adapter	HCA	HBA	CNA	NIC
Convergence - Single L2 network	No	No	Yes	Yes
Separate Network for management	Yes	Yes	No	No
App Changes Req	Yes	No	No	No
Routable	No	No	No	Yes
Kernel Bypass	Yes	No	No	No
Max Frame size	2K	2K	9K	9K
Maturity	High	High	High	High

Current state of the union



- Using All Flash 3PAR 7450
- End to end **Gen 5** (16Gb) Infrastructure
- Latency at host (end-to-end) around ~.5 ms
- 75% reduction in latency compared to 8Gb FC

What are the new options?

- Gen 6 Fibre Channel
- Storage protocols on RDMA
  - iSER
  - SMB Direct
- Transport options for RDMA
  - RoCE
  - iWARP
- NVMe over Fabrics
  - NVMe over RDMA
  - NVMe over FC

Gen 6 Fibre Channel

- 32 Gb/s single lane
- 128 Gb/s multi lane
- How is it lower latency?
  - Lower latency through higher clock rate
  - Possible smaller ASIC geometries

RDMA – Remote Direct Memory Access

- Introduction
  - Accelerated IO delivery model, allowing application software to bypass most layers of software and communicate directly with the hardware
  - Requires new programming model: “verbs” rather than “sockets”
  - Protocol options: Block (iSER) / File (SMB Direct)
  - Transport options: RoCE, iWARP, Infiniband
- RDMA Benefits
  - Low latency, also important is latency jitter
  - High Throughput
  - Zero copy capability, OS / Stack bypass
  - Avoid CPU context switching, interrupt coalescing
- Bulk of the work is done by the Target
  - Read operation, translates to a write by Target
  - Write operation, translates to a read by Target

iSER and SMB Direct over Ethernet

- iSER
  - iSCSI extensions over RDMA
  - Mature protocol
  - Limited OS stacks in both functionality and support. Linux (yes), Windows (yes), VMware (?)
- SMB Direct
  - NFS using Direct I/O
  - Mature protocol
  - Limited OS stack support. Linux (yes), Windows (yes), VMware (?)

RoCE versus iWARP

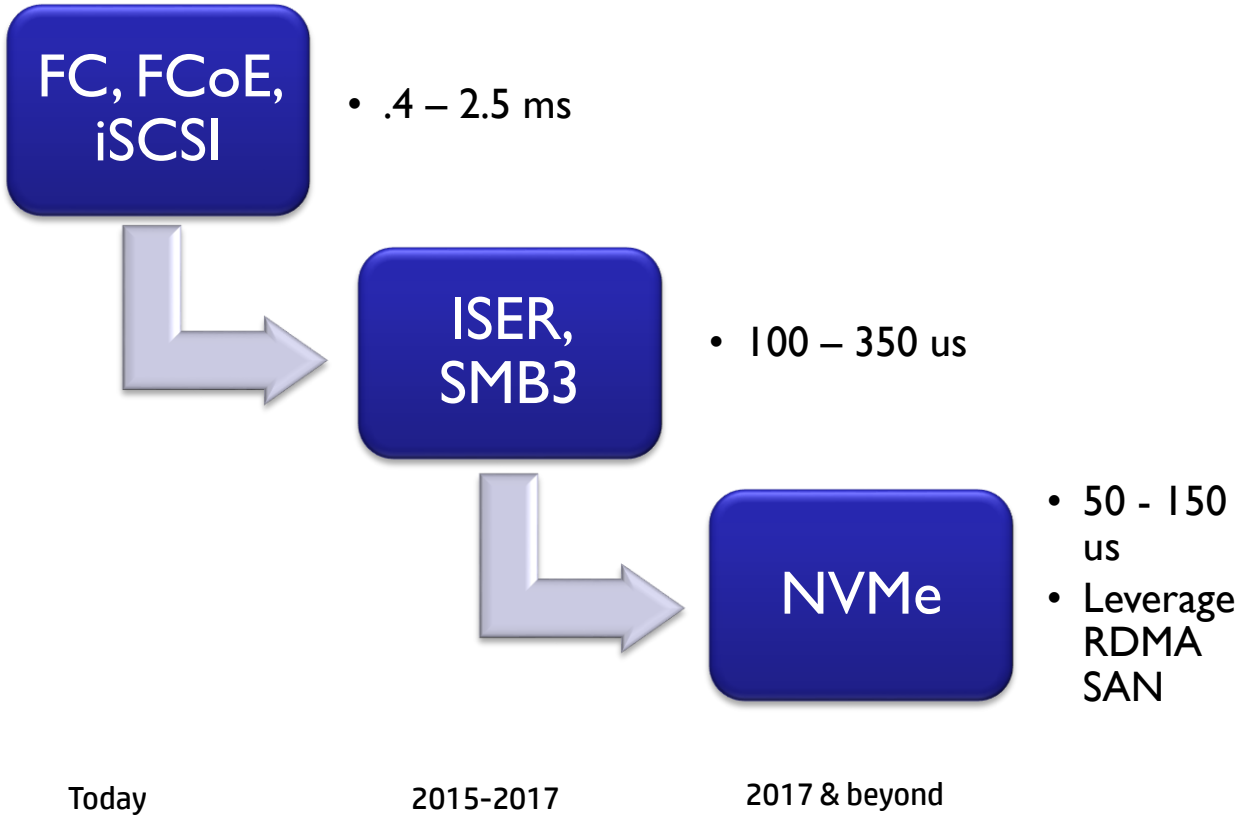
RoCE

- Needs DCB Switching infrastructure
- Routability comes with v2
- End to end congestion management is a big issue.
- Higher cost solution (DCB)
- DCB configuration on switches is cumbersome
- Lowest Latency

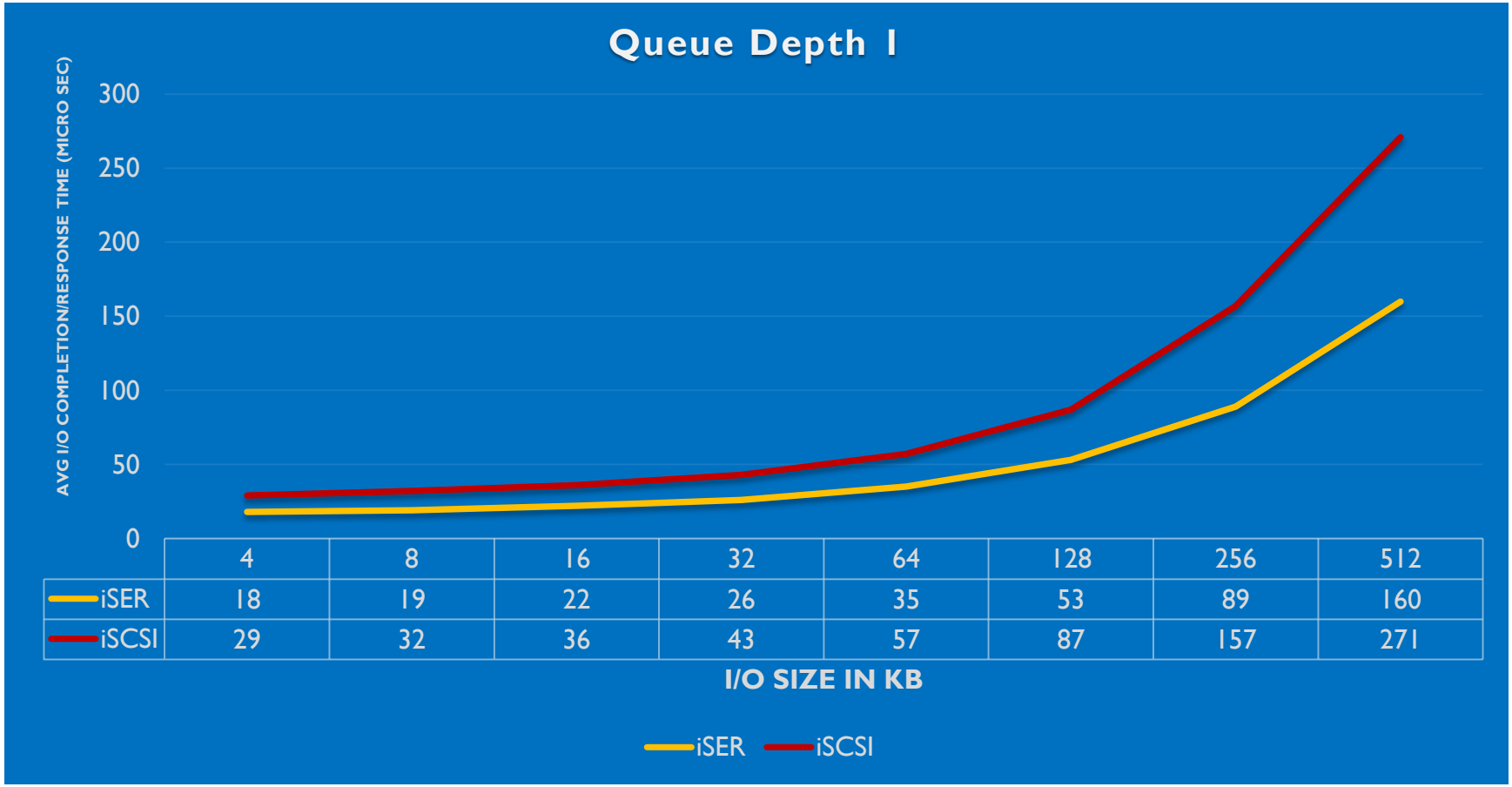
iWARP

- Does not require DCB Switches
- Routable as it runs over TCP/IP
- TCP solves the congestion management issue but adds latency
- Lower cost solution
- Switch config simpler
- Low Latency

Latency Step Function



“READ” Response time with increasing I/O size (Lower is better)



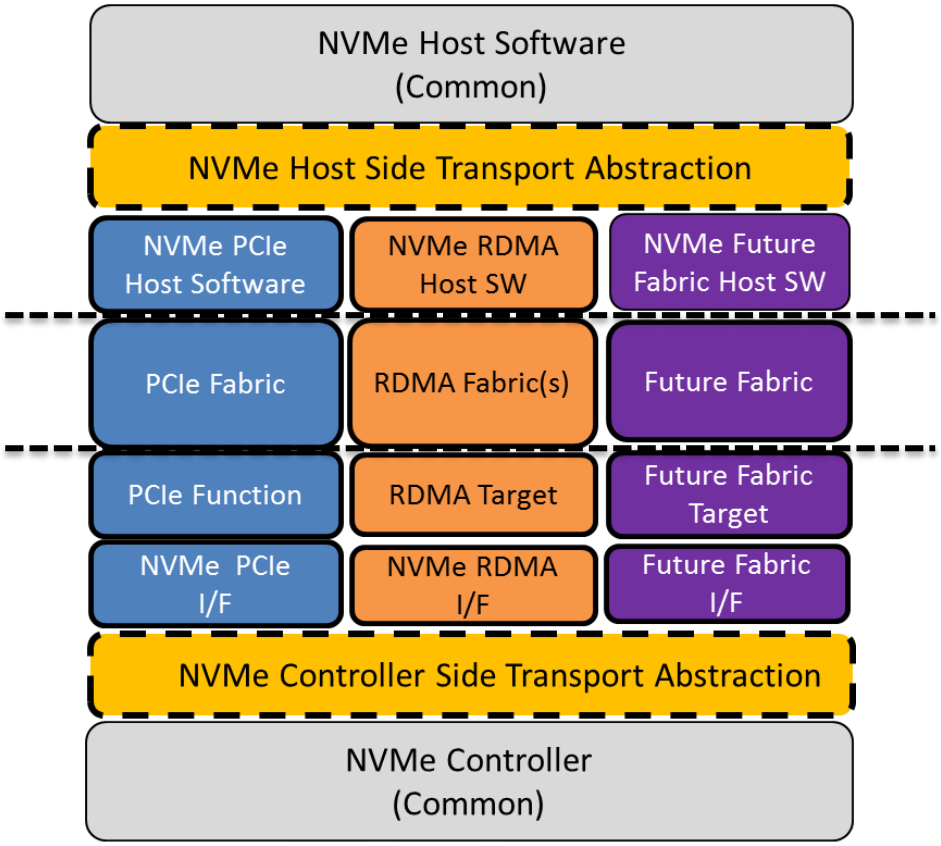
	4	8	16	32	64	128	256	512
iSER	18	19	22	26	35	53	89	160
iSCSI	29	32	36	43	57	87	157	271

NVME OVER FABRICS

NVMe Transports

- Two new fabric transport projects for NVMe
  1. NVMe over Fabrics – Being defined in a subgroup of the NVM Express group
     - NVMe over RDMA
  2. NVMe over FC (FC-NVMe) – New T11 project to define an NVMe over Fibre Channel Protocol

NVMe over Fabrics



- Being defined by a Technical sub-group of the NVM Express group
  - Defined as upper level protocol on top of OFED RDMA verbs
  - Fabric agnostic
    - Supports RDMA fabrics – Ethernet (iWARP, RoCE), Infiniband
    - Support for other fabrics – FC
  - Complies to the NVMe programming model

Key Takeaways

- Fibre Channel is low latency but there is competition
- iSER and SMB Direct are in product development right now
- Gen 6 Fibre Channel is in product development, FC is also working on lowering latency
- NVMe over Fabrics is under standards development now
- Data center networks are key to application latency

Thank You

- Rupin Mohan
  - Rupin.mohan@hp.com
  - +1-774-245-2947