High Performance S2D with Chelsio 100GbE

Chelsio T6 iWARP RDMA solution for Windows Storage Spaces Direct

Overview

Microsoft **Storage Spaces Direct** (S2D) is a feature introduced in Windows Server 2016, 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. Storage Spaces Direct leverages SMB3 for all intra-node communication, including SMB Direct and SMB Multichannel, for low latency and high throughput storage. This guide presents S2D performance results using Chelsio iWARP RDMA technology in a hyper-**converged** deployment scenario.
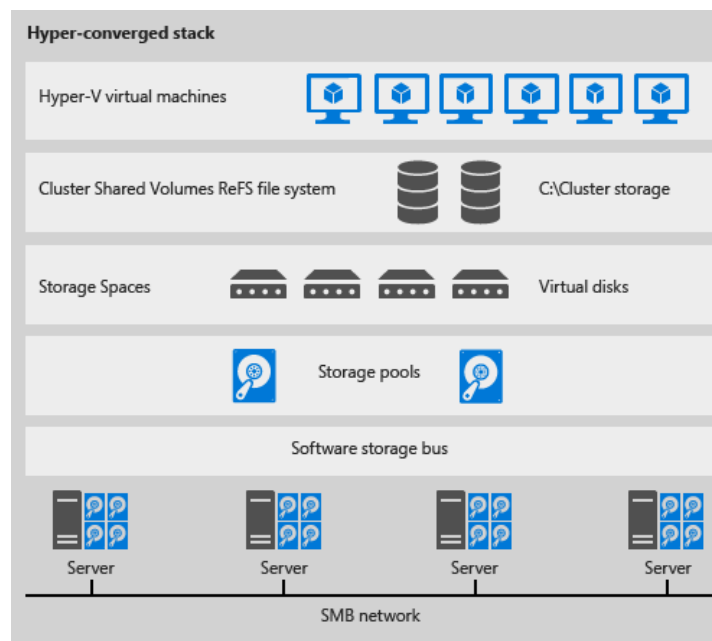

Figure 1 – S2D Hyper-Converged Stack

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 9 years, TCP/IP has been an IETF standard (RFC 793, 791) for 35 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.
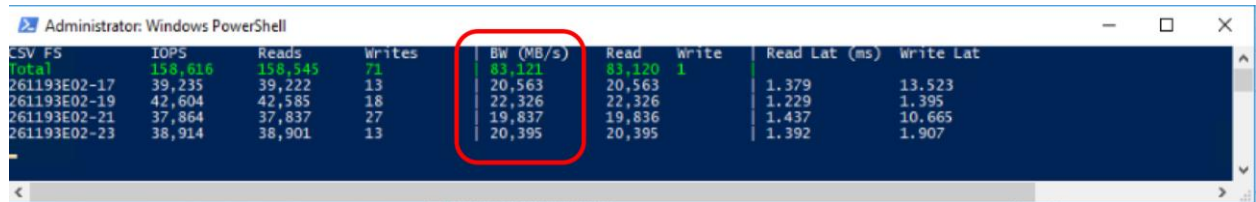
Storage Spaces Direct Performance with iWARP

At Microsoft Ignite 2016, Microsoft did a demo at the 28-minute mark as part of the Meet Windows Server 2016 and System Center 2016 session, demonstrating how Storage Spaces Direct can deliver massive amounts of IOPS to many virtual machines with various storage QoS settings. It showed 6M+ read IOPS using Chelsio 40GbE iWARP adapters.

This paper explains a 4 node S2D cluster connected using the Chelsio 100GbE iWARP T62100-CR adapters and demonstrating aggregated bandwidth exceeding 83GB/s, each VM utilizing around 1GB/s of throughput with average read latency <1.5ms. The following is the hardware configuration:

- **4 nodes of Dell R730xd**
    - 2x E5-2660v3 2.6Ghz 10c/20t
    - 256GiB DDR4 2133Mhz (16 16GiB DIMM)
    - 2x Chelsio T6 100Gb NIC (PCIe 3.0 x16), single port connected/each, QSFP28 passive copper cabling
    - Performance Power Plan
    - Storage:
        - 4x 3.2TB NVME Samsung PM1725 (PCIe 3.0 x8)
        - 4x SSD + 12x HDD (not in use: all load from Samsung PM1725)

- **Windows Server 2016 + Storage Spaces Direct**
    - Cache: Samsung PM1725
    - Capacity: SSD + HDD (not in use: all load from cache)
    - 4x 2TB 3-way mirrored virtual disks, one per cluster node
    - 20 Azure A1-sized VMs (1 VCPU, 1.75GiB RAM) per node
    - OS High Performance Power Plan

- **Workload details:**
    - DISKSPD workload generator
    - VM Fleet workload orchestrator
    - 80 virtual machines with 16GiB file in VHDX
    - 512KiB 100% random read at a queue depth of 3 per VM

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 you can see, the aggregated bandwidth exceeded 83GB/s, which is very impressive. Each VM realized more than 1GB/s of throughput, and notice the average read latency is <1.5ms.

In addition to Storage Spaces Direct, iWARP Protocol also powers 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 and hardware offloaded **iSCSI initiator** for SAN applications.

Summary

Chelsio iWARP RDMA enabled 100Gb Ethernet adapter 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 with Chelsio Ethernet adapters. 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, has allowed for years of testing for a very robust, tested, and efficient deployment with Chelsio iWARP enabled Ethernet adapters.

Related Links

[Storage Spaces Direct throughput with 100GbE iWARP - Microsoft Blog](#)
[S2D throughput using 100GbE iWARP – Microsoft Tweet](#)
[Storage IOPS Update with S2D - Microsoft Blog](#)
[Windows Server 2016 Storage Spaces Direct](#)
[Configuring Storage Spaces Direct](#)
[Hyper-Converged Scale-Unit with Chelsio 40GbE](#)