# dataON

# Deploying a Windows Server 2016 Software-Defined Data Center (SDDC) Solution with Storage Spaces Direct and Storage Replica

DataON and Strategic Online Systems, Inc.
With Intel and Microsoft

August 2017

# Contents

# Abstract

Microsoft released Storage Spaces with Windows Server 2012 as the foundation of a Windows software-defined storage (SDS) stack. DataON is a storage industry veteran and pioneer of cluster-aware storage, delivering certified Windows-based scale-out server (SOFS) platforms and Cluster-in-a-Box (CiB) appliances to over five hundred small and enterprise customers.

As the industry transitions to software-defined data centers (SDDC) with Window Server 2016 and infrastructures built on SDS, hyper-converged and Azure stack hybrid cloud, DataON has emerged as the industry-leading provider of hyper-converged cluster appliances (HCCAs) and storage systems optimized for Microsoft® Windows Server environments. DataON has exclusively focused on customers who have made the "Microsoft choice" to deploy Windows Sever-based storage solutions.

Working with technology industry leaders and ecosystem partners, DataON has delivered more than one hundred petabytes of hyper-converged platform data based on Window Server 2016 Storage Spaces Direct. DataON partners with Intel to power our hyper-converged cluster appliances based on the latest generation of Intel CPUs, Intel Data Center SSDs and Intel Data Center Blocks for Cloud and Microsoft Windows Server 2016. With software-defined networking (SDN) being an integral part of the Microsoft SDDC hyper-converged platform, DataON also partners with Mellanox to fully leverage the technical foundation of SMB3 over RDMA through the switch and network adapter fabric.

Youth Villages is a private non-profit dedicated to helping emotionally and behaviorally troubled children and their families live successfully. Youth Villages partners with Strategic Online Systems, Inc. (SOS) as their managed service provider and they chose DataON to deploy a Windows Server 2016 Storage Spaces Direct solution.  After setup and migration, SOS and Youth Villages saw a significant performance improvement in their new virtualized environment.

In this paper, DataON and Strategic Online Systems, Inc. share their experiences from deploying a DataON S2D-3224i hyper-converged platform with Windows Server 2016 Storage Spaces Direct at Youth Villages.

# Customer Background



## Overview

Youth Villages is a private non-profit dedicated to helping emotionally and behaviorally troubled children and their families live successfully. Helping more than 22,000 children and families every year, Youth Villages is headquartered in Memphis, TN, with more than 2,700 employees across 72 locations in the United States.

Its programs include intensive in-home services, foster care and adoption, residential and intensive residential treatment, an emergency shelter for homeless and runaway teens, and specialized crisis services. Youth Villages' YVLifeSet program has been proven to be one of the country's most effective ways to help former foster youth find success as independent adults. The nonprofit's unique approach produces lasting success for children, with more than 80 percent still living successfully in their homes and communities two years after completing their programs.

Youth Villages partners with Strategic Online Systems, Inc. (SOS) as their managed service provider. SOS focuses on providing strategic planning, network support and equipment deployments for non-profit organizations. Like DataON, SOS understands that hardware is simply a commodity item. The true value lies in a well-integrated solution, which provides an improved user experience to increase customer efficiency.

**IT Challenge: Upgrade Youth Villages' storage environment to a new software-defined solution with an eye for the future.**

SOS used a HP LeftHand iSCSI SAN with Super Micro servers and Cisco switches for Youth Villages' storage infrastructure. Their traditional SAN was rapidly aging, achieving only 5-10K IOPS. System performance was adversely affected during periods of peak demand on the servers.

SOS's goals for an updated Youth Villages' IT infrastructure, included:

- Support for NVMe SSDs for caching
- Support for SQL (EMR database), Hyper-V, Exchange, VDI (9x RDS), and other enterprise applications
- Achieving at least 150K IOPS, to support 2800 users
- Providing 40TB usable storage, growing to 100TB in three years
- Support for RDMA networking
- Replacing the current management and monitoring software

Another consideration for SOS was finding a solution that would still be viable in three years. Most of the leading hyper-converged solution providers have only been in business for a few years. With a rapidly consolidating market, it would be difficult to project which companies would have staying power. Choosing a Microsoft solution could allay that fear.

SOS had successfully deployed Windows Server 2012 R2 Storage Spaces in the past. They learned about the upcoming launch of Windows Server 2016 and were excited about its Storage Spaces Direct solution.

In the end, SOS decided to deploy a Windows Server 2016 Storage Spaces Direct solution with DataON storage for Youth Villages.

# Youth Villages Customer Story At-a-Glance

| | |
|---|---|
| **Customer** | **Youth Villages** (http://www.youthvillages.org/)<br><br>Youth Villages is a private non-profit organization with a mission to help children and family live successfully. |
| **Solution Integrator** | **Strategic Online Systems, Inc.** (http://www.thesosinc.com/)<br><br>SOS is a managed service provider, providing strategic planning, network support and equipment rollout. They specialize in helping medium to large organizations that are making a difference in their local communities, by assisting with:<br>• Management of laptops, workstations, servers and networking equipment<br>• Windows 10 migrations<br>• Service hosting<br>• Email archiving / compliance<br>• Backups / disaster recovery / business continuity<br>• Bring your own device implementations<br>• Enterprise database optimization |
| **Products and Services** | **Hardware**:<br>DataON S2D-3224i (optimized for performance & density)<br><br>**Workloads**:<br>• Windows Server 2016<br>• Storage Spaces Direct<br>• SQL Server<br>• Hyper-V<br>• VDI<br>• Microsoft enterprise applications |
| **Industry** | Nonprofit Organization |
| **Organization Size** | 1,001 to 5,000 employees |
| **Country** | United States |

# Windows Server 2016 Applications

## Overview

Windows Server 2016 is a cloud-ready operating system that delivers new layers of security and Microsoft Azure-inspired innovation for the applications and infrastructure that power your business. From a storage perspective, Windows Server 2016 includes new features and enhancements for software-defined storage, as well as for traditional file servers.

Storage Spaces Direct enables building highly available and scalable storage using servers with local storage, such as DataON S2D hyper-converged appliances. It simplifies the deployment and management of software-defined storage systems and unlocks use of new classes of disk devices, such as SATA SSD and NVMe disk devices.

Health Service improves the day-to-day monitoring, operations, and maintenance experience of cluster resources on a Storage Spaces Direct cluster.

Storage Replica enables storage-agnostic, block-level synchronous replication between servers or clusters for disaster recovery, as well as stretching of a failover cluster between sites.

## Storage Spaces Direct

### Overview

Storage Spaces Direct is a new feature in Windows Server 2016. It uses servers with locally-attached drives (such as DataON S2D hyper-converged appliances) to create highly available, highly scalable software-defined storage at a fraction of the cost of traditional SAN or NAS arrays.
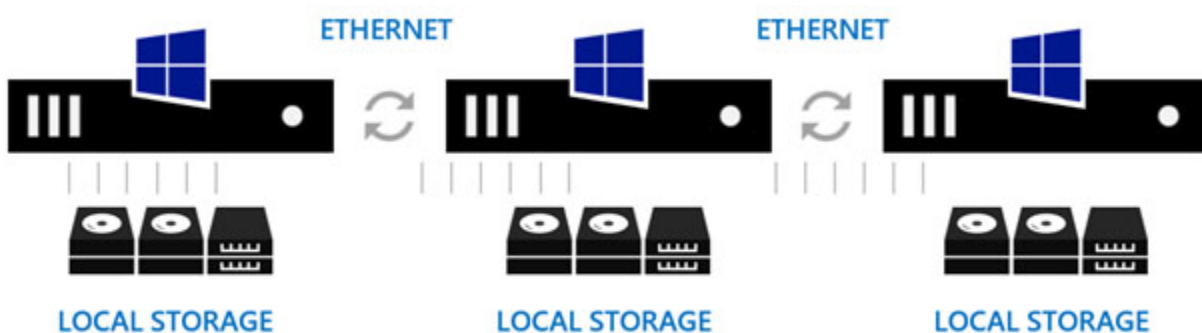
- Storage Spaces utilizes NVMe storage devices for better performance and efficiency. These devices greatly reduce the I/O latency for storage, as well as reduce the CPU utilization to server storage.
- You can also use SATA storage devices to lower the cost of storage.
- In a DataON hyper-converged system, storage uses RDMA over Converged Ethernet (RoCE) as the storage fabric instead of having a shared physical storage fabric behind the servers.
- Storage Spaces Direct is included in Windows Server 2016 Datacenter edition.

## Hardware

The most common configuration is 2-tiered physical storage, combining SSDs with traditional hard drives. You can also have an all-flash configuration with NVMe SSDs plus traditional SSDs where the NVMe SSDs are used for cache and the SATA SSDs are used for capacity.

In a 3-tiered physical storage configuration, NVMe SSDs are used for caching and both SATA SSDs and HDDs are used for additional tiering in the system. The SATA HDDs can be used to store the coldest data.

A traditional 10GbE or better Ethernet network can be used to connect servers/storage via RDMA. RDMA provides significant advantages because it lowers the latency of the storage I/O in the system and reduces the CPU utilization, resulting in higher IOPS performance in the system.

## Feature Highlights

**Built-in Always-on Cache** – Storage Spaces Direct takes the fastest devices in the system and uses them as a write cache so the applications can continue immediately after writing data. It also acts as a read cache for the most frequently read data from slower devices, for faster workloads. The cache automatically configures itself when you enable Storage Spaces Direct.

**Single Pool of Storage** – Storage Spaces Direct automatically groups the available storage devices into a single storage pool, eliminating the need for manual configuration or setting up multiple storage pools.

**Scale from 2 to 16 nodes** – Typically deployments consist of 2, 4, 8, 12 or 16 nodes.

**Scales to 400+ drives** – These servers can accommodate 26 drives per server, for a total of 416 drives. This provides petabytes of storage capacity in a fully scaled system.

**Accelerated Erasure Coding** – Storage Spaces Direct uses erasure coding for parity calculation in smaller deployments for the best possible efficiency and resiliency for two simultaneous failures. With larger clusters, storage efficiency is increased as there can be more data symbols

without increasing the number of parity symbols. Developed in collaboration with Azure, erasure coding is a very efficient way to store data, driving storage efficiency of 50% or higher, compared to 33% for mirroring.

## Deployment Options

Storage Spaces Direct is designed for two deployment options:

**Converged** – In this deployment option, storage and compute resources are in separate clusters. It layers a scale-out file server (SoFS) atop Storage Spaces Direct to provide network-attached storage over SMB3 file share. This allows for scaling compute/workload independently from the storage cluster, essential for larger-scale deployments such as Hyper-V IaaS (Infrastructure as a Service) for server providers and enterprises
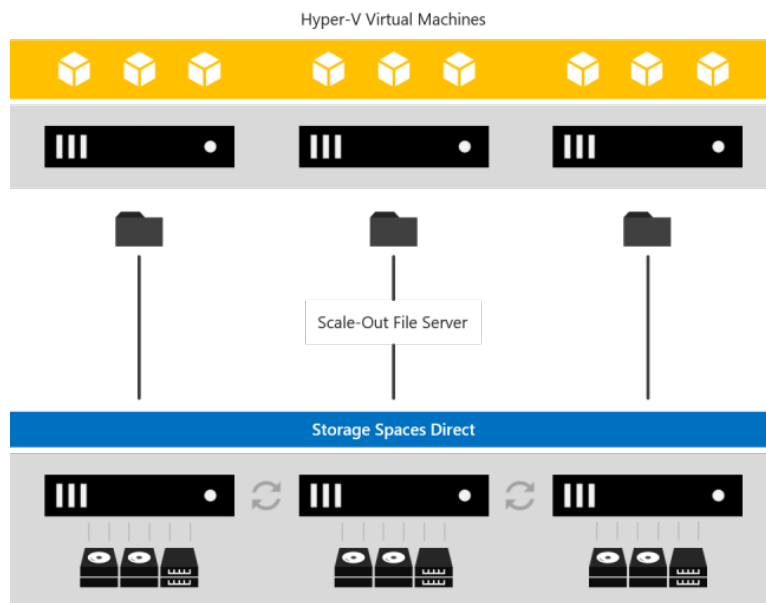


**Hyper-Converged** – This deployment option has one cluster for both compute and storage (such as DataON S2D appliances), and runs Hyper-V or SQL Server databases directly on the servers providing storage, storing their files on the local volumes. This eliminates the need to configure file server access and permissions, and reduces hardware costs for small-to-medium business or remote office/brand office deployments. SQL Server runs natively on Storage Spaces Direct, providing outstanding IOPS and throughput for SQL database operations.

# Storage Health Services Integration

Microsoft believes that partners such as DataON are best positioned to deliver an end-to-end solution user experience for customers. Customers have many different deployment models and use different kinds of hardware so Microsoft provides partners an API foundation to build monitoring and management experiences.

There are two major components of the Microsoft Health Services API:

- **Alerts** – An event-driven model that identifies and surfaces hardware and software problems.

- **Metrics** – Hooks into the core stack in Windows to gather, synchronize, and aggregate statistics.

DataON's exclusive MUST tool is the first to market that leverages the Health Services API to provide visibility, monitoring, and management for Windows Storage deployments.


# Storage Replica

Storage Replica is a new feature of Windows Server 2016.

- It is a storage agnostic system that utilizes HDD, SDD, and NVMe storage.

- It works at a block level, so it doesn't matter which file system you are using (e.g. ReFS, NTFS).

- It uses SMB3 networking, a modern protocol so it doesn't require special networking or expensive HBAs. SMB3 is a fast, wide transport that can be used with Storage Spaces Direct, in addition to Storage Replica.

    o It's used for block placement and block copy engine to transport I/O.

    o It features zero RPO replication which means zero data loss.

Storage Replica supports both synchronous and asynchronous data replication between floors, buildings, campuses, cities, and even continents.

With synchronous replication, there is no possibility of data loss ever. Every I/O is written to two places (source and destination) across two sites, whether it's across the same floor or even across different continents.

With asynchronous replication, I/O is written to the source and makes a best effort to write to the destination. You may experience some data loss in this scenario but your business can still function.

Storage Replica supports two types of scenarios that are relevant to DataON S2D and CiB (Cluster-in-a-Box) solutions: stretch and cluster-to-cluster.

In a stretch cluster, you take a normal cluster, place nodes in two separate locations, and tie them together via Storage Replica. It's an asymmetric cluster that has automatic failover and site awareness that looks like normal clustered storage. You can manage it with PowerShell or the easy-to-use Failover Cluster Manager tool.
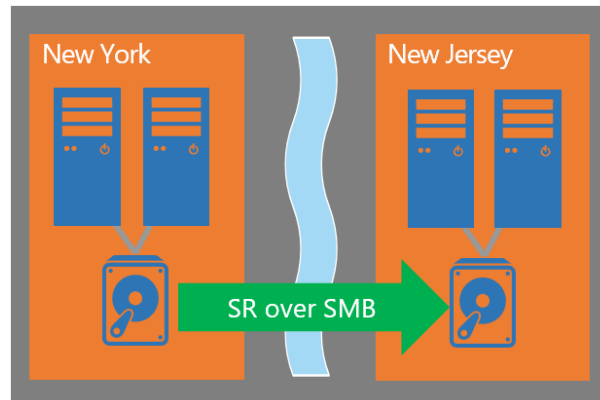


The alternative is cluster-to-cluster, where you take two clusters and tie them together with Storage Replica. This has a more manual orchestration, which may be more advantageous for those customers that prefer not to have automatic failover functionality. They typically want more control over when disaster recovery protocols are triggered since disaster sites usually don't run at full capacity or they are at an alte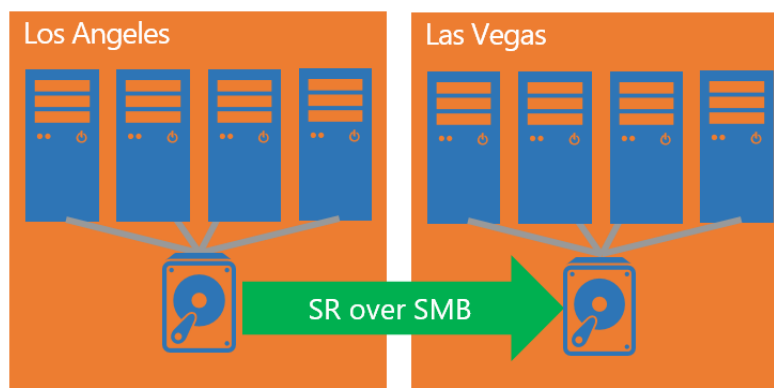rnative location that's only dedicated to disaster recovery. Instantaneous failover is important for certain workloads but not important for most workloads so it gives you time to decide on the appropriate course of action. Cluster-to-cluster can be currently managed with PowerShell, with future support expected via Azure Site Recovery.



Storage Replica has been available for almost two years, with several thousand successful customer deployments. Customers should feel confident that Storage Replica can be used safely as their disaster recovery solution, with good performance.

# Intel Technology for Windows Server 2016
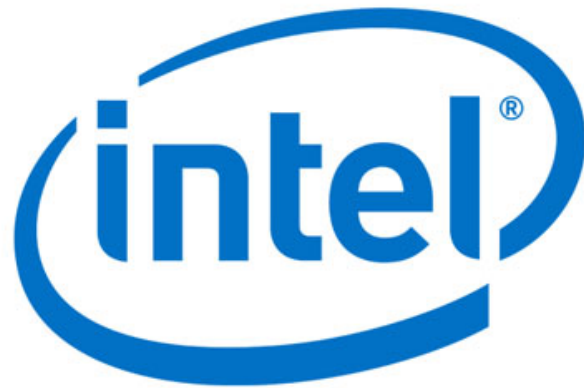
## Intel® Xeon® Processor E5-2600 v4 Family

The Intel Xeon processor E5-2600 v4 family extends the data center class storage features of previous generations. Manufactured using Intel's advanced 14nm processes, it adds updated features to items supported by Windows Server 2016, such as increased memory bandwidth, Intel® Resource Director Technology, Intel® QuickAssist Technology, and Intel® QuickData Technology. When combined with quality storage software, the Intel Xeon processor E5-2600 v4 family enables data centers to run more efficiently and use less power than previous generations.

The DataON S2D-3224i is one example of an Intel-based server system that provides scale-out and scale-up infrastructure and management services for deploying Microsoft Windows Server 2016. It is optimized for performance, density and capacity for the delivery of core Microsoft services and enterprise applications.

## Intel® Solid State Drive Data Center Family

Intel® SSD Data Center family of drives and PCIe storage devices offer full end-to-end data protection, consistent performance with low latencies, high write endurance, and scalability for growing storage needs while helping enterprises and clouds to tackle today's bigger storage challenges. Intel PCIe-based SSDs offer incredible performance and enhanced capabilities, with advanced capacity and performance coming in future Intel® Optane™ SSDs based on 3D XPoint™ Technology.

# Intel Data Center Blocks for Cloud and Microsoft Windows Server 2016

Intel Server Systems are engineered from the CPU out to help meet a wide range of business needs, from virtualization deployments to high-performance computing (HPC) infrastructure. Each design is built to a high specification, delivering server products with maximum processing power, great flexibility, excellent manageability, and high reliability. Stringent design and manufacturing practices, rigorous validation and testing, and excellent warranty and technical support ensure you receive incredible value.

The Intel® Data Center Blocks for Cloud (Intel® DCB for Cloud) and Microsoft Windows Server 2016 include both single node and multi-node server systems.  Server systems within this product family were specifically created to offer Intel customers with pre-configured systems that are Microsoft Windows Server 2016 certified. Intel has extensively tested these systems to ensure best operation and reliability with the Microsoft Windows operating environment.

# Mellanox RDMA over Converged Ethernet (RoCE) Solutions for Windows Server 2016

## Low Latency 10/25/40/50/100 Gigabit Ethernet End-to-End Solutions

Mellanox offers complete 10/25/40/50GbE solutions for Windows Server 2016 data centers with end-to-end RoCE support. These end-to-end solutions deliver high bandwidth and low latency to I/O intensive applications and fast flash storage, enabling data centers to operate with high performance and efficiency. By supporting speeds from 10GbE to 100GbE, Mellanox Spectrum switches and ConnectX-4 network adapter cards give IT managers flexibility in how they deploy higher bandwidth to the servers, providing simple upgrade paths from 10GbE to 40GbE, or to 25/100GbE.

Mellanox provides the default Ethernet switches and adapters for DataON S2D appliances.

## 10/25/40/50/100 Gigabit Ethernet Switching

The Spectrum SN2700 switch leverages Mellanox's most advanced switching architecture to deliver up to 2Tb/s of switching in a single 1U enclosure. The 32 QSFP ports can be configured to be used as 10/25/40/50/100GbE connections. The SN2700 supports thirty-two 40/100GbE ports, up to sixty-four 10/25GbE ports, up to thirty-two 50GbE ports, up any combination, providing flexibility for network installations. The high bandwidth is delivered with port-to-port latency as low as 300ns for improved application response time.

*Mellanox Spectrum SN2700 10/25/40/50/100GbE Switch*

The non-blocking SN2700 switches deliver the most predictable network performance at the line rate with very little variation at any packet size and I/O pattern. Combined with zero packet loss and a dynamically shared buffer mechanism, the SN2700 switch is an ideal switch for bursty and latency-sensitive storage traffic, especially NVMe SSDs.

Besides the SN2700, the Spectrum switch family also include the SN2410 (1RU, 48 10/25G ports with 8 40/100G uplink ports) and the SN2100 (half-width, 1RU, 16 10/25/40/50/100G ports). The compact design of the SN2100 switches make them ideal for high-density rack designs - Two half-width SN2100 switches side-by-side in 1RU space, with typical power consumption of <94 watts each, provide high availability in the TOR and deliver great savings in both CAPEX and OPEX.

# 10/25/40/50/100 Gigabit Ethernet Adapter Cards

ConnectX-4 10/25/40/50/100 Gigabit Ethernet adapters provide exceptional high performance for the most demanding data applications. ConnectX-4 adapters support RoCE specifications delivering low-latency and high-performance over Ethernet networks. Leveraging data center bridging (DCB) capabilities as well as advanced congestion control hardware mechanisms, ConnectX-4 RoCE provides efficient low-latency RDMA services over Layer 2 and Layer 3 networks. In addition to RoCE offload, ConnectX-4 adapters implement a rich set of features, from network function offload (e.g., LSO/LRO/RSS/TSS), SR-IOV support for I/O virtualization and Accelerated Switching and Packet Processing (ASAP[2]) technology to perform OVS data plane and VTEP functions.
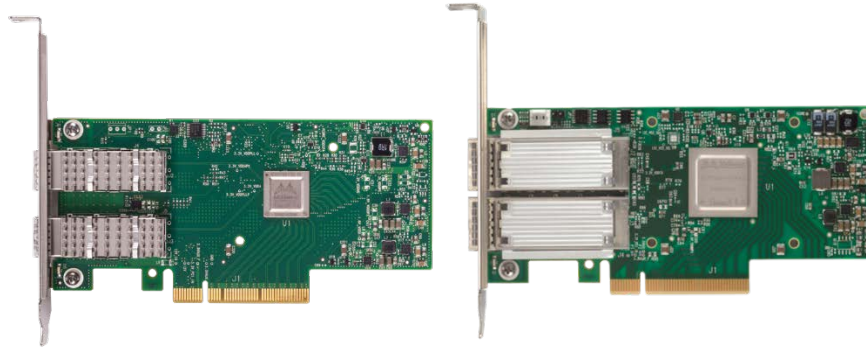
*Mellanox ConnectX-4 10/25/40/50/100GbE Adapters*

With RDMA/RoCE offload in the ConnectX-4 adapters and predictable high performance by the Spectrum switches, the Mellanox Ethernet solution accelerates Microsoft Storage Spaces Direct, unleashes the power of faster storage devices such as NVMe SSDs, and greatly improves server CPU and application efficiency.

# DataON: The Certified Microsoft Platform

## The DataON Difference

DataON is exclusively focused on customers who have made the "Microsoft choice" to deploy a Windows Server-based storage solution. Our team of Microsoft Server experts know how to design, deploy and support Windows Server storage and will work with you to performance tune your workloads. DataON storage solutions are:

- Certified for Windows Server 2012 R2, 2016 SDDC and Windows Server Software-Defined

- Customer-proven with over 500 enterprise installations and 100PB of DataON S2D storage deployments

- Optimized by our team of Microsoft experts to ensure successful deployments into your IT environment, tuned to your workloads

## DataON S2D Hyper-Converged Cluster Appliances (HCCAs)

The DataON S2D HCCAs are built to optimize the full stack of Microsoft Storage Spaces Direct in a hyper-converged platform. They are built with integrated compute, network and storage infrastructure with near-linear scalability to simplify and maximize the deployment of Microsoft applications, virtualization, data protection and hybrid cloud services. Each pre-configured cluster can support 40 Hyper-V VMs per node, for expanded capacity and operational flexibility.

From scale-out file server (SoFS) and software storage bus to storage and networking hardware, this appliance runs on the cluster Shared Volumes Resilient File System (ReFS) and uses high performance NVMe SSDs with SMB 3.0 networking to maximize performance and capacity.

*DataON S2D-3224i Hyper-Converged Cluster Appliance*

The DataON S2D-3224i is an Intel-based server system that provides scale-out and scale-up infrastructure and management services for deploying Microsoft Windows Server 2016. It is optimized for performance, density and capacity for the delivery of core Microsoft services and enterprise applications. The S2D-3224i is design on three core principles:

- Scale-out hyper-converged cluster
- Integrated software-defined services
- Complete visibility and management of the storage infrastructure via DataON's exclusive MUST infrastructure and management software tool

# DataON MUST™ (Management Utility Software Tool) Visibility and Management Tool

The DataON S2D and CiB storage solutions are integrated with the exclusive DataON MUST visibility and management tool. It provides infrastructure visibility and management for Microsoft's new suite of software-defined storage stack technologies like Storage Spaces Direct, Storage Replica, and storage quality of service (QoS) which, based on your policy, monitors hardware and software storage infrastructure to identify potential problems. Using an event-driven model for rapid detection with minimal overhead, MUST also provides on-demand access to curated collections of hyper-converged clusters, storage performance, and capacity metrics. The MUST dashboard display is designed to efficiently and dynamically connect the dots to help provide root cause analysis.
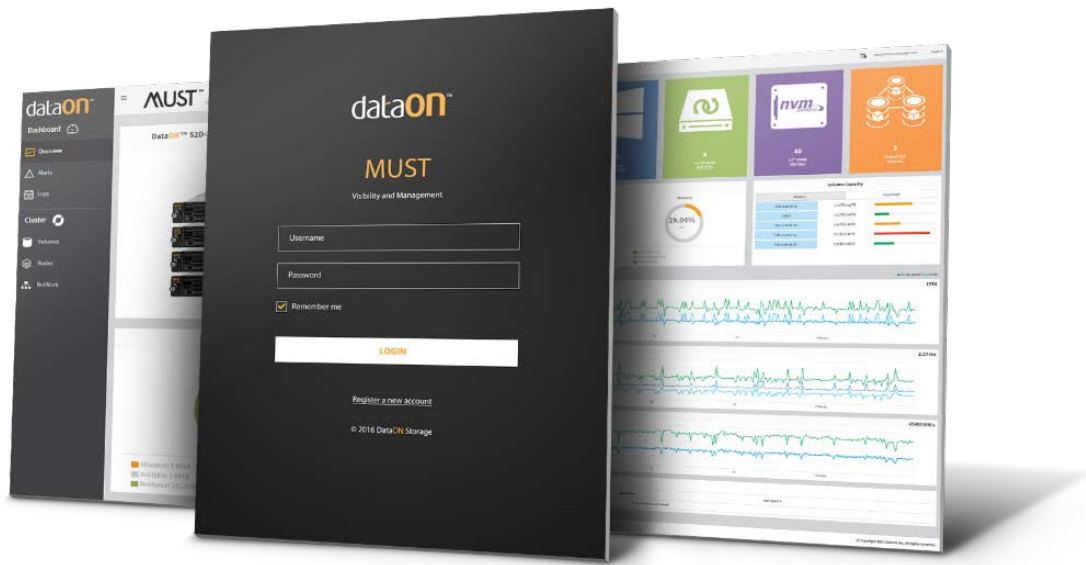
MUST offers complete integration with Microsoft Storage Health Services for Windows Server 2016. It is built to provide multiple tiers of storage visibility and monitoring.

**SDDC & Hyper-Converged Infrastructure Tier** – System-level information on performance, capacity, hardware inventory and faults/alerts. MUST gives you a dashboard level view of your operations, analytics, infrastructure health management, storage systems metrics and event logging insights.

**Systems and Storage Services Audit Log Tier** – Detailed logging level visibility for events, so you can perform root cause analysis and export source data for analytics.

**Hyper-Converged Cluster Appliance (HCCA)/Node Tier** – Pool, volume and device-level performance, health and operational analytics for your HCI cluster. This enables you to proactively perform systems maintenance and better understand requirements for workload migrations.

**SAN-like Call Home Service Support** – Leveraging the Health Services Faults in Windows Server 2016, administrators can have automated email alerts sent to key contacts. You can also leverage third party SNMP monitoring traps to alert you when you need disk or hardware replacements.

SOS has been using DataON MUST with their S2D deployment. "MUST has been very valuable and was a big selling point. To us, it demonstrated a commitment from DataON to support storage as a service," said Benjamin Clements, President, Strategic Online Systems, Inc. "The inclusion of MUST with the S2D appliance is what completes the solution with Storage Spaces Direct as a viable SAN replacement."

# Deploying a Windows Server 2016 Software-Defined Solution with Youth Villages and Strategic Online Systems Inc.



## Finding the Solution

Once SOS decided on a Windows Server 2016 Storage Spaces Direct solution, they researched offerings from Lenovo, HP, Dell, and Super Micro. However, they didn't think any of them were fully committed to supporting a Storage Spaces Direct solution.

SOS's search led them to DataON. They found that DataON had a growing reputation in the industry as the only experts in Storage Spaces Direct and designing and deploying Microsoft-based storage. "Anybody can put out guidelines for how to deploy Storage Spaces Direct but customers don't view that as a complete solution. DataON is the ONLY real Microsoft Storage Spaces Direct vendor in this market," said Benjamin Clements, President, Strategic Online Systems, Inc. Multiple mentions about DataON at Microsoft's Ignite conference confirmed what they were hearing and they decided on a Storage Spaces Direct and DataON solution.

DataON proposed their S2D-3224i hyper-converged appliance (HCCA) to fulfill Youth Villages' requirements:

- Hybrid NVMe + SATA SSD storage
- Runs enterprise apps, including SQL and VDI
- Achieves over 1.0M IOPS

- Supports over 40 VMs for Hyper-V
- Leverages SMB 3.0 over RDMA for low latency and high CPU efficiency
- Includes DataON's exclusive MUST visibility and management software tool

SOS purchased four DataON S2D-3224i hyper-converged cluster appliances (HCCAs) for Youth Villages' data centers. Optimized for balanced IOPS and capacity, this all-flash NVMe and SATA SSD solution provides maximum flash performance in a 4-node cluster. It has been tested and certified to be part of the Windows Server Software-Defined (WSSD) program, following Microsoft requirements and best practices for a Windows Server 2016 Software-Defined infrastructure. The S2D-3224i is also certified for Windows Server 2016 SDDC and Windows Server 2012 R2.

SOS was also looking for a good monitoring system for their new hyper-converged cluster. DataON was able to respond to that requirement with their exclusive MUST (Management Utility Software Tool) visibility and management software, which is pre-configured with their S2D storage solutions. SOS was impressed with its powerful features, such as:

- Dashboard-level metrics from a single pane of glass
- System alerts based on Windows Health Service faults with built-in root cause analysis
- Call-home support that alert system administrators based on fault settings and severity levels

# Deploying the Solution

DataON worked very closely with SOS for the design, testing, validation, proof-of-concept, and deployment.

There were some initial challenges with deploying the DataON and Storage Spaces Direct solution. SOS found that their existing Cisco switch did not support RDMA, and needed to replace them with Mellanox 40GbE switches. Being a newer technology, there was also a learning curve with deploying RDMA. In SOS's previous Microsoft deployments, Storage Spaces only supported a SAS infrastructure. With Storage Spaces Direct, it leverages an Ethernet network and takes advantage of RDMA over Ethernet for a loss-less network with no packet drops.

**Windows Server 2016 Converged Cluster Appliance**
S2D-3224i Connection Diagram

Hyper-V Cluster

Hyper-V Server #1
to A5
to B5

Hyper-V Server #2
to A6
to B6

Hyper-V Server #3
to A7
to B7

Hyper-V Server #4
to A8
to B8

Hyper-V Server #5
to A9
to B9

Hyper-V Server #6
to A10
to B10

Hyper-V Server #7
to A11
to B11

Global Network

Global Network Switch

S2D Converged Cluster Appliance with RDMA

Windows Server 2016 Certified

Switch A
Switch B

Interconnect
Interconnect
Global
Global

to A1    to B1
S2D-3224i #1

to A2    to B2
S2D-3224i #2

to A3    to B3
S2D-3224i #3

to A4    to B4
S2D-3224i #4

# Hardware Configuration and Deployment Tips

Here are some basic configuration and deployment tips to help you get started.

| Hardware – Windows Server 2016 Server and Storage Certifications | • Make sure your system is certified for Microsoft Server, Microsoft Software-Defined Data Center, and Window Server Software-Defined, for the most optimized infrastructure, operational visibility and reliability for their Windows Server 2016 based software-defined data center.<br>• Check that all the S2D hardware components are certified for Microsoft Windows Server 2016.<br>• Intel Processor selection – Keep in mind the CPU clock speed or frequency and cache has a direct impact on the performance of workloads like SQL and VDI.<br>• Make sure you have enough memory for each node/system because it will impact application memory allocation to application workloads such as SQL.<br>• For the Windows Server Storage Spaces Direct storage bus **cache tier** selection, select high endurance and make sure you have enough storage capacity (see Intel Data Center SSDs).<br>• For the Windows Server Storage Spaces Direct storage bus **performance tier** selection, you should mirror to match at least your most demanding workload. You should also make sure you have enough storage capacity (see Intel data center SSDs).<br>• For the Windows Server Storage Spaces Direct storage bus **capacity tier** selection, you can select mirror or RS 2+2 with multi-resilient volume or erasure coding to meeting your storage capacity needs (see HGST HDDs).<br>• Deploy SMB3 over RDMA networking with DCBx enabled switches. |
| --- | --- |
| SMB3 over RDMA Networking Fabric | • With S2D storage, use SMB3 multi-channel RDMA networking for consistent performance and low latency.<br>• Make sure your DCB switch supports RDMA, as well as supports priority flow control (PFC).<br>• Make sure you setup your DCBx switch with the right parameters. Determine if either lowest latency or highest bandwidth is your priority, which will affect some settings such as jumbo frame size.<br>• In choosing a 1-port RDMA NIC versus a 2-port RDMA NIC, consider your bandwidth saturation and need for dedicated data paths for PCIe 3.0 lanes<br>• For best performance, use a 40GbE end-to-end network |

| | |
|---|---|
| | • Using a 40GbE to 10GbE splitter is not advised.<br>• Need to develop customized S2D network deployment charts with subnet, gateways, and VLAN ID for your SMB fabric, host, cluster, live migration and others (refer to the **DataON S2D deployment checklist**). |
| **DataON S2D Storage Setup** | • Make sure the VHDX files for the VMs are configured with 4096 physical sector size bytes (instead of 512) to ensure good performance and low write latency.<br>• Make sure you understand your workloads' requirements before configuring S2D storage. For example, SQL databases require low latency writes so you should place tempdb and log files on 3-way mirror volumes. Read-heavy loads can be placed on MRV volumes.<br>• Evaluate the endurance of each tier for daily writes. Microsoft's Cosmos Darwin has a great blog to help you [understand SSD endurance](#) for Storage Spaces Direct.<br>• Make sure you use the **DataON S2D platform checklist** to ensure proper system configuration.<br>• Make sure you have our customized deployment guide and **S2D installation PowerShell script** for your S2D deployment with correct IP for the infrastructure. |

With the purchase of any DataON S2D appliance, we will provide a detailed deployment guide, customized PowerShell scripts, and a driver pack to help you get you running with Storage Spaces Direct.

The configuration checklist includes:

- Cabling diagram
- Switch configuration
- DataON S2D application configuration
    - System
    - Networking
- Windows features installation
    - Hyper-V
    - Failover Clustering
    - File Services
    - Data Center Bridging
- Quality of Service setup
- Virtual switches setup
- Virtual networking setup

- Cluster creation & validation
- Storage Spaces Direct configuration
- Performance testing
  - VM Fleet setup
  - Task examples
- DataON MUST visibility and management tool
  - Installation
  - Configuration
- Testing
  - Testing failover
  - Validate volumes failover
  - Using MUST to test resiliency
  - Simulate drive failures
  - Simulate node failures

The driver pack can be found on the C: drive includes the latest drivers for:

- Intel NVMe SSD
- Intel Data Center Tool
- Intel Onboard Chipset
- Intel Onboard 1G
- Mellanox ConnectX-4

The customized PowerShell scripts can also be found on the C: drive.

# Benchmarks and Results

After setup and migration, SOS and Youth Villages saw a significant performance improvement in their new virtualized environment. Youth Villages' database was now seeing:

- Improvement from 13-30ms delay to less than 1ms delay write on the storage
- Ability to accommodate bursts of up to 200,000 IOPS
- Ability to withstand huge bandwidth bursts while still delivering 15,000 IOPS and low latency to current VMs
- Reduction in latency from 30-150ms to 1ms.
- Reduction in disk queuing from 100-400 to under 2.

A detailed performance benchmark was done with Microsoft VM Fleet, testing for IOPS, throughput, latency and resiliency. The setup was configured for 3-way mirror, MRV, and parity.

## 3-Way-Mirror

VM Fleet testing (20 virtual machines per node)

```
Volume   Filesystem Capacity Used Resiliency    Size (Mirror) Size (Parity) Footprint Efficiency
------   ---------- -------- ---- ----------    ------------- ------------- --------- ----------
CCSD-N1 ReFS        24TB     1%   3-Way Mirror  24TB          0             71TB      33%
CCSD-N2 ReFS        24TB     1%   3-Way Mirror  24TB          0             71TB      33%
CCSD-N3 ReFS        24TB     1%   3-Way Mirror  24TB          0             71TB      33%
CCSD-N4 ReFS        24TB     1%   3-Way Mirror  24TB          0             71TB      33%
CCSD-N5 ReFS        24TB     1%   3-Way Mirror  24TB          0             71TB      33%
CCSD-N6 ReFS        24TB     1%   3-Way Mirror  24TB          0             71TB      33%
```

## Random Reads and Writes

Block size 4Kb, 8 threads, 8 outstanding I/O (0% write / 100% read)

```
Administrator: Windows PowerShell

CSV FS       IOPS        Reads       Writes    | BW (MB/s)   Read     Write  | Read Lat (ms)  Write Lat
Total        2,020,563   2,020,224   338       | 8,277       8,274    3      |
CCSD-N1      330,240     330,172     67        | 1,353       1,352    1      | 0.234          0.457
CCSD-N2      302,786     302,716     70        | 1,240       1,240    1      | 0.236          0.415
CCSD-N3      328,944     328,866     77        | 1,347       1,347    1      | 0.231          0.437
CCSD-N4      415,866     415,864     1         | 1,703       1,703           | 0.288          0.487
CCSD-N5      309,566     309,475     91        | 1,268       1,268    1      | 0.243          0.430
CCSD-N6      333,161     333,130     32        | 1,365       1,364           | 0.233          0.396
--------------------
SYS          CPU (%)
Total        462
CCSD-N1      76
CCSD-N2      80
CCSD-N3      78
CCSD-N4      70
CCSD-N5      80
CCSD-N6      79
--------------------
```

Block size 4Kb, 8 threads, 8 outstanding I/O (100% write / 0% read)

```
Administrator: Windows PowerShell

CSV FS       IOPS        Reads     Writes    | BW (MB/s)   Read     Write  | Read Lat (ms)  Write Lat
Total        441,180     1         441,179   | 1,815                1,815  |
CCSD-N1      75,300                75,300    | 310                  310    | 0.000          12.527
CCSD-N2      72,014                72,014    | 296                  296    | 0.000          14.193
CCSD-N3      72,526                72,526    | 298                  298    | 0.000          13.567
CCSD-N4      73,172                73,172    | 302                  302    | 2.825          13.610
CCSD-N5      78,270                78,270    | 322                  322    | 2.057          12.295
CCSD-N6      69,898                69,898    | 287                  287    | 0.000          14.443
--------------------
SYS          CPU (%)
Total        321
CCSD-N1      56
CCSD-N2      56
CCSD-N3      57
CCSD-N4      35
CCSD-N5      62
CCSD-N6      55
--------------------
```

Block size 4Kb, 8 threads, 8 outstanding I/O (30% write / 70% read)

```
Administrator: Windows PowerShell
CSV FS      IOPS        Reads       Writes      | BW (MB/s)   Read    Write   | Read Lat (ms)  Write Lat
Total       1,091,077   763,639     327,438     | 4,470       3,128   1,342   |
CCSD-N1     187,552     131,173     56,379      | 768         537     231     | 1.909          3.853
CCSD-N2     183,780     128,703     55,077      | 753         527     226     | 1.668          2.106
CCSD-N3     174,835     122,431     52,404      | 716         501     215     | 3.180          4.111
CCSD-N4     214,081     149,809     64,272      | 877         614     263     | 1.851          5.406
CCSD-N5     178,258     124,755     53,503      | 730         511     219     | 2.611          3.309
CCSD-N6     152,572     106,768     45,804      | 625         437     188     | 4.257          4.745
---------------------
SYS         CPU (%)
Total       447
CCSD-N1     77
CCSD-N2     83
CCSD-N3     75
CCSD-N4     61
CCSD-N5     80
CCSD-N6     71
---------------------
```

Block size 4Kb, 8 threads, 8 outstanding I/O (50% write / 50% read)

```
Administrator: Windows PowerShell
CSV FS      IOPS        Reads       Writes      | BW (MB/s)   Read    Write   | Read Lat (ms)  Write Lat
Total       851,232     425,080     426,151     | 3,486       1,740   1,746   |
CCSD-N1     160,993     80,434      80,559      | 660         329     330     | 1.550          4.135
CCSD-N2     117,891     58,905      58,986      | 483         241     242     | 7.095          7.551
CCSD-N3     155,108     77,532      77,576      | 635         317     318     | 1.978          6.721
CCSD-N4     140,598     70,210      70,389      | 576         287     288     | 2.861          9.228
CCSD-N5     134,887     67,231      67,656      | 552         275     277     | 2.775          8.820
CCSD-N6     141,755     70,769      70,986      | 581         290     291     | 3.098          5.639
---------------------
SYS         CPU (%)
Total       431
CCSD-N1     77
CCSD-N2     72
CCSD-N3     76
CCSD-N4     49
CCSD-N5     77
CCSD-N6     80
---------------------
```

## Sequential Reads and Writes

Block size 512Kb, 1 thread, 1 outstanding I/O (0% write / 100% read)



```
Administrator: Windows PowerShell

CSV FS        IOPS      Reads     Writes    | BW (MB/s)  Read      Write     | Read Lat (ms)  Write Lat
Total         44,329    44,325    4         | 23,236     23,236              |
CCSD-N1       7,931     7,930     1         | 4,157      4,157               | 2.388          1.972
CCSD-N2       5,967     5,967               | 3,128      3,128               | 3.217          0.000
CCSD-N3       7,172     7,171     1         | 3,759      3,759               | 2.652          2.059
CCSD-N4       8,522     8,522               | 4,468      4,468               | 2.200          1.587
CCSD-N5       7,299     7,297     1         | 3,826      3,826               | 2.607          6.668
CCSD-N6       7,437     7,437               | 3,899      3,899               | 2.550          0.000
-------------------
SYS           CPU (%)
Total         36
CCSD-N1       6
CCSD-N2       7
CCSD-N3       7
CCSD-N4       5
CCSD-N5       6
CCSD-N6       6
-------------------
```

Block size 512Kb, 1 thread, 1 outstanding I/O (100% write / 0% read)



```
Administrator: Windows PowerShell

CSV FS        IOPS      Reads     Writes    | BW (MB/s)  Read      Write     | Read Lat (ms)  Write Lat
Total         4,423     1         4,421     | 2,209                2,209     |
CCSD-N1       732                 732       | 363                  363       | 0.000          27.951
CCSD-N2       788       1         787       | 390                  390       | 0.693          25.980
CCSD-N3       708                 708       | 351                  351       | 0.000          29.164
CCSD-N4       748                 748       | 372                  372       | 0.000          26.944
CCSD-N5       776                 776       | 393                  393       | 0.000          26.152
CCSD-N6       671                 670       | 340                  340       | 1.144          30.203
-------------------
SYS           CPU (%)
Total         21
CCSD-N1       3
CCSD-N2       4
CCSD-N3       4
CCSD-N4       1
CCSD-N5       3
CCSD-N6       5
-------------------
```

# Benchmark & Testing Tips

| DataON S2D testing & validation | • Use Microsoft's **Diskspd** utility for testing and benchmarking. Use the cluster health PowerShell command line to validate your cluster setup (at the system and virtual disk levels).<br>• Use VM Fleet for performance tuning.<br>• DataON MUST is the ideal tool for real-time diagnostics. |
|---|---|
| Measure latency from the application perspective | • On the servers using Storage Spaces Direct, open the **Resource Monitor** (go to **Task Manager** and click on **Resource Monitor**).<br>• Setup steps<br>   o Open **Task Manager**<br>   o Click **Resource Monitor** at the bottom of the screen<br>   o Click **Disk tab** at the top of the screen<br>   o Expand **Disk Activity**<br>   o Click **Total (B/sec)** to sort from highest to lowest<br>• **Testing**<br>   o See latency in the "Response Time" column for the processes that are the highest **total bytes/sec** AND "normal" I/O priority.<br>• **Real World Results**<br>   o Before, Youth Villages was seeing 30-150ms consistently for latency on tempdb, normal log files and database.<br>   o After implementing Storage Spaces Direct, Youth Villages is consistently seeing 1ms latency.<br>   o Before Youth Villages saw that disk queuing (also through Performance Monitor) would consistently reach 100-400. After implementing Storage Spaces Direct, disk queuing stays below 2 consistently. |
| Measure performance outside of the application, from the server perspective | • Download and extract **diskspd** (current version is 2.0.17)<br>• Open an elevated command prompt:<br>   o run D:\Diskspd-v2.0.17\amd64fre\diskspd -b64k -d15 -h -L -o8 -t8 -r -w100 -c50M T:io.dat<br>   o Where: |

| d: | Drive installed the application **diskspd** is installed. |
|---|---|
| -b64k | Testing with 64K blocks (use only numbers divisible by 4). |

| | | -d15 | Test for 15 seconds. | |
|---|---|---|---|---|
| | | -h | Disable software caching and set write-through I/O (to force the storage to use the disks and not just pull from cache). | |
| | | -L | Measure latency statistics. | |
| | | -o8 | Defines 8 outstanding I/O requests per target per thread (if monitoring disk queueing on performance monitoring, you will see the disk queue average around 8x the number of threads. | |
| | | -t8 | Defines 8 threads per target. In this case, disk queuing will be around 64, if looking at performance monitoring under the Disk tab/Storage/Disk Queue Length column. | |
| | | -r | Specifies random I/O. | |
| | | -w100 | Determines how much of the workload is writes. In this case, looking to see 100% of writes, which is admittedly the hardest load to put on the storage. | |
| | | -c50M | Create a test file that is 50 Mbytes. | |
| | | t:io.dat | Tells **diskspd** what drive needs to be tested. This can be a drive letter or a UNC path. If testing the host, use the UNC path pointing to the SOFS share. If testing the VM, use the drive letter that is encapsulated by the VHDX. | |

# Summary

SOS was looking to replace Youth Villages' aging SAN with a new hyper-converged solution. They had previous success with Windows Server 2012 R2 Storage Spaces and were excited about the upcoming Windows Server 2016 Storage Spaces Direct. With an eye for the future, they decided on a Storage Spaces Direct solution.

SOS's search for a vendor that truly supports Storage Spaces Direct solutions led them to DataON. DataON's reputation in the industry as Microsoft experts gave SOS confidence in a DataON and Microsoft SDS solution.

SOS worked with DataON to deploy four DataON S2D-3224i hyper-converged clusters. The all-flash NVMe and SATA SDD solution was certified for the Windows Server Software-Defined program, as well as Windows Server 2016 SDDC and Windows Server 2012 R2. The new Windows SDS solution utilized RDMA networking with 40GbE Mellanox switches, for low latency performance and increased CPU efficiency.

The S2D-3224i clusters were pre-configured with DataON's exclusive MUST visibility and monitoring tool to provide visibility, monitoring, and management for Windows SDS. In SOS's view, the inclusion of MUST completes the solution with Storage Spaces Direct as a viable SAN replacement.

The new solution ended up providing about 30x reduced latency and the ability to withstand huge bursts of demand with no impact to latency. In addition, they were able to save about 50% over the cost of a traditional SAN.

In addition to the increased performance, SOS is happy with the reduced cost of their SAN replacement. They were able to save about 50% over the cost of a traditional SAN. Because Youth Villages is a nonprofit organization, SOS was also able to get a reduction in the price for the Microsoft Server 2016 license.

As Youth Villages' storage needs increase, SOS will be able to easily expand the S2D solution by adding storage nodes as needed. With traditional SANs, upgrading can be a more complex process and may include the need for rip-and-replace.

As a result of the successful deployment of the DataON and Storage Spaces Direct solution, SOS is now looking at creating a backup solution for Youth Villages, using DataON CiB appliances and Storage Replica.

# Trademarks

DataON and the DataON logo, and MUST are trademarks of DataON in the United States.

The following terms are trademarks of other companies: Intel, Xeon, and the Intel logo are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries. Active Directory, Hyper-V, Microsoft, SQL Server, Windows, Windows Server, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both. Other company, product, or service names may be trademarks or service marks of others.

# Appendix

## System Specifications

- 4x DataON S2D-3224i Nodes
- 8x Intel Xeon E5-2620v4 CPU with 128G RAM per node
- 4x Mellanox ConnectX-4 dual-port 40GbE RDMA NIC
- 8x SanDisk 128GB Boot drive
- 8x Intel P3700-800GB NVMe (high endurance cache)
- 80x Intel S3520-1.6TB SATA SSD (capacity SSD)
- 2x Mellanox 40GbE Switch
- Fault Tolerance = 2; 3-Way Mirror ~ 9%, MRV ~ 48%
- Usable Capacity~ 60TB

# Firmware

| Component | Platform | Category | Device Part Number (P/N) | Supported Version |
|---|---|---|---|---|
| BIOS | S2D-3224i | BIOS | N/A | SE5C610.86B.01.01.0022 |
| BMC | S2D-3224i | KVM over IP | N/A | 1.50.10802 |
| Expander Storage Backplane | S2D-3224i | Storage - Backplane | N/A | 1300 |
| Intel® Ethernet Controller X540 Series / 10GbE | S2D-3224i | Network / On-board | N/A | N/A<br>Driver Version 4.0.215.0 |
| Mellanox ConnectX-4 EN /40GbE | S2D-3224i | Network / RDMA | MCX414A-BCAT | 12.18.2000 |
| Broadcom HBA 9300-8i | S2D-3224i | Storage - HBA | LSI00344 | 14.00.00.00 |
| SanDisk X400 / 256GB | S2D-3224i | Storage – SATA SSD (Boot) | SD8SN8U-256G-1122 | X4140000 |
| HGST Ultrastar He10 / 10TB 4Kn | S2D-3224i | Storage – SAS HDD | 0F27402 | A21D |
| Intel SSD DC S3520 Series / 1.6TB | S2D-3224i | Storage – SATA SSD | SSDSC2BB016T701 | N2010101 |
| Intel SSD DC P3700 Series / 800GB | S2D-3224i | Storage – PCIe SSD/ NVMe | SSDPE2MD800G401 | 8DV101F0 |

## Switches

| Component | Type | Platform | Category | Supported Version |
|---|---|---|---|---|
| Mellanox SN2100 | MLNX-OS | SN2100 | Network Switch | MLNX-OS 3.6.3508 |
| Mellanox SN2700 | MLNX-OS | SN2700 | Network Switch | MLNX-OS 3.6.3508 |

# Hostname & IP Configuration

| Name | Description | SMB A | SMB B | Host Mgmt |
|---|---|---|---|---|
| Example-N1 | S2D Node1 | 172.16.10.11 | 172.16.20.11 | 172.16.30.11 |
| Example-N2 | S2D Node2 | 172.16.10.12 | 172.16.20.12 | 172.16.30.12 |
| Example-N3 | S2D Node3 | 172.16.10.13 | 172.16.20.13 | 172.16.30.13 |
| Example-N4 | S2D Node4 | 172.16.10.14 | 172.16.20.14 | 172.16.30.14 |

| Networks | Subnet | Gateway | VLAN ID | Description | Preferred DNS |
|---|---|---|---|---|---|
| SMB Fabric A | 172.16.10.0/24 | | 10 | Storage Fabric | |
| SMB Fabric B | 172.16.20.0/24 | | 20 | Storage Fabric | |
| Host Mgmt | 172.16.30.0/24 | 172.16.30.1 | 30 | S2D Node Management Network | |

# Collateral & Resources

## Collateral

[Microsoft Windows Server 2016 datasheet](#)

[DataON S2D-3224i HCCA](#)

[DataON MUST datasheet](#)

[Mellanox Ethernet Switches](#)

[Mellanox Ethernet Adapters](#)

## Videos

[Storage Spaces Direct in Windows Server 2016 presentation](#)

[DataON MUST demo video](#)