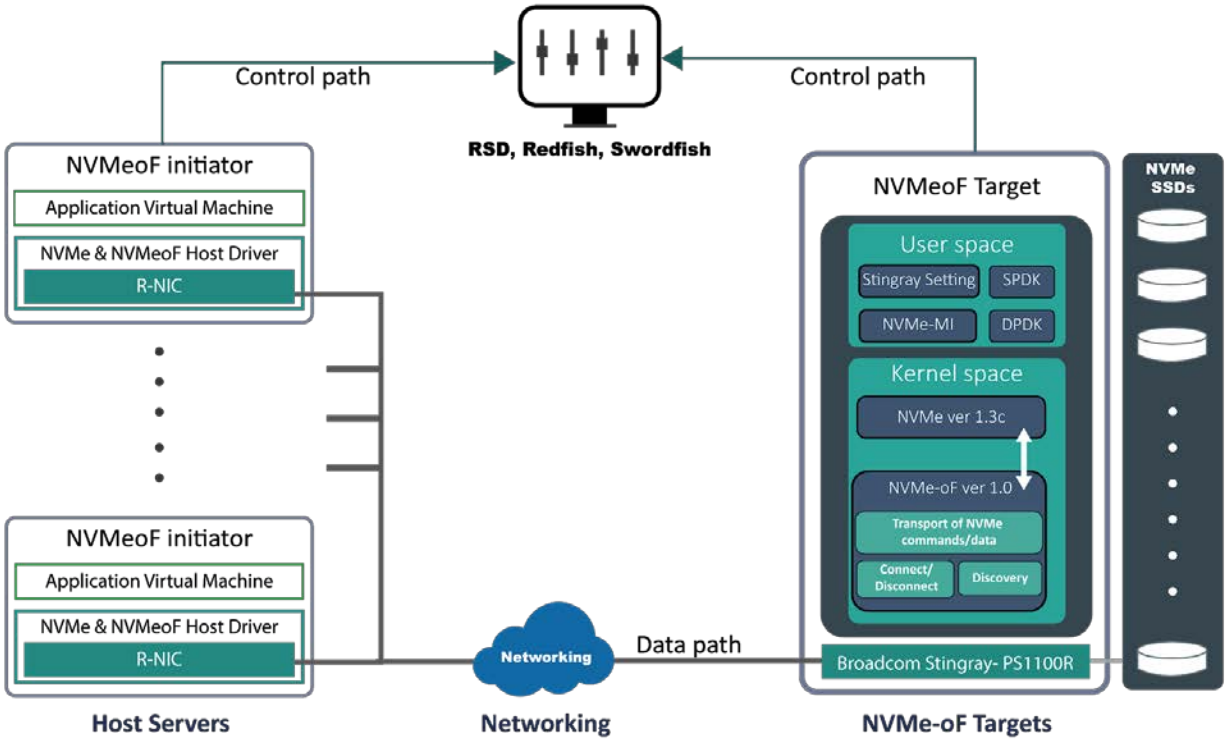
NVMe-oF Through PCIe Gen4

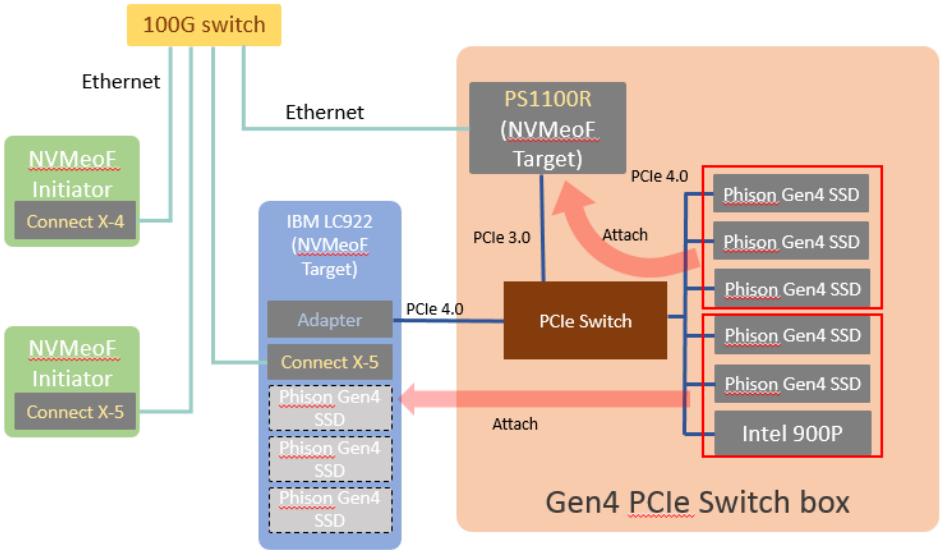
H3 Platform

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

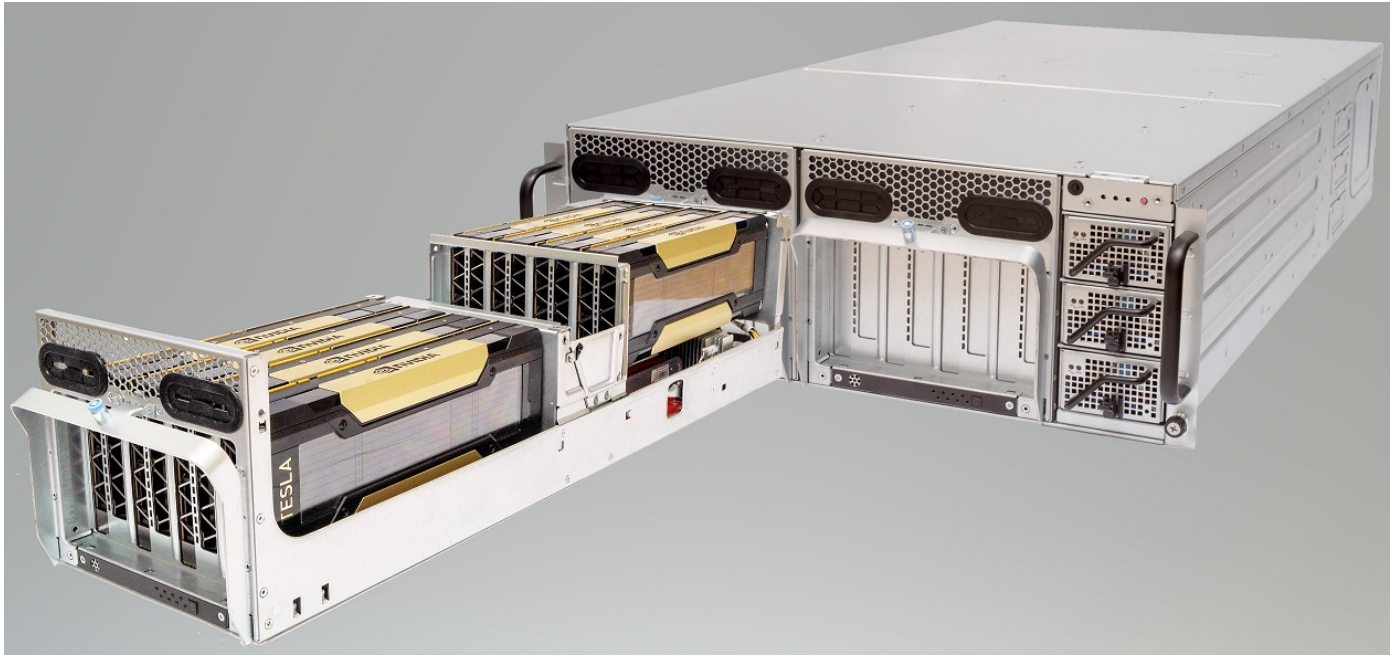
JBOf+ Broadcom Stingray 100G Smart NIC

- **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→ PCIe switch→ IBM P9 CPU→ Mellanox CX5→ 100G switch
- Broadcom Stingray+ PCIe Gen4 switch
  - PCIe Gen3 smart NIC+ PCIe Gen4 switch
  - NVMe SSD→ PCIe switch→ Stingray→ 100G switch

NVMe-oF– 19,209 MB/s

Bandwidth - NVMe over Fabric

Transfer Size (Sequential)	Initiator_1		Initiator_2	
	MB/s	IOPS	MB/s	IOPS
128K read	9,260	71.2k	9,949	75.4k
128K write	8,537	64.7k	7,577	57.2k

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 assessing with NVMeoF targets

Direct-attached (Gen4 x16)– 23,772MB/s

Bandwidth - Direct Attached

Transfer Size (Sequential)	Server_1 (NVMe x2)		Server_2 (NVMe x3)	
	MB/s	IOPS	MB/s	IOPS
128K read	9,664	75.5k	14,108	113k

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

Latency - NVMe over Fabric

Transfer Size (Sequential)	Initiator_1 Avg. (usec)	Initiator_2 Avg. (usec)
4K read	41.3	38.6
4k write	39.7	35.9

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

Latency - Direct Attached (concurrent)

Transfer Size (Sequential)	Server_1 Avg. (usec)	Server_2 Avg. (usec)
4K read	18.2	17.7
4k write	23.8	18.9

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