Accelerating NVMe™ over Fabrics with Hardware Offloads at 100Gb/s and Beyond

Rob Davis, Mellanox
Ilker Cebeli, Samsung
Disclaimer

This presentation and/or accompanying oral statements by Samsung representatives collectively, the “Presentation”) is intended to provide information concerning the SSD and memory industry and Samsung Electronics Co., Ltd. and certain affiliates (collectively, “Samsung”). While Samsung strives to provide information that is accurate and up-to-date, this Presentation may nonetheless contain inaccuracies or omissions. As a consequence, Samsung does not in any way guarantee the accuracy or completeness of the information provided in this Presentation.

This Presentation may include forward-looking statements, including, but not limited to, statements about any matter that is not a historical fact; statements regarding Samsung’s intentions, beliefs or current expectations concerning, among other things, market prospects, technological developments, growth, strategies, and the industry in which Samsung operates; and statements regarding products or features that are still in development. By their nature, forward-looking statements involve risks and uncertainties, because they relate to events and depend on circumstances that may or may not occur in the future. Samsung cautions you that forward looking statements are not guarantees of future performance and that the actual developments of Samsung, the market, or industry in which Samsung operates may differ materially from those made or suggested by the forward-looking statements in this Presentation. In addition, even if such forward-looking statements are shown to be accurate, those developments may not be indicative of developments in future periods.
NVMe™ over Fabrics Market Success

Driving New Level of Performance for Your Multi-cloud and AI Workloads with NVMe Storage Solutions

For Storage end users, the advent of the Non-Volatile Memory Express (NVMe) protocol simply validates the vision and engineering objectives IBM has pursued for decades.

- February 2018: IBM will announce NVMe over Fabrics (NVMe-F) capabilities for IBM FlashSystem 900, IBM Spectrum Accelerate, and IBM Spectrum Virtualize families.

The bottom line: IBM is committed to providing end-to-end support for NVMe, to prepare our customers for the future.

NetApp announces the first available NVMe-over-InfiniBand enterprise-class systems

SUNNYVALE, Calif. — September 19, 2017 — NetApp (NASDAQ: NTAP) today introduced a powerful system that has set new records for price/performance. NetApp has also added new software to its big data analytics portfolio, including the first available NVMe-enabled solution on the market and a NetApp® Converged Infrastructure Solution for Data Analytics. These new offerings provide ultrafast performance and the building blocks for analytics data management that is fast, secure, efficient, and future-proof.

Mellanox and Huawei Advance RDMA Technology with Leading-Edge NVMe Over Fabrics Solution

SHANGHAI, CHINA – Aug. 31, 2016 – Mellanox® Technologies, Ltd. (NASDAQ: MLNX), a leading supplier of high performance cloud and storage networking solutions, today announced that it will preview a new leading-edge NVMe-oF™ (NVMe Express® over Fabrics) solution, jointly developed by Mellanox and Huawei, at this week’s HUAWEI CONNECT Conference (HCC 2016) in Shanghai, Aug. 31 – Sept. 2.
NVMe™ over Fabrics Maturity

UNH-IOL provides a neutral environment for multi-vendor interoperability and conformance to standards testing since 1988

In May 2017 and again in October they hosted the first and second test for NVMe-oF™

Test plans called for participating vendors to mix and match their NICs in both Target and Initiator positions

Testing was successful with near line rate performance at 25Gb/s achieved at the first test
Time for the Next Level of Performance

- Current Performance
  - 6M IOPs, 512B block size
  - 2M IOPs, 4K block side
  - 50% CPU utilization
  - ~15μsec latency difference from local

- How do we lower the latency difference and CPU utilization?
Some of the use cases for NVMe™ Over Fabrics

1. Software-Defined Storage (SDS)
2. Hyper-Converged
3. Disaggregated JBOF Storage
4. Classic SAN
5. Direct Attached JBOF SAS DAS Replacement
Performance Test Configuration – 2016

- 1x NVMe-oF™ target
  - 24x NVMe 2.5” SSDs
  - 2x 100GbE NICs
  - Dual x86 CPUs

- 4x initiator hosts
  - 2x25GbE NICs each

- Open Source NVMe-oF kernel drivers
Local vs. Remote Latency Comparison – 2016

<table>
<thead>
<tr>
<th></th>
<th>Read Gap</th>
<th>Write Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>~17 us</td>
<td>~9 us</td>
</tr>
</tbody>
</table>
Performance Test Configuration – 2017

- **1x NVMe-oF™ target**
  - 36x NF1 SSDs
  - 2x 100GbE NICs, 2x 50GbE NICs
  - Dual x86 CPUs

- **6x initiator clients**
  - 2x25Gb/s each

- **Open Source NVMe-oF kernel drivers**
  - Ubuntu Linux 16.04/4.9 on Target
Local vs. Remote Latency Comparison - 2017

2017 Tests

<table>
<thead>
<tr>
<th>Reading Gap</th>
<th>Writing Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>~14 us</td>
<td>~10 us</td>
</tr>
</tbody>
</table>

2016 Tests

<table>
<thead>
<tr>
<th>Reading Gap</th>
<th>Writing Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>~17 us</td>
<td>~9 us</td>
</tr>
</tbody>
</table>
SSDs Will Continue to get Faster

2017 Tests

<table>
<thead>
<tr>
<th></th>
<th>Read Gap</th>
<th>Write Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDDs</td>
<td>~14 us</td>
<td>~10 us</td>
</tr>
</tbody>
</table>

2016 Tests

<table>
<thead>
<tr>
<th></th>
<th>Read Gap</th>
<th>Write Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDDs</td>
<td>~17 us</td>
<td>~9 us</td>
</tr>
</tbody>
</table>
Closing the Local vs. Remote Performance Gap

~15usec latency (not including SSD)  ~5usec latency (not including SSD)
50% CPU utilization                 0.01% CPU utilization
No Offload - Initiator Requests and Responses to Target Go Through Software
How Offload Works

Offload

- Only control path, management and exceptions go through Target CPU software
- Data path and NVMe™ commands handled by the network adapter
Results – CPU Utilization – Latency

Offload Local vs. Remote Latency
~5usec
NVMe-oF™ Offload Magnifies CMB Value

- Use SSD and Network Adapter with NVMe-oF™ protocol offload to bypass CPU memory controller completely.
Summary

- NVMe™ over Fabrics is taking off
  - Large and small vendors in productions
  - Multi-vendor interoperability
- NVMe over Fabrics protocol offload moves the performance even closer to local SSD
- Control Memory Buffer (CMB) value is dramatically enhanced with NVMe over Fabrics protocol offload on the network adapter

8M IOPs, 512B block size
5M IOPs, 4K block size
0.01% CPU utilization
~5usec latency (not including SSD)