High Performance 25G S2D for AMD EPYC

Using AMD EPYC 7551 Platform & Chelsio T6 iWARP RDMA Adapter

Overview

AMD EPYC, industry's first hardware-embedded x86 server security solution, is a system on chip (SoC) which provides exceptional processing power coupled with high-end memory and I/O resources to meet workload demands of any scale, from virtualized infrastructures to cloud-era datacenters. The combination of the AMD EPYC 7551 servers with Chelsio's industry-leading Unified Wire adapter solution delivers compelling performance, power and total cost of ownership (TCO) advantages. This enables innovative topologies and networked computing models to address the most demanding processing needs.

Microsoft **Storage Spaces Direct** (S2D) is a feature which enables building highly available and scalable storage systems by pooling local server storage. You can now build HA Storage Systems using storage nodes with only local storage, which can be disk devices that are internal to each storage node. This not only eliminates the need for a shared SAS fabric and its complexities, but also enables using devices such as SATA solid state drives, which can help further reduce cost or NVMe solid state devices to improve performance. S2D leverages SMB3 for all intra-node communication, including SMB Direct and SMB Multichannel, for low latency and high throughput storage. This paper presents S2D performance results on EPYC servers running Windows Server 2019 using Chelsio iWARP RDMA technology in a hyper-**converged** deployment scenario.

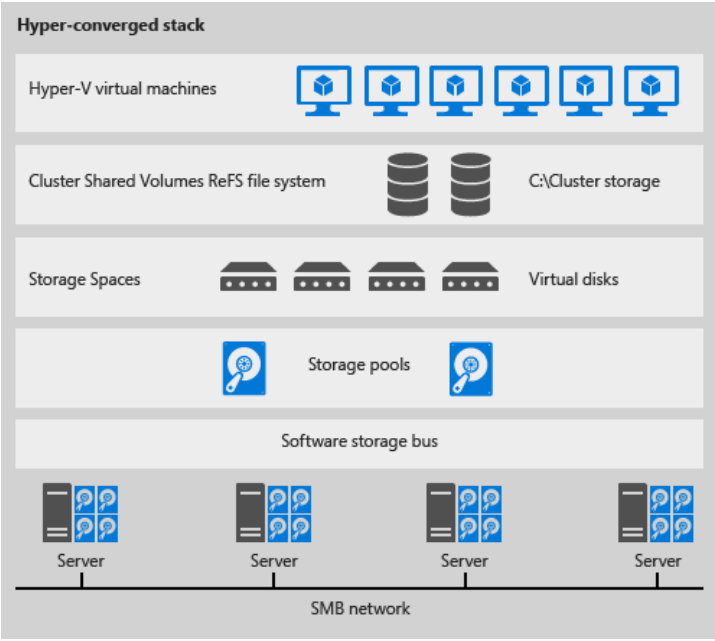

Figure 1 – S2D Hyper-Converged Stack

Why Chelsio iWARP RDMA Solution for S2D

Chelsio's sixth generation (T6), high performance iWARP RDMA 1/10/25/40/50/100GbE adapters enable incremental, non-disruptive server installs, and support the ability to work with any legacy (non-DCB) switch infrastructure. This enables a decoupled server and switch upgrade cycle and delivers a brownfield strategy to enable high performance, low cost, scalable S2D deployment.

iWARP has been an IETF standard (RFC 5040) for 11 years, TCP/IP has been an IETF standard (RFC 793, 791) for 37 years. iWARP Inherits the loss resilience and congestion management from underlying TCP/IP stack and enables a very high performance, extremely low latency, high bandwidth and high message rate solution. iWARP presents no surprises, no fine print, and is a plug and play solution. It is scalable to wherever the datacenter can scale to.

Network QoS is used in **hyper-converged** configuration to ensure that the Software-Defined-Storage system has enough bandwidth to communicate between the nodes to ensure resiliency and performance. Chelsio's iWARP RDMA enabled Unified Wire Ethernet adapters with enhanced rate-limiting (network QoS) features offload bandwidth allocation to the adapter bypassing the operating system. This eliminates the need for a DCB enabled Ethernet switch to implement Storage Spaces Direct (S2D) in a hyper-converged mode, resulting in reduced total ROI and simplified management.

Microsoft also recommends and prefers to use iWARP RDMA as it is easier to configure/setup, scalable, routable and works with any standard ethernet switches.

- Microsoft recommends iWARP for S2D
- Microsoft Recommendation on the RDMA alternatives in Windows
- Hyper-converged solution using Storage Spaces Direct in Windows Server 2016

Storage Spaces Direct Performance with iWARP

This paper demonstrates S2D performance on a 3 node AMD EPYC cluster using Chelsio 25GbE iWARP T6225-CR/T6225-SO-CR adapters.

Results

Below is a screenshot from the VMFleet Watch-Cluster window, which reports IOPS, bandwidth and latency.

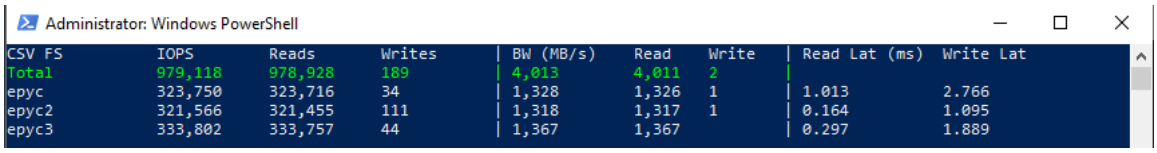

Figure 2 – Storage Spaces Direct IOPS Numbers

As seen from the above screenshot, this setup was able to demonstrate over **978K IOPs** for 4KiB random read in aggregate using 60 virtual machines, which means **16,300+ IOPs per virtual machine**.

The following graph presents the READ, WRITE, READ-WRITE (70/30) IOPs and BW results for I/O sizes varying from 4 to 512 KBytes. *Please note that BW numbers for higher IO-sizes are limited by SSD performance.*

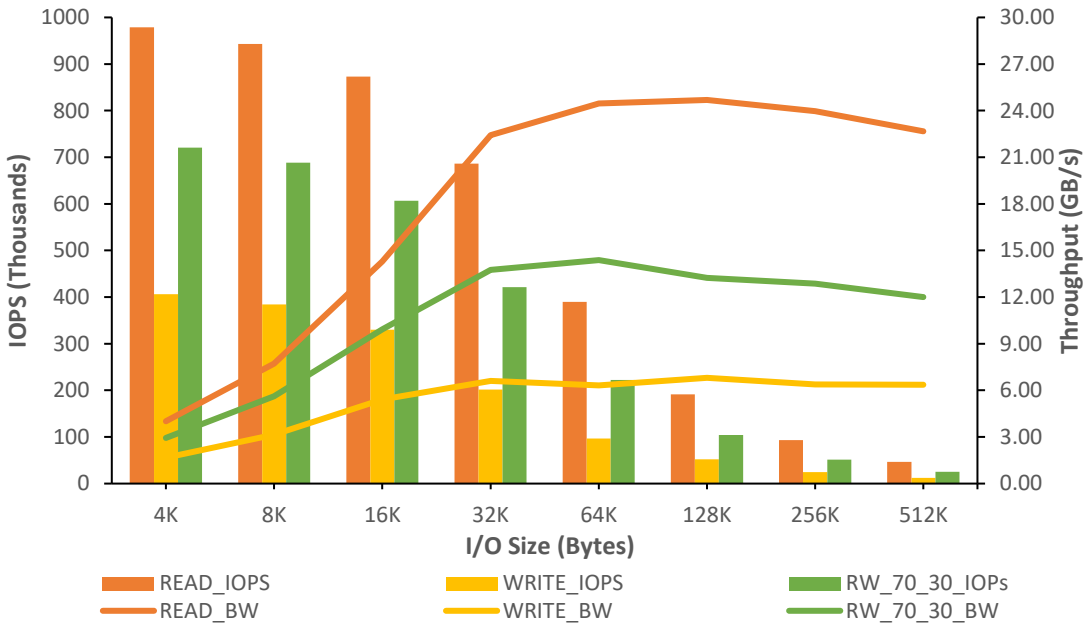

Figure 3 – Storage Spaces Direct: IOPS and BW vs. I/O size

The following is the hardware configuration:

- **3 nodes of AMD EPYC**
  - 1x AMD EPYC 7551 32-Core Processor @ 2.00GHz
  - 128GiB DDR4 2667MHz
  - Windows Server 2019 (build:17763.195) + Storage Spaces Direct
  - Motherboard: Supermicro H11DSi-NT 1.01
  - Performance Power Plan
  - 1x Chelsio T6225-CR/T6225-SO-CR (Dual port 25Gb PCIe 3.0 x8) per node
    - Driver: v6.13.1.0
    - Firmware: v1.22.8.0
    - Dual port connected to 25G Switch
    - MTU 1500
- **Total Storage:**
  - 4x 1.6TB Intel NVMe SSD DC P3600 (PCIe 3.0 x8)
  - 4x 2.4TB Micron NVMe SSD 9100 (PCIe 3.0 x4)

A 20 virtual machines per node, for a total of 60 virtual machines configuration was used for this setup. Used A1 default virtual machine size (i.e., 1 vCPU and 1.75GiB RAM). VMFleet tool was used to run DISKSPD in each of the virtual machines with various multi-threaded workloads for IO-sizes 4KiB to 512KiB, with 24 threads and queue depth of 4 per VM.

In addition to S2D, iWARP enabled Chelsio adapters power other aspects of Microsoft Windows installations such as **Storage Replica** for disaster recovery, **SMB Direct** for high performance file access, **Client RDMA** for bringing RDMA benefits to Windows 10 deployments, **Network Direct** for Windows HPC deployments, hardware offloaded **iSCSI, iSER, NVMe-oF and FCoE initiators** for SAN applications, **d.VMMQ, NVGRE and VxLAN** encapsulation offload, **Guest RDMA, SR-IOV** for Virtual environments.

Summary

AMD EPYC cluster with Chelsio iWARP RDMA enabled 25Gb Unified Wire and Converged Network adapters (CNAs) delivers a high-performance Storage Spaces Direct (S2D) solution using standard Ethernet infrastructure and enables datacenters to deploy S2D now by leveraging all-inboxed drivers. The ability to work with any non-DCBX switch, enables an immediate plug and play deployment. Support of iWARP protocol is enabled since Windows Server 2012-R2 release, and in boxed support in Windows Server 2016 and 2019, has allowed for years of testing for a very robust, tested, and efficient deployment with Chelsio iWARP enabled Ethernet adapters.

Related Links

Axellio demos WSSD cluster with Chelsio 100GbE
XIO demos WSSD cluster at Microsoft Ignite
Storage Spaces Direct throughput with 100GbE iWARP - Microsoft Blog
High Performance S2D with Chelsio 100GbE
Storage IOPS Update with S2D - Microsoft Blog
Windows Server 2016 Storage Spaces Direct