High Performance NVMe Over 100G iWARP RDMA

Using White Box JBOF Storage Platform & Chelsio T6 Adapter

Executive Summary

NVMe over Fabrics specification extends the benefits of NVMe to large fabrics, beyond the reach and scalability of PCIe. NVMe enables deployments with hundreds or thousands of SSDs using a network interconnect, such as RDMA over Ethernet. T6 iWARP RDMA provides a low latency, high throughput, plug-and-play Ethernet solution for connecting high performance NVMe SSDs over a scalable, congestion controlled and traffic managed fabric, with no special configuration needed. This paper presents the performance results of Chelsio NVMe-oF over 100GbE iWARP fabric in a White Box JBOF Storage Platform setup with a Microsemi PCIe Switch. The Chelsio NVMe solution delivers line-rate throughput performance of 93 Gbps and 2.4M IOPS (at 4K I/O size).

Test Results

The following graph presents NVMe-oF READ, WRITE IOPS and throughput results of Chelsio iWARP solution using null block device as storage array. The results are collected using the **fio** tool with I/O size varying from 4k to 512k bytes with an access pattern of random READs and WRITEs.

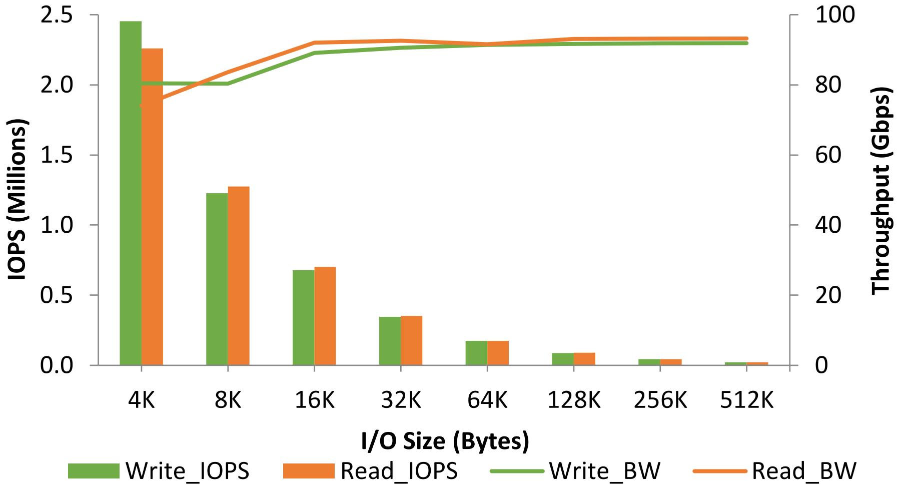


Figure 1 – Null Block Device READ, WRITE Throughput & IOPS vs. I/O size

Using null block device as storage array, T6 solution delivers 93 Gbps line-rate throughput for both READ and WRITE operations. The WRITE IOPS exceeds 2.4M at 4K I/O size.

The following graph presents READ, WRITE IOPS and throughput results using SSD as storage array. *Please note that WRITE throughput and IOPS numbers are limited by SSD performance.*

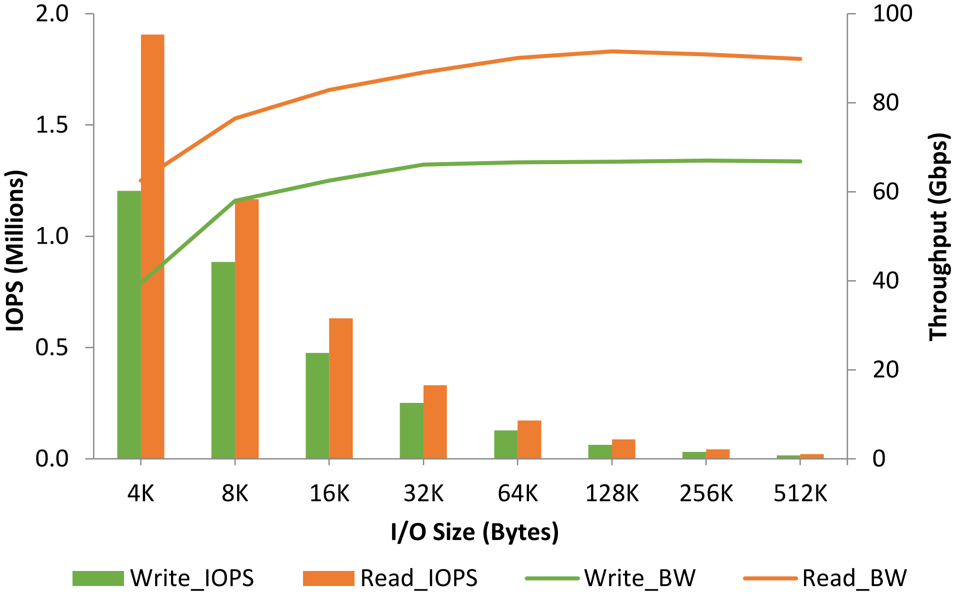
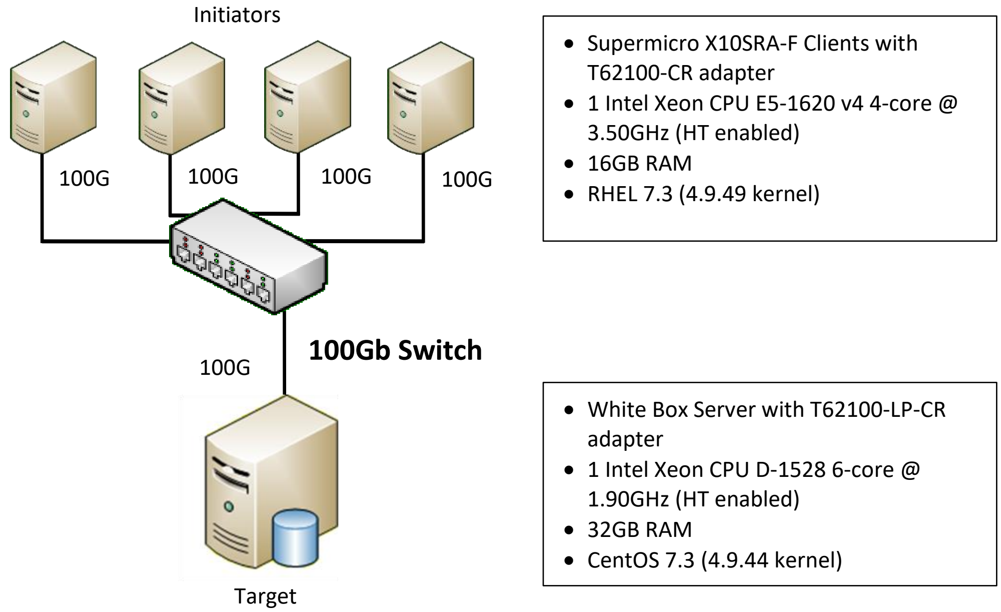


Figure 2 - SSD READ, WRITE Throughput & IOPS vs. I/O size

Chelsio's T6 solution delivers a commendable performance using SSD with READ Throughput of 91.5 Gbps and READ IOPS of 1.9M. *Further performance tuning is in progress.*

Test Setup

Topology



- Supermicro X10SRA-F Clients with T62100-CR adapter
- 1 Intel Xeon CPU E5-1620 v4 4-core @ 3.50GHz (HT enabled)
- 16GB RAM
- RHEL 7.3 (4.9.49 kernel)

- White Box Server with T62100-LP-CR adapter
- 1 Intel Xeon CPU D-1528 6-core @ 1.90GHz (HT enabled)
- 32GB RAM
- CentOS 7.3 (4.9.44 kernel)

Figure 3 – Test Setup

The setup consists of a target machine connected to 4 initiator machines through a 100GbE switch using single port on each system. Standard MTU of 1500B is used. Chelsio Unified Wire v3.6.0.3 is installed on each machine.

Storage configuration

For null block setup, 2 targets are configured, each with 1 null block device of 1GB size. 2 Initiators are used and each of them connects to 1 target.

For the SSD setup, 8 targets are configured, each with 1 Samsung NVMe PCIe SSD of 800GB size. 4 Initiators are used and each of them connects to 2 targets.

Commands used

Null block:

```
[root@host~]# nvme connect -t rdma -a <Target IP> -n nvme-nullbX -s 4420
[root@host~]# fio --name=random --iodepth=32 --rw=randwrite/randread --size=900m
--invalidate=1 --direct=1 --numjobs=8 --bs=<blocksize> --runtime=30 --time_based
--ioengine=libaio --fsync_on_close=1 --group_reporting --filename=<device1>
```

SSD:

```
[root@host~]# nvme connect -i 2 -t rdma -a <Target IP> -n nvme-nullbX -s 4420
[root@host~]# fio --name=random --iodepth=64 --rw=randwrite/randread --size=900m
--invalidate=1 --direct=1 --numjobs=16 --bs=<blocksize> --runtime=30 --time_based
--ioengine=libaio --fsync_on_close=1 --group_reporting --
filename=<device1:device2>
```

Conclusion

This paper showcases the remote storage access performance capabilities of Chelsio T6 NVMe-oF over 100GbE iWARP fabric solution in a White Box JBOF Storage Platform setup with T62100-LP-CR Unified Wire adapter and Microsemi PCIe Switch. Using iWARP RDMA enables the NVMe based storage to be shared, pooled and managed more effectively across a low latency, high performance network. The results show that Chelsio's NVMe over iWARP RDMA solution achieves:

- Line-rate throughput of 93 Gbps for READ using null block device and a high 91.5 Gbps using SSD.
- READ IOPS exceeding 2.4M using null block device and 1.9M using SSD.

Related Links

100G NVMe over Fabrics JBOF
100G NVMe over Fabrics for AMD EPYC
T6 100G NVMe-oF demonstration