NVMe-oF Through PCIe Gen4

H3 Platform

Brian Pan
Drive 200G NVMe-oF by Using PCIe Gen4 Solution
Architecture of NVMe-oF through PCIe Gen4

**IBM P9 with 100G NIC**
- IBM LC922
- Mellanox CX5

**JBOF+ Broadcom Stingray 100G Smart NIC**
- Broadcom Atlas PCIe Gen4 switch
- Broadcom Stingray

**NVMe SSD**
- Phison PCIe Gen4 NVMe SSD
- Intel PCIe Gen3 NVMe SSD
Testing System– PCIe Gen4 JBOf
JBOF Specification

- PCIe switch
  - Broadcom 88096 Atlas PCIe switch with internal Synthetic mode
- Host connection
  - 1x PCIe Gen4 x16 to LC922
  - 1x Broadcom Stingray 100G Smart NIC
- NVMe SSD
  - 5x Phison 5016-E16 NVMe SSD
  - 1x Intel 900P
NVMe-oF Target Setup

- LC922 with Mellanox Connect X5
  - End to end PCIe Gen4
  - NVMe SSD $\rightarrow$ PCIe switch $\rightarrow$ IBM P9 CPU $\rightarrow$ Mellanox CX5 $\rightarrow$ 100G switch

- Broadcom Stingray+ PCIe Gen4 switch
  - PCIe Gen3 smart NIC $\rightarrow$ PCIe Gen4 switch
  - NVMe SSD $\rightarrow$ PCIe switch $\rightarrow$ Stingray $\rightarrow$ 100G switch
NVMe-oF— 19,209 MB/s

<table>
<thead>
<tr>
<th>Transfer Size (Sequential)</th>
<th>Initiator_1</th>
<th>Initiator_2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MB/s</td>
<td>IOPS</td>
</tr>
<tr>
<td>128K read</td>
<td>9,260</td>
<td>71.2k</td>
</tr>
<tr>
<td>128K write</td>
<td>8,537</td>
<td>64.7k</td>
</tr>
</tbody>
</table>

NOTES:
1. Performance measured using FIO rev 3.1, with 8 workers with Queue Depth of 64 and using Linux in-box NVMe driver.
2. Initiator_1 is accessing to Target_1 with 3x NVMe
3. Initiator_2 is accessing to Target_2 with 2x NVMe
4. Initiator_1 and Initiator_2 are simultaneously accessing with NVMeoF targets
# Bandwidth - Direct Attached

<table>
<thead>
<tr>
<th>Transfer Size (Sequential)</th>
<th>Server_1 (NVMe x2)</th>
<th>Server_2 (NVMe x3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MB/s</td>
<td>IOPS</td>
</tr>
<tr>
<td>128K read</td>
<td>9,664</td>
<td>75.5k</td>
</tr>
</tbody>
</table>

## NOTES:

1. Performance measured using FIO rev 3.1, with 8 workers with Queue Depth of 64 and using Linux in-box NVMe driver.
2. Server_1 is assigned with 2x NVMe
3. Server_2 is assigned with 3x NVMe
4. Server_1 and Server_2 are simultaneously assessing with 5x NVMe
## Latency of NVMe-oF

<table>
<thead>
<tr>
<th>Transfer Size (Sequential)</th>
<th>Initiator_1 Avg. (usec)</th>
<th>Initiator_2 Avg. (usec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4K read</td>
<td>41.3</td>
<td>38.6</td>
</tr>
<tr>
<td>4k write</td>
<td>39.7</td>
<td>35.9</td>
</tr>
</tbody>
</table>

**NOTES:**
1. Performance measured using FIO rev 3.1, with 1 worker with total Queue Depth of 1 and using Linux in-box NVMe driver.
# Latency of Direct-attached

<table>
<thead>
<tr>
<th>Transfer Size</th>
<th>Server_1 Avg. (usec)</th>
<th>Server_2 Avg. (usec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4K read</td>
<td>18.2</td>
<td>17.7</td>
</tr>
<tr>
<td>4k write</td>
<td>23.8</td>
<td>18.9</td>
</tr>
</tbody>
</table>

**NOTES:**
1. Performance measured using FIO rev 3.1, with 1 worker with total Queue Depth of 1 and using Linux in-box NVMe driver.
Benefits of PCIe Gen4 Solution

- **Performance**
  - Almost double performance compared PCIe Gen3

- **Cost saving**
  - By using PCIe Gen4 solution, NVMe-oF can support more initiators
  - Only one PCIe Gen4 x16 support 200Gbps ethernet connection
Brian Pan | H3
GM
huaiyangpan
www.h3platform.com
brian.pan@h3platform.com
+886 2 2698 3800#110