S2D Performance with Network QoS

Chelsio 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 networked 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.

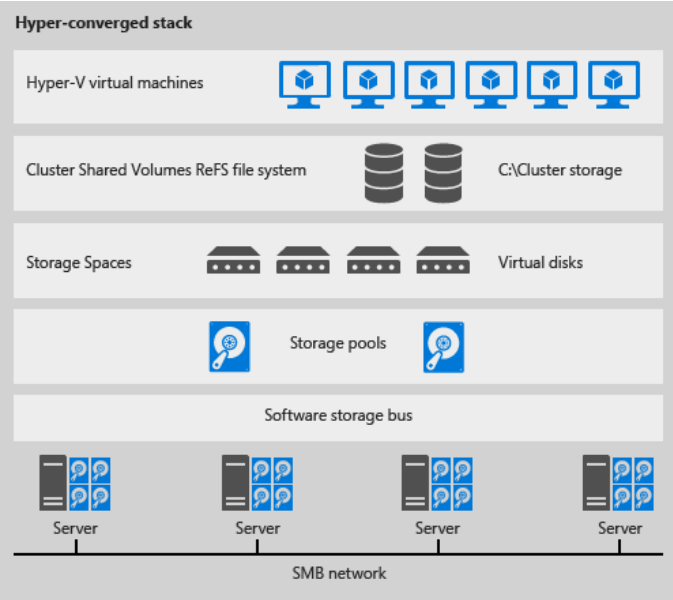

Figure 1 – S2D Hyper-Converged Stack

Network QoS is used in this **hyper-converged** configuration to ensure that the Software Defined Storage system has enough bandwidth to communicate between the nodes to ensure resiliency and performance. This paper presents Network (NIC) and iWARP RDMA bandwidth results, with and without Network QoS enabled in an S2D environment. The results showcase Chelsio adapter's excellent rate-limiting capabilities with a converged network traffic. The results also show how bandwidth allocation can be easily offloaded onto Chelsio Unified Wire adapters, bypassing the operating system and expensive switch hardware (DCB capable), thus resulting in reduced total ROI and simplified management.

Test Results

The following results present Network (NIC) and iWARP RDMA throughput performance numbers collected with and without network QoS enabled. NTttcp benchmarking tool is used to collect NIC results and VMFleet tool is used to run Diskspd to collect iWARP RDMA numbers.

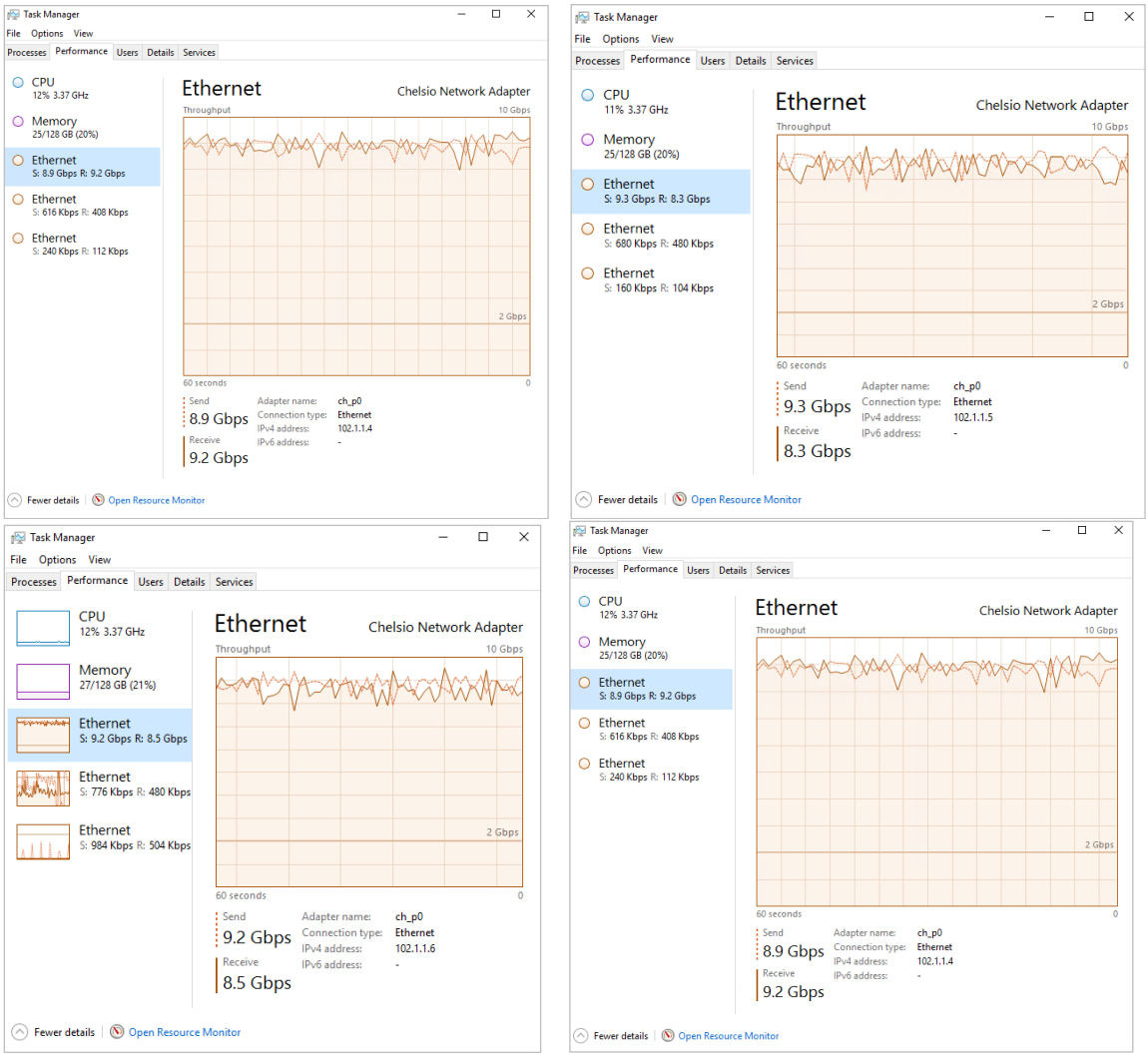

Figure 2 - NIC Only Throughput Results without network QoS

Figure 2 shows NIC only bandwidth results from each host, reaching line-rate.

	IOPS	Reads	Writes	BW (MB/s)	Read	Write	Read Lat (ms)	Write Lat
Total	7,728	1	7,727	2,008		2,008		
CHANDRU	1,906		1,906	492		492	0.000	21.062
BUDHA	1,918	1	1,917	498		498	2.697	20.800
MANGAL	1,982		1,982	516		516	0.000	20.127
SURYA	1,923		1,923	501		501	0.000	20.767

Figure 3 – iWARP RDMA Only Throughput results without network QoS

Figure 3 shows iWARP RDMA only bandwidth results.

	IOPS	Reads	Writes	BW (MB/s)	Read	Write	Read Lat (ms)	Write Lat
Total	10,178	4,635	5,543	1,447	4	1,443		
CHANDRU	5,927	4,545	1,382	364	4	360	0.001	28.922
SUDHA	1,419	45	1,374	358		358	0.001	28.993
MANGAL	1,473	45	1,428	371		371	0.001	28.064
SURYA	1,359		1,359	354		354	0.000	29.399

Figure 4 – iWARP RDMA and NIC Throughput results without network QoS

Figure 4 shows, iWARP RDMA consuming 75% of bandwidth when both iWARP RDMA and NIC traffic are run concurrently with no network QoS configured.

	IOPS	Reads	Writes	BW (MB/s)	Read	Write	Read Lat (ms)	Write Lat
Total	6,004		6,004	1,563		1,563		
CHANDRU	1,497		1,497	389		389	0.000	26.673
SUDHA	1,506		1,506	392		392	0.000	26.507
MANGAL	1,532		1,532	400		400	0.000	26.056
SURYA	1,469		1,469	382		382	0.000	27.110

Figure 5 – iWARP RDMA Throughput with network QoS set 80:20

Figure 5 shows iWARP RDMA bandwidth with Network QoS configured as 80:20 (RDMA: NIC).

	IOPS	Reads	Writes	BW (MB/s)	Read	Write	Read Lat (ms)	Write Lat
Total	819		819	210		210		
CHANDRU	211		211	54		54	0.000	112.730
SUDHA	187		187	48		48	0.000	113.721
MANGAL	206		206	53		53	0.000	113.998
SURYA	215		215	55		55	0.000	96.011

Figure 6 – iWARP RDMA Throughput with network QoS set 10:90

Figure 6 shows iWARP RDMA bandwidth with Network QoS configured as 10:90 (RDMA: NIC).

Observing the bandwidth numbers, it is clear how Chelsio's Unified Wire adapter was utilized to allocate bandwidth efficiently to different protocols without any loss in performance.

Test Setup

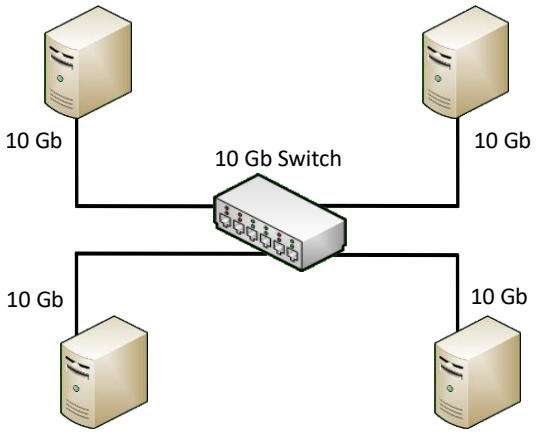

Figure 7 - Topology

Network Configuration

The test setup consists of 4 Ivy Bridge Supermicro servers (nodes) connected via a 10GbE switch using single port on each system. Each node is configured with 2 Intel Xeon E5-2678v2 8-core processors clocked at 3.4Ghz (HT enabled) and with 64GB of RAM. T520-CR adapter is installed

in each node with Windows Server 2016 Data Center edition. Standard MTU of 1500B is used. 20 virtual machines are created per node, 80 in total. Each VM is configured with 1 vCPU and 1GB of RAM with Windows Server 2016 as guest operating system.

No additional configuration is needed on the Switch when configuring S2D with iWARP RDMA. Network QoS can be configured on the hosts to rate-limit the RDMA and NIC bandwidth accordingly. In addition, the QoS limit can be configured/changed dynamically without changing switch configuration.

Commands Used

Setting QoS limit

i. Enable QoS:

```
# Enable-NetAdapterQos ch_p0
```

ii. Create new policy:

```
# New-NetQosPolicy -Name rdma -PriorityValue 5 -NetDirectPortMatchCondition
445
```

iii. Create new ETS rule specifying the bandwidth percentage:

```
# New-NetQosTrafficClass -Name rdma -Priority 5 -Algorithm ETS -
BandwidthPercentage 80 -Verbose
```

To change limit of an existing ETS rule:

```
# Set-NetQosTrafficClass -Name rdma -Priority 5 -Algorithm ETS -
BandwidthPercentage 10 -Verbose
```

NIC:

Server

```
# ntttcp -r -m 2,*,test-ip -l 128k -rb 512k -p 10000 -t 300
```

Client

```
# ntttcp -s -m 2,*,test-ip -l 128k -sb 512k -p 20000 -t 300
```

RDMA:

```
# diskspd.exe -Z20M -z -h -t1 -o2 -b256k -p256k -w100 -W30 -C30 -d300 -D -L
C:\run\testfile1.dat
```

Network QoS with SET and RDMA vNICs

Switch Embedded Teaming (SET) can group iWARP RDMA enabled Chelsio ports into one or more software-based virtual network adapters (vNICs). RDMA can be enabled on the vNICs to provide fast performance and fault tolerance in the event of a network adapter failure. The following commands configure Network QoS on a host with SET and RDMA vNICs configured:

i. Enable QoS:

```
# Enable-NetAdapterQos
```

ii. Create a vSwitch with SET:

```
# New-VMSwitch -Name "sw0" -NetAdapterName ch_p0,ch_p1 -AllowManagementOS
$false -EnableEmbeddedTeaming $true
# Set-VMSwitchTeam -Name "sw0" -LoadBalancingAlgorithm Dynamic -TeamingMode
SwitchIndependent
```

iii. Create the required number of vNICs (4 for RDMA traffic and 1 for Management traffic):

```
# Add-VMNetworkAdapter -SwitchName "sw0" -Name "rdmanic1" -ManagementOS
# Add-VMNetworkAdapter -SwitchName "sw0" -Name "rdmanic2" -ManagementOS
# Add-VMNetworkAdapter -SwitchName "sw0" -Name "rdmanic3" -ManagementOS
# Add-VMNetworkAdapter -SwitchName "sw0" -Name "rdmanic4" -ManagementOS
# Add-VMNetworkAdapter -SwitchName "sw0" -Name "vnic1" -ManagementOS
```

iv. Enable RDMA for vNICs:

```
# Get-NetAdapter | where {$_.Name -like "*rdmanic*"} | Enable-NetAdapterRdma
```

v. Create new policy for RDMA:

```
# New-NetQosPolicy -Name rdma -PriorityValue 5 -NetDirectPortMatchCondition
445
```

vi. Create new ETS rule specifying the bandwidth percentage:

```
# New-NetQosTrafficClass -Name rdma -Priority 5 -Algorithm ETS -
BandwidthPercentage 80 -Verbose
```

Conclusion

This paper presented the rate-limiting capabilities of Chelsio Unified Wire Adapters in an S2D environment. Chelsio's Unified Wire adapters with enhanced rate-limiting features offer data centers the benefits of creating a unified network with low operational cost, consolidated storage, server and networking resources, reduced heat and noise, and less power consumption. A lossless Ethernet is guaranteed with simplified administration since the specifications enable transport of storage and networking traffic over a single unified Ethernet network.

Related Links

S2D Performance with iWARP RDMA
iWARP RDMA – Best Fit for Storage Spaces Direct
High Performance S2D with Chelsio 100GbE
S2D throughput with 100GbE iWARP - Microsoft Blog
Hyper-Converged Scale-Unit with Chelsio 40GbE
High Performance S2D with Chelsio 40GbE
Windows Server 2016 Storage Spaces Direct