Optimizing NVMe over Fabrics Performance with Different Ethernet Transports: Host Factors

Live Webcast
September 16, 2020
10:00 am PT
Today’s Presenters

Tom Friend
Moderator
Illuminosi

Fred Zhang
Presenter
Intel

Eden Kim
Presenter
Calypso Systems

David Woolf
Presenter
University of New Hampshire
SNIA Legal Notice

- The material contained in this presentation is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification.
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA.
- Neither the authors nor the presenters are attorneys and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel.
  - If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the authors' personal opinion and current understanding of the relevant issues involved.
  - The authors, the presenters, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.

NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.
SNIA-at-a-Glance

185 industry leading organizations

2,000 active contributing members

50,000 IT end users & storage pros worldwide

Learn more: snia.org/technical  @SNIA
Technologies We Cover

- Ethernet
- iSCSI
- NVMe-oF
- InfiniBand
- Fibre Channel, FCoE
- Hyperconverged (HCI)
- Storage protocols (block, file, object)
- Virtualized storage
- Software-defined storage
Introduction to NVMe over Fabrics

David Woolf, University of New Hampshire
Agenda

- Recap of NVMe over Fabrics
- Factors Impacting Different Ethernet Transport Performance for NVMe over Fabrics
- Data Comparison of NVMe over Fabrics test with iWARP, RoCEv2 and TCP
- Conclusion
NVMe-oF What is it?

- NVMe is a storage protocol optimized for flash memory.
  - The transport used for SSDs is PCIe
- NVMe-oF maps that protocol onto fabric transports.
  - iWARP
  - RoCEv2
  - TCP
  - Fibre Channel
  - Infiniband

We’ll focus our presentation here.
Know your NVMe-oF Transports: What’s the difference?

- **iWARP: RDMA over TCP**
  - Low latency, Scales well on large datacenter networks
  - Requires iWARP capable adapters
  - Increase performance with TCP Offload

- **RoCEv2: RDMA over Converged Ethernet**
  - Low latency, best suited for Rack scale
  - Requires RoCE capable NICs/Switches

- **NVMe/TCP: NVMe mapped directly on TCP**
  - Low latency, Scales well on large datacenter networks
  - Easily supported on simple NICs/Switches
  - Increase performance with TCP Offload
NVMe-oF: How Mature is It?

- **Specification Maturity**
  - NVMe-oF v1.1 specification released in October 2019
  - NVMe-oF v1.0 specification released in June 2016

- **Driver support**
  - Linux support for Initiator and Target drivers
  - Starwind produced NVMe-oF Windows Initiator

- **Testing**
  - Dozens of products on UNH-IOL Integrators List for NVMe-oF and continuing plugfests

- **Many Ethernet based NVMe-oF product launches in 2020**
Factors Impacting Different Ethernet Transport Performance

Fred Zhang, Intel
Scope of the Discussion of Factors Impacting NVMe-oF Performance: Host Factors

- Many factors impact NVMe over Fabric performance
  - On Host: CPU, NVMe drive
  - Switch: settings for congestion management
  - Network: over-subscribed, different fan-in ratio

- This presentation will focus on Host factors
  - Offload vs. non-offload
  - NVMe drive attributes
  - MTU (Maximum Transmission Unit or Frame)
  - Others, e.g. Workload, Thread Count, Queue Depth, pre-conditioning, test flow

- Focusing on 100GbE throughput

- Additional factors will be discussed in future topic:
  - Switch setup
  - Number of nodes in the storage network
  - Network topology
  - Congestion
Offload vs. Non-Offload

- RDMA is a host bypass and offload technology that need less CPU utilization

- Traditional TCP relies on protocol stack and consumes CPU cycles, especially for high speed Ethernet, e.g. 100GbE
  - New technologies are coming up to optimize on top of standard TCP stack to archive high performance with less CPU utilization, e.g. Application Device Queues

- Other software like SPDK work in user space, using polling, to get the high performance, with dedicated CPU cores

- Offloaded TCP engine will save CPU cycles and get to the desired performance with less CPU utilization
NVMe Drives with different Read/Write IOPs

- NVMe over Fabric performance is very much reliant on NVMe drives performance for directly attached storage target, especially when network speed moved up to 100GbE.

- For storage target with head nodes connected to storage clients, or storage array on network, there are additional layers of complexity: RAID, QoS, etc.

- NVMe drive characteristics:

<table>
<thead>
<tr>
<th>Manufacturer Specs</th>
<th>Random Reads</th>
<th>Random Writes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSD 1</td>
<td>550K IOPS</td>
<td>550K IOPS</td>
</tr>
<tr>
<td>SSD 2</td>
<td>550K IOPS</td>
<td>250K IOPS</td>
</tr>
<tr>
<td>SSD 3</td>
<td>1M IOPS</td>
<td>130K IOPS</td>
</tr>
<tr>
<td>SSD 4</td>
<td>285K IOPS</td>
<td>41K IOPS</td>
</tr>
</tbody>
</table>

- Real world workload IO patterns are very different:
  - Heavy read/light write
  - Balanced read/write
  - Light read/heavy write
  - Small IO size vs. large IO size
MTU – ~1500B vs. ~9000B

- MTU (Maximum Transmission Unit), in the context of Internet Protocol, is the maximum size of IP packet allowed without fragmentation
  - Default MTU is 1500B, Ethernet frame adds 18B, if not tagged, to make it 1518B maximum Ethernet frame
  - If Jumbo Frame is supported, maximum Ethernet frame can be 9000B

- Many modern OSes force the MTU transmission, regardless of upper level data size, for efficiency

- Caveat: Jumbo Frame needs to be enabled on both Initiator, Target and all network devices along the path
Test Comparison iWARP, ROCEv2 & TCP

Eden Kim, Calypso
This Preliminary Review is intended to report Phase I data on comparison testing of iWARP, ROCEv2 and TCP transports with MTU Regular 1500B v Jumbo 9000B frames and comparison of high performance six drive LUNs identified as SSD-1 and lower performance six drive LUN identified as SSD-2. Data summary slides are set forth in the presentation.

The first pass tests are comprised of six test runs, each of which has 4 tests. SSD-1 runs take 6 hours while SSD-2 take 9 hours. Variables are MTU setting [1500B v 9000B], Transport [iWARP v ROCEv2 v TCP] and SSD [SSD-1 v SSD-2]

The intent of testing is to measure differences in performance for iWARP, ROCEv2 & TCP
Each transport is tested at MTU 1500B & 9000B to determine if there is a difference
Each transport is also tested against Optane 6 drive LUN and NVME 6 drive LUN

1. Synthetic RND 4K RW & SEQ 128K RW (corner case stress tests)
2. TC/QD Sweep using GPS Nav 9 IO Stream composite workload (to determine OIO saturation)
3. Replay test – Replay of GPS Nav real world workload (sequence of IO Streams & QDs)
4. Individual Streams test – Running each of the 9 IO Streams to Steady State

Test results presented herein represent first pass testing to compare Transport, MTU settings and Comparison of SSDs. Data has not yet been repeated and further testing, adjustment of test settings and optimization of platforms is intended in Phase II. Accordingly, some data points may represent anomalous conditions and/or system or software that needs optimization.
Test Set Up: Objectives, Test Platform & Test Settings

- **Objectives:** Saturate various Transports across 100Gb wire with different workloads & storage to compare performance
  1. Compare MTU IOPS & QoS: 1500 byte v 9000 byte with Synthetic 4K/128K
  2. Compare Transport: iWARP, RoCEv2, TCP with GPS Nav Demand Intensity & CPU Usage - TC/QD Sweep
  3. Compare Real World GPS Nav Portal Workload: Replay Test & Individual Streams
  4. Compare Drives: SSD-1 (high R/W IOPs) x6, SSD-2 (low R/W IOPs) x6

- **Test Platform – High Performance (CPU, RAM):**
  - CPU: Intel® Xeon® Platinum 8280L CPU @ 2.70GHz
  - Memory on Target: Two different types of RAM: 12 x 32GB @ 2933MHz; 12 x 256GB @ 2666MHz.
  - Memory on Initiator: 12 x 32GB @ 2933MHz
  - NIC – 100Gb – NIC, No Switch, Link Flow Control ON

- **Tests: Synthetic & Real World Workloads**
  - Asynchronous IO libaio stimulus generator & IO Traffic
    - Calypso CTS Stimulus Generator is on the host initiator server and applies test IOs across 100Gb ethernet wire to target storage server
    - IO traffic on the 100Gb wire are test IOs. There is no other application or driver IO traffic during the test
  - Synthetic Single IO Stream Workloads
    - Synthetic RND 4K, SEQ 128K RW – T4Q32, Total OIO=128
    - GPS Workload 9 IO Streams - T4Q32, Total OIO=128
  - Real World GPS Nav Workload
    - Replay test - Replay observed combinations of IO Streams and QDs
    - TC/QD Sweep (DIRTH) – Fixed 9 IO Streams at TC/QD sweep 1 – 576
**Phase I – Set-up and Test Plan**

**CTS Control Server**
- Test SW, DB, Test Scripts

**Host Server**
- IO Stimulus Generator
- Logical Storage

**NIC-Wire-NIC**
- MTU Setting
- Test I/Os

**Target Server**
- Target Storage
- Logical Storage

---

**Workload Tests:**
2. TC/QD Sweep – GPS Nav OIO 1-576
3. Replay Test – GPS Nav IO Sequence
4. Ind. Streams – Ind. IO Stream to SS

---

**Host Server (Initiator)**
- MTU

**Target Server (Target)**
- NIC

**100 Gb Ethernet**
- RoCE
- TCP

---

**Logical Storage**
- NIC

---

MTU 1500 b – Standard Frame
MTU 9000 b – Jumbo Frame

---

Higher Perf 6 SSD LUN
Lower Perf 6 SSD LUN
Workloads & Tests:  Real World GPS Nav Portal & Synthetic Corner Case

Real World GPS Nav Replay Test, TC/QD Sweep Test, Synthetic Corner Case Test, GPS Nav Individual Streams Test

- **Queue Depths**
- **IOPS**
- **IO Stream Combinations**

### Workload Capture
Each step of the workload has a different combination of the 9 IO Streams and QDs

### Replay Test
Sequence and Combination of IO Streams and QDs are replayed

### TC/QD Sweep Test
- Applies a fixed composite of all 9 IO Streams for each step of the test while running a range of TC/QDs from 1 - 16

### Synthetic Tests
- Run a single, or few, IO Stream at a fixed QD and duration after pre-conditioning and steady state

---

<table>
<thead>
<tr>
<th>IO Stream Combinations</th>
<th>IOPS</th>
<th>QD 1 ~ 16</th>
<th>TC 1 - 16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed 9 IO Streams</td>
<td>QD</td>
<td>TC</td>
<td></td>
</tr>
<tr>
<td>9 IO Streams = 78% of Total IO Streams</td>
<td>Ave QD = 15</td>
<td>Total Max OIO = 576</td>
<td></td>
</tr>
<tr>
<td>100% Write IOs</td>
<td>Median QD = 8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replay Sequence &amp; Combination IO Stream &amp; QDs</td>
<td>Max QD = 368</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>AO</th>
<th>GPS 9 IO IO Streams</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ 4K W</td>
<td>24.8%</td>
</tr>
<tr>
<td>RND 16K W</td>
<td>17.7%</td>
</tr>
<tr>
<td>SEQ 16K W</td>
<td>14.4%</td>
</tr>
<tr>
<td>SEQ 0.5K W</td>
<td>13.9%</td>
</tr>
<tr>
<td>SEQ 1K W</td>
<td>5.8%</td>
</tr>
<tr>
<td>RND 4K W</td>
<td>4.5%</td>
</tr>
<tr>
<td>RND 1K W</td>
<td>2.65%</td>
</tr>
<tr>
<td>RND 28K W</td>
<td>2.54%</td>
</tr>
<tr>
<td>SEQ 1.5K W</td>
<td>2.44%</td>
</tr>
</tbody>
</table>

---

- **21 | ©2020 Storage Networking Industry Association. All Rights Reserved.**
1. Compare MTU IOPS & QoS: 1500B v 9000B – SSD-1 - iWARP v ROCEv2 v TCP

- iWARP v ROCEv2 v TCP – 1500B v 9000B
  - Comparison Plot - Synthetic RND 4K RW; SEQ 128K RW

**Observations**

- **IOPS** – Substantially equivalent for all Transports
- **QoS** – High QoS spikes are observed for Read workloads
- **CPU** – CPU System Usage % are very low – typically less than 2%

Note: Demand Intensity OIO=128 (T4/Q32) Pre-conditioned to SNIA PTS Steady State Each Workload Segment = 5 Minutes
### 1a. MTU: 1500B v 9000B IOPS & QoS – SSD-1: All Workloads, iWARP

<table>
<thead>
<tr>
<th>Drive Family</th>
<th>Auto-generated Multi WSAT</th>
<th>IOPS</th>
<th>TP</th>
<th>ART</th>
<th>5 sec</th>
<th>9 sec</th>
<th>MRT</th>
<th>Power</th>
<th>Auto DIRTH</th>
<th>IOPS</th>
<th>TP</th>
<th>ART</th>
<th>5 sec</th>
<th>9 sec</th>
<th>MRT</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVMe0.3xF-1500</td>
<td>SSD-1 MTU 1500 iWARP 2250 GB</td>
<td>RND 4K/ SEQ 128K RW</td>
<td>TC/QD Sweep</td>
<td>9 Individual IO Streams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVMe0.3xF-9000</td>
<td>SSD-1 MTU 9000 iWARP 2250 GB</td>
<td>RND 4K/ SEQ 128K RW</td>
<td>TC/QD Sweep</td>
<td>9 Individual IO Streams</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Key Pts.**

- **IOPS** – Generally no significant difference
- **QoS** – Higher RTs for RND 4K/SEQ 128K Reads

**Replay Test**

- Replay Workload QD Varies between 7 – 368
- TC/QD Sweep DIRTH QD range 1 – 576
- Pre-conditioned to SNIA PTS Steady State

Note: Synthetic Demand Intensity = 128 (T4/Q32)
2. Compare Transport: Demand Intensity TC/QD Sweep, SSD-1, iWARP v ROCEv2 v TCP

- TC/QD Sweep DIRTH – Demand Intensity comparison
  - Optimal Demand Intensity, MB/s, QoS comparison (red box) where QoS is the Figure of Merit
  - QoS Ceiling (purple dotted line), IOPS & MB/s dips (teal circle)

**Observations**

**iWARP 1500B: MB/s, QoS & ART v Total OIO**

**ROCEv2 1500B: MB/s, QoS & ART v Total OIO**

**TCP 1500B: MB/s, QoS & ART v Total OIO**

**TC/QD test:** MB/s & QoS are plotted as Demand Intensity increases over a range of OIO 1 – 576  
**Figure of Merit:** QoS is just before QoS dramatically increases – here at OIO 72-144  
**Saturation:** QoS saturates where RTs exceed 1 mS Ceiling – here at RDMA OIO 288 and TCP OIO 144  
**Note:** RDMA continues to increase is MB/s with small increase in QoS whereas TCP has high QoS across all OIO
3. Compare Real World Workload GPS Nav: Replay Test – iWARP v ROCEv2 v TCP

Replay Test: Replays the Sequence of IO Stream combinations and QDs observed in the workload capture

iWARP v ROCEv2 v TCP
- iWARP & ROCEv2 substantially similar IOPS & QoS
- TCP has lower IOPS and higher QoS than iWARP & ROCEv2
- Note: Replay Test Values are average across total Replay Test. i.e. Variations due to different IO Stream combinations and QDs are averaged

GPS - 9 IO Streams
- Replay Test: 720 IO Capture Steps are applied after Steady State for a 1 min interval or 9 min duration

**Key Pts.**
- **iWARP & ROCEv2:** Substantially similar IOPS & QoS
- **TCP:** Lower IOPS & Higher QoS than iWARP & ROCEv2

<table>
<thead>
<tr>
<th>IO Stream</th>
<th>iWARP</th>
<th>ROCEv2</th>
<th>TCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEQ 4K W</td>
<td>24.8%</td>
<td>28.0</td>
<td></td>
</tr>
<tr>
<td>RND 16K W</td>
<td>17.7%</td>
<td>19.9</td>
<td></td>
</tr>
<tr>
<td>SEQ 16K W</td>
<td>14.4%</td>
<td>16.2</td>
<td></td>
</tr>
<tr>
<td>SEQ 0.5K W</td>
<td>13.9%</td>
<td>15.7</td>
<td></td>
</tr>
<tr>
<td>SEQ 1K W</td>
<td>5.8%</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>RND 4K W</td>
<td>4.5%</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>RND 1K W</td>
<td>2.65%</td>
<td>2.98</td>
<td></td>
</tr>
<tr>
<td>RND 28K W</td>
<td>2.54%</td>
<td>2.86</td>
<td></td>
</tr>
<tr>
<td>SEQ 1.5K W</td>
<td>2.44%</td>
<td>2.75</td>
<td></td>
</tr>
</tbody>
</table>
3a. Compare GPS Nav: Individual Streams Test – iWARP v ROCEv2 v TCP

Individual Streams Test: Each IO Stream measured at SNIA PTS Steady State for 5 min

iWARP v ROCEv2 v TCP

IOPS:
- TCP Lower for all block sizes than iWARP & ROCEv2
- For 0.5K, 1K, 1.5K – iWARP slightly higher than ROCEv2

QoS:
- ROCEv2 Lower for all block sizes than iWARP & TCP
- TCP higher for 4K, 16K

Key Pts.

IOPS – iWARP generally higher ROCEv2, TCP Lower especially for smaller block sizes
QoS – ROCEv2 generally lower than iWARP. TCP higher for 4K & 16K block sizes
Note: Smaller block IO Streams negatively affect TCP IOPS & QoS, less impact on RDMA
### 4. Compare Drives: Synthetic RND 4K, SEQ128K RW - iWARP v ROCEv2 v TCP

- **SSD-1 vs SSD-2: RND 4K RW & SEQ 128K RW**

<table>
<thead>
<tr>
<th></th>
<th>SSD-1</th>
<th>SSD-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTU</td>
<td>1500</td>
<td>1500</td>
</tr>
<tr>
<td>iWARP</td>
<td>2250 GB</td>
<td>24004 GB</td>
</tr>
<tr>
<td>ROCEv2</td>
<td>2250 GB</td>
<td>24004 GB</td>
</tr>
<tr>
<td>TCP</td>
<td>2250 GB</td>
<td>24004 GB</td>
</tr>
</tbody>
</table>

**Key Pts.**

- **IOPS** – SSD-1 has higher or substantially equivalent IOPS compared to SSD-2
- **QoS** – SSD-1 has lower or similar QoS compared to SSD-2 except for iWARP Reads
4a. TC/QD Sweep: SSD-1 v SSD-2, MTU 1500B, iWARP

Optimal OIO (red box), QoS Ceiling (purple dotted line), IOPS & MB/s dips (teal circle)

Optimal OIO = 144
- IOPS, ART, QoS increase to Max OIO
- SSD-1 - 2,740 MB/s, 1.1 mS QoS
- SSD-2 - 449 MB/s, 27.68 mS QoS

MB/s Dips
- SSD-1 - MB/s dip at OIO 16
- SSD-2 – MB/s dips at OIO 8,16,32,40

QoS Saturation
- SSD-1 - OIO=288 at 2.05 mS
- SSD-2 - OIO=288 at 37.55 mS

Max OIO=576
- SSD-1 - 2,988 MB/s, 4.13 mS QoS
- SSD-2 – 624 MB/s, 49.01 mS QoS

NOTE: SSD-2 shows lower MB/s and higher QoS across all OIO

IOPS & MB/s continue to increase at Max OIO for both SSDs (absence of plateau or leveling off)

SSD-1 shows higher MB/s and lower QoS. Max OIO=576 2,988 MB/s, 4.13 mS
SSD-2 shows lower MB/s and higher QoS. Max OIO=576 624 MB/s, 49.01 mS

NOTE: SSD-2 shows lower MB/s and higher QoS across all OIO
4b. TC/QD Sweep: SSD-1 v SSD-2, MTU 1500B, ROCEv2

Optimal OIO (red box), QoS Ceiling (purple dotted line), IOPS & MB/s dips (teal circle)

Optimal OIO = 144
- IOPS, ART, QoS increase to Max OIO
- SSD-1 - 2,977 MB/s, 1.15 mS QoS
- SSD-2 - 427 MB/s, 29.52 mS QoS

MB/s Dips
- SSD-1 - MB/s dip at OIO 16
- SSD-2 – MB/s dips at OIO 8,16,32,40

QoS Saturation
- SSD-1 - OIO=288 at 2.2 mS
- SSD-2 - OIO=288 at 39.18 mS

Max OIO=576
- SSD-1 - 2,977 MB/s, 5.5 mS QoS
- SSD-2 – 590 MB/s, 52.16 mS QoS

TC/QD Sweep test runs a fixed composite 9 IO Stream combination from the GPS Nav workload across a range from OIO=1 to OIO=576

SSD 1 ROCEv2 shows higher MB/s and lower QoS. Performance substantially similar to iWARP SSD-1
SSD 2 ROCEv2 shows lower MB/s and higher QoS. Performance substantially similar to iWARP SSD-1

NOTE: ROCEv2 SSD-2 shows lower MB/s and higher QoS across all OIO
4c. TC/QD Sweep: SSD-1 v SSD-2, MTU 1500B, TCP

Optimal OIO (red box), QoS Ceiling (purple dotted line), IOPS & MB/s dips (teal circle)

**Optimal OIO = 72**
- IOPS, ART, QoS increase to Max OIO
- **SSD-1** - 2,839 MB/s, 1.29 mS QoS
- **SSD-2** - 390 MB/s, 18.31 mS QoS

**MB/s Dips**
- **SSD-1** - MB/s dip at OIO 16,32
- **SSD-2** - MB/s dips at OIO 8,16,32,40

**QoS Saturation**
- **SSD-1** - OIO=144 at 3.89 mS
- **SSD-2** - OIO=80 at 28.9 mS

**Max OIO=576**
- **SSD-1** - 2,945 MB/s, 60.41 mS QoS
- **SSD-2** - 607 MB/s, 50.84 mS QoS

**Key Pts.**

**SSD-1** TCP shows higher MB/s and lower QoS. Performance substantially Lower to RDMA SSD-1

**SSD-2** TCP shows lower MB/s and higher QoS. Performance substantially Similar to RDMA SSD-1

**NOTE:** TCP SSD-2 shows lower MB/s and higher QoS across all OIO

TC/QD Sweep test runs a fixed composite 9 IO Stream combination from the GPS Nav workload across a range from OIO=1 to OIO=576
5. Test Data: Preliminary Observations & Conclusions

CPU: Overall low CPU utilization on storage target: 1%-4%, regardless of offload (iWARP, RoCEv2) or non-offload(TCP)

MTU: There appears to be nominal difference between 1500B & 9000B MTU
MTU 1500B has better response times for small block IO Streams (0.5K W, 1K W, 1.5K W)

SSD-1 v SSD-2: SSD-1 has significantly Higher IOPS, MB/s & Lower QoS than SSD-2
However, SSD-2 shows higher performance for small block 0.5K, 1K & 1.5K and SEQ 128K Reads
Note: additional testing is planned to validate findings, iterate test settings and conditions and to optimize test platform set up

TC/QD Sweep: GPS Nav Workload shows increasing IOPS & MB/s at highest OIO
However, ART and QoS increase at high OIO, probably higher than an acceptable application RT Ceiling
Synthetic workload OIO are up to 576 for TC/QD, 128 for RND 4K/SEQ 128K & Ind Streams, and 576 for Replay test.
Note: GPS Nav is a 100% W workload. Future tests should run a real world workload with 35% Reads (e.g. Retail Web Portal 68:32 RW)

Transport: 1. iWARP & ROCEv2 are substantially similar and faster than TCP
2. iWARP optimal OIO=144, ROCEv2 and TCP optimal OIO=72
3. OIO saturation occurs at 288 – 576 for GPS Nav 9 IO Stream workload
4. iWARP & ROCEv2 have lower QoS response times than TCP for GPS Nav workload
5. ROCEv2 has lower QoS response times than iWARP
Questions?

• Please submit to Q & A panel during the session
• Otherwise, please submit to Brighttalk Portal
• Contact fred.zhang@Intel.com or info@calypsotesters.com

Thank You!
Additional SNIA Resources on NVMe-oF

- SNIA Video “Intro to NVMe-oF” YouTube Playlist:
  - https://www.youtube.com/playlist?list=PLH_ag5Km-YUapfuug7nnwCpaegJVO2DZE

- SNIA Educational Library:
  - https://www.snia.org/educational-library?search=NVMe+over+Fabrics
After this Webcast

- Please rate this webcast and provide us with your feedback
- This webcast and a copy of the slides will be available at the SNIA Educational Library [https://www.snia.org/educational-library](https://www.snia.org/educational-library)
- A Q&A from this webcast, including answers to questions we couldn’t get to today, will be posted on our blog at [https://sniansfblog.org/](https://sniansfblog.org/)
- Follow us on Twitter [@SNIANSF](https://twitter.com/SNIANSF)
Thank You