Micron® 5100 Is Ready for Next Generation of Microsoft® Storage Spaces Direct

Using HPE® Proliant servers with Micron 5100 series SATA SSDs provides top-tier, cost-effective HCI solutions for your Windows®-based data center

Overview

One of the interesting technologies that is part of the Windows Server architecture is hyper-converged software-defined storage. As a major provider of storage technologies such as DRAM and SSDs, Micron wanted to get an early peek at the future of Storage Spaces Direct (S2D) to see how Micron SSDs could contribute to a high-performance S2D solution. With the pending release of RedStone 3 (RS3), the development builds of RS3 offered us a preview of what will be coming with the future release of S2D. While we don't know when the next version of S2D will be released by Microsoft, what they have done so far in the early access leading up to RS3 is extremely interesting.

KEY FINDINGS

- 625,000 4 KiB mixed random IOPS in a 4-node S2D cluster (12.7 cents per IOPS[1]
- 9 GiB per second 100% sequential read throughput
- 2.1 GiB per second 100% sequential write throughput
- All-Flash S2D will meet the most demanding enterprise workloads

Background

Our goal was to better understand the future of Microsoft S2D technology and how Micron's storage products can help you meet your HCI performance needs. To accomplish that, Micron has been working in cooperation with HP Enterprise® (HPE), running a series of tests built upon HPE ProLiant DL Gen 9 Series servers and using pre-release versions of Microsoft's upcoming Windows Server platform. The test results documented in this technical brief provide early information you can use to develop your plans to successfully build high-performance, all-flash S2D clusters which can be used as building blocks for larger-scale HCI solutions. Once Microsoft releases the final version of S2D to the public, Micron will be starting a series of projects that will dig deeper into the Microsoft S2D software and highlight the benefits that Micron's solid state technology can provide for your solution. We plan to provide results of this testing in the form of a formal reference architecture that can be used to scale your Hyper-V and S2D infrastructure.

Testing

Our early test results with the pre-release version of Hyper-V, S2D and Micron SSDs show impressive performance for an all-SATA, all-flash storage solution.

For this technical brief, we ran a series of tests using build 16278 of Windows 2016 Core configured as a 4-node Hyper-V HCI cluster with eight virtual machines per node, using the cluster configuration and a single-capacity tier of 8 960GB Micron 5100 MAX SATA SSDs per node (See Figure 5 later in this brief).

Testing was performed using various read-write ratios at various queue depths using small (4 KiB) and large (64 KiB) block sizes in random and sequential patterns. Typical random I/O workloads are 4 KiB in size to simulate transactional database solutions such as Microsoft SQL Server®; typical sequential I/O uses larger block sizes (64 KiB) to simulate workloads such as video streaming and transcoding, database transactions logs or data warehouse ingest operations. (See How We Test.)

Impressive All-Flash HCI Performance for OLTP

Using randomized 4 KiB I/O requests to represent small-block solutions, our test results show that SATA all-flash solutions can be very effective when used within hyper-converged S2D environments (Figure 1). Typical OLTP transactional use cases experience mixed read-write access patterns with small block sizes and are represented by the 10% and 30% write results shown in Figure 1. The test shows impressive results even at low queue depths, but the solution scales very well as we increase queue depths from 1 to 32, with the result for a 70/30 mixed workload peaking at over 625,000 IOPS from 32 SATA SSDs and 4 S2D nodes, as measured at the virtual server layer. Since we are using triple-mirrored volumes, that 625,000 IOPS translates to over 1 million IOPS, as measured at the drive interface[2] based on the 70/30 read/write mix.

When we drill into the 70/30 read/write results, we can see that this excellent IOPS performance comes with very low latencies (Figure 2). A couple of observations can be made by looking at the results. First, there was a significant drop in performance at queue depth 4 for some unknown reason. We will need to investigate this further, but it could simply be an anomaly of using a pre-release versions of S2D. Second, we saw a clear plateau in performance moving from queue depths of 16 to 32. Considering the 100% increase in the latency values between those two queue depths, it would appear that —

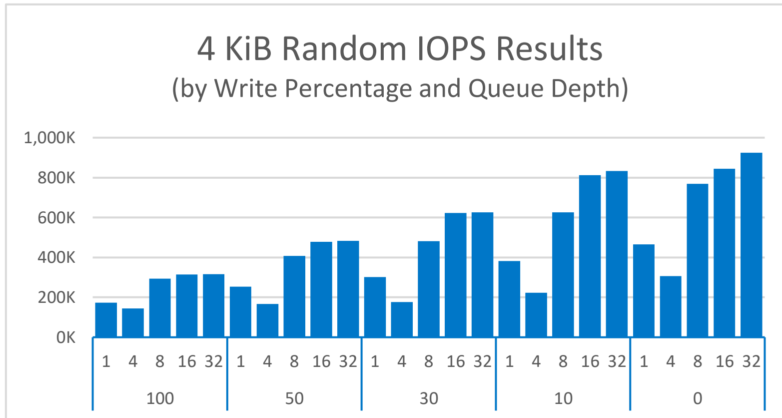

Figure 1: Micron SATA All-Flash S2D 4 KiB Random IO Results

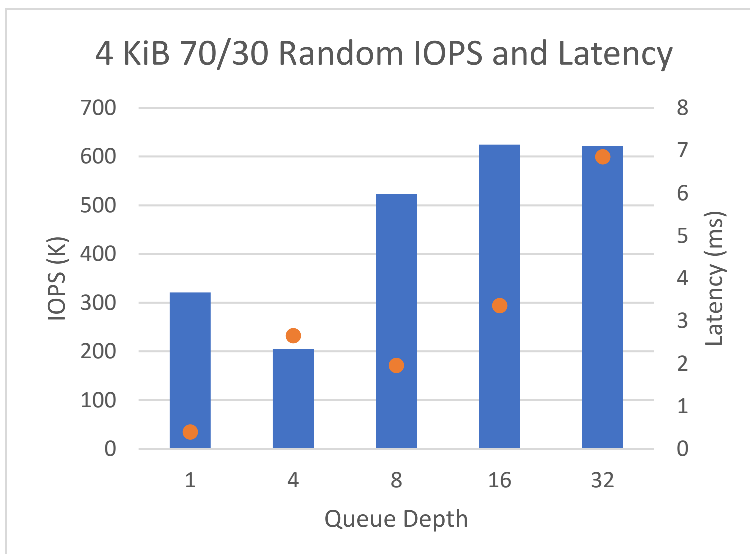

Figure 2: Low-Latency Random I/O for Mixed Read/Write Workloads

at least with this pre-release of the software — that the sweet spot would be operating at queue depths of 16. More investigation is warranted with the final release of S2D, but it is a good starting point for your planning. Again, once we receive the release version of S2D, we will be able to do a more comprehensive analysis.

Maximized Throughput for Your Data

Large block read and write I/O is representative of workloads such as business intelligence and video streaming/transcoding, and we wanted to look at the maximum throughput we could achieve with this solution for both reads and writes. To measure throughput, we ran 100% sequential read and write workloads using 64 KiB block size across multiple queue depths.

Our 100% read test results show that S2D with Micron 5100 SSDs can extract over 8 GiB per second of throughput at higher queue depths from our SATA all-flash test environment (Figure 3).

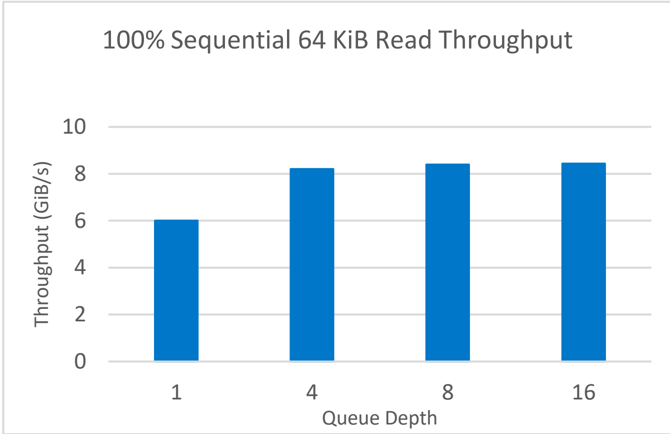

Figure 3: 100% Read Throughput Results

For 100% write workloads, our test configuration achieved between 1.1 and 2.1 GiB per second of write throughput depending on queue depth (Figure 4). Of note is that throughput decreased at higher queue depths. This degradation in throughput demands further investigation with the final S2D release, but it is important to remember additional I/O is required due to triple mirroring. Triple mirroring adds additional latency, and each of the 2.1 GiB being sent to three different SSDs thus represents 6.3 GiB of drive level I/O.

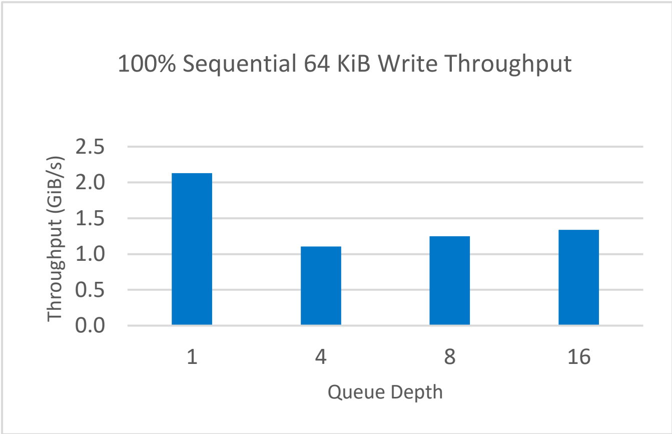

Figure 4: 100% Write Throughput Results

More investigation is necessary, and we plan on doing a comprehensive suite of virtual machines definitions as part of our reference architecture project once we have the final version of S2D.

Testing Configuration

The test cluster (Figure 5) consisted of four HPE DL380 G9 servers configured as a hyper-converged cluster. Each DL380 G9 was equipped with eight Micron 5100 MAX 1.92TB SATA SSDs. All SSDs were added to a single S2D storage pool, and triple-mirrored logical volumes were configured using Storage Spaces. Eight virtual machines were deployed to each node in the cluster.

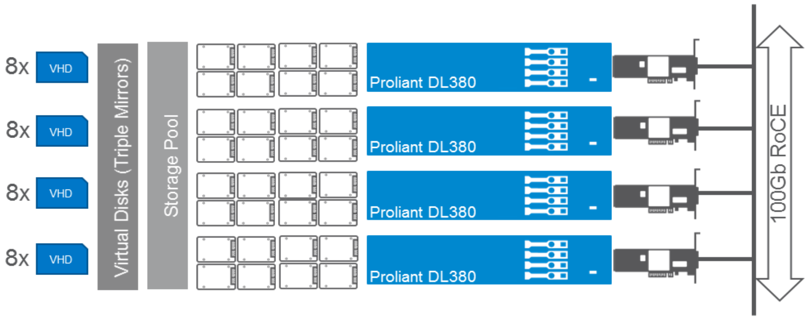

Figure 5: Single-Tier Hyper-V With S2D Test Configuration

Each node in the cluster was interconnected with a 100 GbE network configured to use RDMA over Converged Ethernet (RoCE) for all data movement within the S2D environment. This combination delivered excellent performance in a flexible and cost-effective solution for hyper-converged deployments.

The following tables describe the test environment in detail.

Server Configuration (Hewlett-Packard Enterprises DL380 Gen 9

Component	Description	Quantity
Server platform	Hewlett-Packard Enterprises DL380 Gen 9	4
Processor	Intel E5-2690v4 (817959-B21 + 817959-L21)	1+1
Memory	32 Micron DDR4 RDIMMs	8
HBA Controller	HP H240ar FIO Smart HBA	1
Boot device	HP Dual 120GB RI Solid State M.2 (777894-B21)	2
Capacity tier	Micron 5100 MAX 960GB SSDs	8
Network	HPE IB EDR/EN 100GB Adapter	1

Total cost for one DL380 test server was $19,800[3] based on approximate pricing from an HPE reseller rounded to the nearest dollar. This cost was used to determine the value of the solution relative to its performance. The network switch was not included in this value estimate with the assumption that most readers would already have a network infrastructure in place. The typical cost for a 32-port Mellanox Spectrum SN2700 100 GbE switch is $14,000–$20,000[3] depending on the reseller.

VM Definition

Component	Description
VM Size (Total/Active)	40GB VHDX (10GB Active Data Set)
vProcessor	8 cores
vRAM	16GB

Software Definition

Component	Description
Microsoft Windows Core	Fall Update Pre-Release (Build 16278)
DiskSPD/VMFleet	Version 0.8

DiskSPD Switches Used During Testing

Parameter	Small-Block Random Test	Large-Block Sequential Tests
Block Size	-b 4	-b 64,128
IO Pattern	-p r	-p si
Write Percentage	-w 0,10,30,50,100	-w 0,100
Duration	-d 120	
Warmup	-W 30	
Cool Down	-C 30	
Other Parameters	-Z20M, -z, -h, -D, -L, -R xml	

The Bottom Line

Our pre-release demonstration across a 4-node cluster using Storage Spaces Direct, Hyper-V, Windows Clustering, local high-performance SATA SSD storage and 100 GbE configured with RDMA over Converged Ethernet (RoCE), shows the potential value of eliminating complex and costly external shared storage requirements from Windows server solutions. The demonstration used DiskSpd to generate nearly 20,000 random mixed workload I/Os from each of the 32 VMs in the 4-node cluster.

Overall performance was excellent, with 70/30 mixed small-block I/O, exceeding our expectations based on our past experience with various virtual server solutions generating over 625,000 IOPS. In addition, our 100% sequential read/write workloads showed that S2D can drive many enterprise data center workloads with over 9 GiB of cluster throughput.

Several observations from our testing could improve your Fall Update S2D planning:

- Overall S2D performance can be greatly impacted by dataset size relative to free memory. Typically, Windows will leverage as much free memory as possible as a cache. Smaller data sets will result in more I/O being executed out of cache rather than disks.
- Write performance is directly impacted by the data protection scheme selected. While we ran 3-way mirroring on our 4-node cluster, Microsoft recommends using erasure coding with any cluster larger than 3 nodes. This should improve overall performance for write heavy workloads relative to our 3-way mirror configuration.
- There seems to be a sweet spot balance between performance and latency in our mixed workload testing around queue depth 16 at a VM level which represents a kernel-level queue depth of 1024.

With our 5100 SSD family, you get the consistent performance, reliability and low TCO you've come to expect from Micron SSDs. The 5100 MAX drives provide the best endurance of the 5100 family specifically required for high write I/O environments, along with the read performance you would expect from solid state technologies.

Visit our web site for more information about the Micron 5100 family.

How We Test

Micron believes in providing useful, real-world test results based on solution environment configuration values that our customers would typically use for their solutions. It is our goal to provide as much data as possible to provide you with enough information to customize your solution to meet your performance goals.

In addition, we always execute multiple test runs for every configuration change and publish the statistical average of the results across all test runs.

For this technical brief, we ran each test configuration five times.

micron.com

1. Price per IOPS calculated using total cost detailed in Test Configuration section for 4 server nodes divided by 625,000 IOPS.
2. Drive IOPS = (0.3 x 625,000)*3 + (0.7 x 625,000).
3. Pricing as of September, 2017.