Implementing Windows Server 2019 Storage Spaces Direct using HPE ProLiant servers
## Contents

1.0 Introduction.................................................................................................................................................. 3  
2.0 HPE Azure Stack HCI solutions ...................................................................................................................... 3  
   2.1 HPE hybrid.................................................................................................................................................. 3  
   2.2 HPE all-flash............................................................................................................................................... 3  
3.0 Configuration best practices............................................................................................................................ 4  
   3.1 Environment preparation............................................................................................................................. 4  
   3.2 Network design......................................................................................................................................... 5  
   3.3 Deploying hyperconverged solutions......................................................................................................... 10  
   3.4 Deploying converged solutions.................................................................................................................. 12  
4.0 Troubleshooting........................................................................................................................................... 12  
   4.1 Cluster verification.................................................................................................................................... 12  
   4.2 Disk management..................................................................................................................................... 13  
   4.3 Disk rebuild.............................................................................................................................................. 13  
   4.4 Disk clean................................................................................................................................................ 13  
   4.5 Cluster management................................................................................................................................ 14  
   4.6 Network management............................................................................................................................... 14  
   4.7 Remove Storage Spaces Direct.................................................................................................................. 14  
5.0 Additional resources..................................................................................................................................... 14  
Appendix 1: Performance validation method...................................................................................................... 15  
   A1.1 VMFleet for Storage Spaces Direct.......................................................................................................... 15  
   A1.2 How to monitor the performance of VMFleet .......................................................................................... 16
1.0 Introduction
This document discusses Azure Stack HCI solution implementation using HPE ProLiant Gen9 and Gen10 servers. Windows Server® 2019 Storage Spaces Direct (S2D) is a software-defined storage feature based on Storage Spaces technology. It uses internal storage devices such as SATA and SAS HDDs, SSDs, and NVMe disks to create shared disk volumes. Distributed among several cluster nodes, these volumes can withstand disk, node, and network failures, and continue to operate normally without downtime. Different levels of resiliency can be achieved using mirror, parity, or a combination of the two. S2D is mainly targeted at cloud applications, Hyper-V, storage for backup and replication, and hyperconverged/converged infrastructures.

2.0 HPE Azure Stack HCI solutions
HPE Azure Stack HCI solutions are configured to deliver solid performance, improved reliability, and high availability. They can be implemented as 2 to 16 node configurations, providing you with the ability to deploy an engineered, tested, and validated private cloud environment to meet your most demanding workloads. All HPE Azure Stack HCI solutions are available as hyperconverged or converged solutions.

2.1 HPE Hybrid
HPE hybrid solutions start with two nodes of HPE ProLiant servers using RDMA-capable network adapters to provide high performance with minimum processor utilization. In order to make use of HPE solutions using RDMA networking, an RDMA compatible networking infrastructure is required. HPE hybrid solutions combine different types of drives (HDD/SSD/NVMe) to allow read/write caching and are available in both LFF and SFF form factors.

2.2 HPE All-Flash
HPE all-flash solutions are configured entirely with either SSD or NVMe flash drives and 25 Gb or faster RDMA network adapters to support performance-intensive workloads with HPE ProLiant and HPE Synergy Gen9/Gen10 servers.
### 3.0 Configuration best practices

This section provides some best practices for configuring Storage Spaces Direct. The procedures outlined in the examples are based on a 4-node HPE DL380 Gen10 SFF configuration. However, the same procedures, with some minor adjustments, can be applied to other configurations.

#### 3.1 Environment preparation

##### 3.1.1 System configuration

1. For each HPE DL380 server, install the SSD/HDD drives, network adapters, and storage controller as shown in the following figure.

![Figure 2: HPE DL380 Gen10 front/rear view](image)

2. The SSD and HDD drives intended for use in your S2D storage pool should be placed in zones that are in physically approximate for the ease of management and maintenance. In the example above showing the Gen10 DL380 SFF solution, they should all be placed in frontal all the S2D drives are placed in the front drive slots. All the S2D storage pool drives should must be connected to one or more Smart Array Controller running in HBA mode only.

3. For OS boot drives, you can choose between SSD, HDD, or Embedded M.2 depends depending on your design choices and system capabilities. You can choose to create either a RAID boot volume or simple AHCI boot volume. Boot drives must be connected to a Smart Array Controller or the onboard SATA controller that is not connected to any S2D storage pool drives. In the example above showing a Gen10 DL380 SFF solution, a recommend configuration is to place the boot drives in the rear cage and connected to the onboard S100i controller which can be configured as a RAID volume by configuring UEFI options.

##### 3.1.2 Firmware update

Download the latest Service Pack for ProLiant (SPP) and boot from this ISO image file to update the system firmware.

##### 3.1.3 Storage controller configuration

1. Make sure each node has the default RBSU settings.
2. For Gen10 storage controllers, HBA mode is enabled by default.
   a. S2D drives may only reside on a storage controller configured exclusively in HBA mode.
   b. S2D drives cannot reside on a storage controller configured in either RAID or mixed mode (RAID logical drives and HBA physical drives simultaneously).
3. For Gen9 storage controllers, enable HBA mode on the storage controllers as follows:
   a. Power on the server and press F10 during POST to enter the intelligent provisioning utility.
   b. On the next screen, select HPE Smart Storage Administrator.
   c. From the main HPE Smart Storage Administrator screen, select HPE P840 controller.
   d. Under Actions, click the Configure button.
   e. If the controller is already in HBA mode, you will see the Enable RAID Mode button. In this case, skip to Step g.
   f. If the controller is in RAID mode, select Enable HBA Mode, then click OK and Finish.
   g. Click the X button in the upper right-hand corner followed by OK to exit the application.
   h. Click the green power icon in the upper right-hand corner followed by Reboot.
4. Use HPE SSA to create an RAID 1 volume with the boot drives.
### 3.1.4 Operating system installation and configuration

2. Apply the Service Pack for ProLiant (SPP). Make sure to install the chipset, HPE iLO, network, and storage drivers. All others are optional. Reboot the system after completion.
3. Use PowerShell to install the required Windows® features:
   ```
   PS C:\> Install-WindowsFeature -Name File-Services, Failover-Clustering, Hyper-V, Data-Center-Bridging -IncludeManagementTools
   -Restart
   ```
4. Run Windows Update.

### 3.2 Network design

#### 3.2.1 Network connections

Connect each server node to each switch as shown in Figure 3. For added resiliency, we recommend using dual switches. Each network adapter has its ports connected to each of the two switches. There are two S2D network adapters in this configuration. If switch A or B fails, the S2D network will continue to operate using the remaining healthy switch. If one network adapter fails in a node, the S2D network will continue normal operation using the remaining healthy network adapter.

![Recommended network configuration](image1)

**Figure 3.** Recommended network configuration

For configurations consisting of only two nodes, connect the servers as shown in Figure 4.

![Two-node network configuration](image2)

**Figure 4.** Two-node network configuration
3.2.2 Active Directory
Configure each node to join a Windows Active Directory domain. For the sample deployments in this document, we assume the following network configuration.

![Network Configuration Diagram](image)

**Figure 5.** Storage Spaces Direct network configuration diagram

3.2.3 RDMA networking options
The remote direct memory access (RDMA) feature in HPE network adapters allows lower latency, reduces processor read/write cycles, and improves performance. RDMA can be implemented using either RDMA over Converged Ethernet (RoCE) or iWARP. RoCEv2 implementations require a lossless or lossy network configuration on the switch while iWARP implementations require minimal configuration and are easier to deploy.

3.2.3.1 iWARP networking
For iWARP deployments, set the RDMA Operational Mode in RBSU and Device Manager -> Advanced Properties as shown in Figures 6 and 7.

![iWARP Setting in RBSU](image)

**Figure 6.** iWARP setting in RBSU
3.2.3.2 RoCEv2 networking
For RoCEv2 deployments, configure Virtual LAN (VLAN) and Priority Flow Control (PFC) on the switches and network adapters.

Switch configuration
Use the following procedure for the HPE FlexFabric 5900AF-48XG-4QSFP+ 48-port switch:

1. Use an RJ-45 to RS-232 cable to connect the console port of the switch to the serial port of your PC.

2. Open a terminal emulation program (for example, PuTTY) on your PC.
3. Set your serial connection to 9600bps/8-N-1.
4. The following screen will appear: Press Enter to continue.
5. For each port, create and assign VLAN and Priority Flow Control values. In the following example, we use VLAN 4 and a Priority Flow Control value of 3, but the actual values will depend on your particular network configuration.

```
<HPE>system-view
[HPE]vlan 4
[HPE-vlan4]quit
[HPE]interface Ten-GigabitEthernet 1/0/1
[HPE-Ten-GigabitEthernet1/0/1]port link-type trunk
[HPE-Ten-GigabitEthernet1/0/1]port trunk permit vlan 4
[HPE-Ten-GigabitEthernet1/0/1]priority-flow-control auto
[HPE-Ten-GigabitEthernet1/0/1]priority-flow-control no-drop dot1p 3
[HPE-Ten-GigabitEthernet1/0/1]qos trust dot1p
[HPE-Ten-GigabitEthernet1/0/1]quit
```

6. Repeat Step 5 for each S2D node port interface on the switch (that is, Ten-GigabitEthernet 1/0/1 to Ten-GigabitEthernet 1/0/16).

7. After all ports are configured, use “display vlan” command (in our example, “display vlan 4”) to verify that the required ports are tagged properly.
Use the following commands for the HPE StoreFabric SN2700M 100GbE switch:

```yaml
switch (config) # vlan 4
switch (config) # interface ethernet 1/1 switchport mode trunk
switch (config) # interface ethernet 1/1 switchport trunk allowed-vlan 4

# Repeat this command for ports 1/1 to 1/32
switch (config) # interface ethernet 1/32 switchport mode trunk
switch (config) # interface ethernet 1/32 switchport trunk allowed-vlan 4

switch (config) # interface ethernet 1/1-1/32 traffic-class 3 congestion-control ecn minimum-absolute 150 maximum-absolute 1500
switch (config) # interface ethernet 1/1-1/32 ingress-buffer iPort.pg0 bind switch-priority 6
switch (config) # interface ethernet 1/1-1/32 ingress-buffer iPort.pg3 bind switch-priority 3

switch (config) # interface ethernet 1/1-1/32 ingress-buffer iPort.pg3 map pool iPool1 type lossless reserved 67538 xoff 18432 xon 18432 shared alpha 2
switch (config) # interface ethernet 1/1-1/32 ingress-buffer iPort.pg0 map pool iPool0 type lossy reserved 10240 shared alpha 8

switch (config) # interface ethernet 1/1-1/32 egress-buffer ePort.tc3 map pool ePool1 reserved 1500 shared alpha infinity

switch (config) # interface ethernet 1/1-1/32 traffic-class 6 dcb ets strict
switch (config) # interface ethernet 1/1-1/32 qos trust l3
switch (config) # dcb priority-flow-control enable force
switch (config) # dcb priority-flow-control priority 3 enable

switch (config) # interface ethernet 1/1-1/32 dcb priority-flow-control mode on force
```

**Network adapter configuration**

Configure PFC and create an SMB Direct policy to work with RDMA using PowerShell as follows:

```powershell
# Set QoS policy for SMB-Direct
PS> Set-NetQosOcbxSetting -Willing $false
PS> New-NetQosPolicy "SMB" -SMB -PriorityValue8021Action 3
PS> New-NetQosPolicy "DEFAULT" -Default -PriorityValue8021Action 0

# Enable PFC 3
PS> Enable-NetQosFlowControl -Priority 3
PS> Disable-NetQosFlowControl -Priority 0,1,2,4,5,6,7

# Apply the policy to all adapters
PS> Enable-NetAdapterQos -InterfaceAlias "Embedded FlexibleLOM 1 Port 1", "Embedded FlexibleLOM 1 Port 1 2"
PS> Enable-NetAdapterQos -InterfaceAlias "PCIe Slot 2 Port 1", "PCIe Slot 2 Port 2"

# Configure VLAN ID 4
PS> Set-NetAdapter -Name "Embedded FlexibleLOM 1 Port 1" -VLanID 4
PS> Set-NetAdapter -Name "Embedded FlexibleLOM 1 Port 1 2" -VLanID 4
PS> Set-NetAdapter -Name "PCIe Slot 2 Port 1" -VLanID 4
PS> Set-NetAdapter -Name "PCIe Slot 2 Port 2" -VLanID 4

# Assign 95% bandwidth for SMB-Direct
PS> New-NetQosTrafficClass "SMB" -Priority 3 -BandwidthPercentage 95 -Algorithm ETS
```
Note
“Embedded FlexibleLOM 1 Port x” and “PCIe Slot x Port x” are the friendly names of the RDMA network adapters.

Use the Performance Monitor to verify RDMA is set up correctly. The performance counters RDMA Activity and SMB Client Shares can list relevant objects such as active RDMA connections, inbound bytes per second, and SMB data bytes per second.

Figure 1. Performance Monitor

3.3 Deploying hyperconverged solutions
In a hyperconverged environment, the compute and storage components are located in the same cluster. For this reason, the recommended hardware requirements are higher in order to handle the load. Storage Spaces Direct volumes are accessible from all the nodes under a local directory named \ClusterStorage.

Figure 2. Hyperconverged solution

3.3.1 Sample procedure
The following procedure creates a 3-way mirror volume on a sample 4-node cluster. For more information about different volume types, see the following document: docs.microsoft.com/en-us/windows-server/storage/storage-spaces/plan-volumes. These commands should be executed from one of the cluster nodes using the domain administrator account.

1. Create cluster
   Create a cluster named MYCLUSTER using the four nodes N1, N2, N3, N4, and assign an unused IP address to the cluster.
   PS C:\> New-Cluster –Name MYCLUSTER -Node N1, N2, N3, N4 –NoStorage –StaticAddress 192.168.100.250
   Assign the file share witness as follows:
   PS C:\> Set-ClusterQuorum –FileShareWitness \{{Witness Server Name}}\{Share Name}
2. Enable Storage Spaces Direct

Set the storage system to Storage Spaces Direct mode with the following PowerShell command.

```powershell
PS C:\> Enable-ClusterStorageSpacesDirect
```

![Enable Storage Spaces Direct](image1.png)

3. Create disk volume

Volumes in Storage Spaces Direct provide resiliency, using the following mechanisms: **mirroring** and **parity**. We strongly recommend use of **mirroring** for optimal performance. The following example shows how to create a 1 TB mirror volume on the Storage Spaces Direct pool named “S2D”:

```powershell
PS C:\> New-Volume -StoragePoolFriendlyName "S2D" -FriendlyName MyDisk -FileSystem CSVFS_ReFS -Size 1TB
```

![Create disk volume](image2.png)

This example shows how to create a 1 TB parity volume on Storage Spaces Direct:

```powershell
PS C:\> New-Volume -StoragePoolFriendlyName "S2D" -FriendlyName MyDisk -FileSystem CSVFS_ReFS -Size 1TB
```

![Create disk volume](image3.png)
3.4 Deploying converged solutions

A converged environment separates compute and storage, placing them in different clusters. This allows the compute and storage functions to use independent hardware resources and provide independent scalability. The S2D volumes are accessed through the network using the SMB3 protocol. To understand more about this deployment option, see the following documentation: docs.microsoft.com/en-us/windows-server/storage/storage-spaces/storage-spaces-direct-overview.

![Hyper-V cluster](image)

![SMB3 storage network fabric](image)

![Hyper-V cluster with local storage](image)

Figure 15. Converged solution

4.0 Troubleshooting

4.1 Cluster verification

Use the Test-Cluster cmdlet to verify your system configuration. A full detailed report will identify the areas where the requirements are not met.

PS C:\> Test-Cluster –Node N1, N2, N3, N4 –Include “Storage Spaces Direct”, Inventory, Network, “System Configuration”

![Test-Cluster output](image)

Figure 16. A full detailed report after using the Test-Cluster cmdlet

Use a browser to view the report that is generated in the specified folder location.
4.2 Disk management

In the following report, there were no eligible disks found on Node 3 for S2D. Replace the storage with the appropriate media to fix the problem.

Ensure the proper HPE Smart Array controller driver is installed using the SPP. The Windows inbox storage controller driver will make your SSDs/HDDs not eligible for S2D. The following are some useful tips for disk management:

Disks cannot contain any partitions and should be cleaned using the “Clean” command in DiskPart.

Use the following cmdlet to view the list of disks used by the storage pool:

```
PS C:\> Get-StoragePool –FriendlyName “S2D” | Get-PhysicalDisk
```

Use the following cmdlet to delete a virtual disk:

```
PS C:\> Remove-VirtualDisk –FriendlyName “MyDisk”
```

4.3 Disk rebuild

When replacing failed disks with new ones, verify data reconstruction status using Get-StorageJob as follows.

```
PS C:\> Get-StorageJob
```

4.4 Disk clean

All disks need to be empty before enabling S2D. You can clean all disks using PowerShell scripts. For more information, see the Clean drives section in following documentation: docs.microsoft.com/en-us/windows-server/storage/storage-spaces/hyper-converged-solution-using-storage-spaces-direct#step-3-configure-storage-spaces-direct
4.5 **Cluster management**

Adding a new node to your existing cluster can extend your storage pool size. However, only volumes created after adding the new node will be able to make use of the additional storage space available in the new node. Add a new node using the Get-Cluster cmdlet as follows:

```
PS C:\> Get-Cluster MYCLUSTER | Add-ClusterNode NEWNODE
```

![Cluster management](image)

**Note**

Each cluster may contain up to a maximum of 16 nodes.

4.6 **Network management**

Use the following cmdlets to disable RDMA:

```
PS C:\> Disable-NetAdapterRdma "Embedded FlexibleLOM 1 Port 1"
PS C:\> Disable-NetAdapterRdma "Embedded FlexibleLOM 1 Port 1 2"
PS C:\> Set-NetOffloadGlobalSetting -NetworkDirect Disabled
```

Alternatively, enable RDMA again using:

```
PS C:\> Enable-NetAdapterRdma "Embedded FlexibleLOM 1 Port 1"
PS C:\> Enable-NetAdapterRdma "Embedded FlexibleLOM 1 Port 1 2"
PS C:\> Set-NetOffloadGlobalSetting -NetworkDirect Enabled
```

4.7 **Remove Storage Spaces Direct**

Before you remove a Storage Spaces Direct Cluster, you should remove all virtual disks created in the Storage Spaces Direct storage pool (see Disk Management). Once all virtual disks are removed, you can run the following commands to remove the Storage Spaces Direct Cluster.

```
PS C:\> Disable-ClusterStorageSpacesDirect
PS C:\> Get-Cluster MYCLUSTER | Remove-Cluster
```

Reboot all nodes and clean all used disks.

5.0 **Additional resources**

Microsoft® Storage Spaces Direct

Microsoft Storage Replica

Microsoft Azure Stack HCI home page

HPE Azure Stack HCI home page
Appendix 1: Performance validation method

A1.1 VMFleet for Storage Spaces Direct

In this section, we describe the standardized environment and the procedure used to test our performance validation. Unless explained otherwise, all the tests are based on following criteria:

1. Four-node isolated S2D cluster
2. Three-way mirror volume resilience setup for use by the VM guests
3. VMFleet running 20 VM guests per node. Each VM guest runs one instance of diskspd

VMFleet is part of diskspd and provides performance stress against an S2D deployment. Our validations are carried out in the following manner:

1. Download diskspd from GitHub. The VMFleet files are located in diskspd-master\Frameworks\VMFleet folder
2. Copy the VMFleet folder onto one of the S2D nodes
3. Run following commands:
   - Get-ClusterNode % { New-Volume -StoragePoolFriendlyName S2D* -FriendlyName S_ -FileSystem CSVFS_ReFS -Size 12TB }
   - New-Volume -StoragePoolFriendlyName S2D* -FriendlyName C -Collect -FileSystem CSVFS_ReFS -Size 1TB
4. Go to the VMFleet folder and run ".\install-vmfleet.ps1 -source."
5. Copy DISKSPD.exe into C:\ClusterStorage\Collect\Control\Tools
6. Open Hyper-V Manager and install Windows Server 2019 Core VM for VMFleet. Copy VM file (YourVM.vhdx) to C:\ClusterStorage\Collect
   - For testing purposes, we recommend using a fixed size VHDX (20 GB).
7. To create 20 running VMs, run the following command:
   - \create-vmfleet.ps1 -basevhd "C:\ClusterStorage\collect\YourVM.vhdx" -adminpass xxx -connectpass xxx -connectuser “Administrator”
   - vms 20
   - adminpass and –connectpass are the same password as the VM administrator.
   - Allow some time for 20 VM deployment.
8. The following command is optional for VM configuration:
   - \set-vmfleet.ps1 -ProcessorCount 1 -MemoryStartupBytes 8GB -DynamicMemory $false
9. All VMFleet files are ready; run \\test-vmfleet.ps1 to start all VMs.
   - Run \\test-clusterhealth.ps1 to check for any problems in the cluster
10. To monitor the storage performance, run \\Watch-Cluster.ps1 (Ctrl-C to stop)
11. Run test cases on VMs; the following commands run the test for 600 seconds using seven threads to drive 0%/10%/30% random 64 KiB writes at a depth of 40 overlapped I/Os to a regular file.
   - \\start-sweep.ps1 -b 64 -i 7 -o 40 -w 0 -p r -d 600
   - \\start-sweep.ps1 -b 64 -i 7 -o 40 -w 10 -p r -d 600
   - \\start-sweep.ps1 -b 64 -i 7 -o 40 -w 30 -p r -d 600
12. Log files will be generated when all tests are completed in the VMs.
13. To collect the report files, copy *.xml files from C:\ClusterStorage\Collect\Control\report to a new folder.
   - Copy process-diskspd.ps1 from diskspd-master folder to the folder you created
   - Run process-diskspd.ps1, report files will be collected into single file named result.tsv
14. To stop all VMs, run \\stop-vmfleet.ps1
15. To destroy all VMs, run \\destroy-vmfleet.ps1
A1.2 How to monitor the performance of VMFleet

Performance Monitor provides counters and recorders for monitoring processor utilization, disk read/write bytes, and network activity.

The following steps outline how to monitor specific counters.

1. Open Performance Monitor, browse to **Data Collector Sets** and create a **new Data Collector Set**.

2. Name the collector set and choose **Create manually (Advanced)**. Click **Next**.

---

Figure 20. Create a new Data Collector Set in Performance Monitor

Figure 21. Name the new Data Collector Set
3. Select **Performance counter** and click **Next**

![Create new Data Collector Set](image)

**Figure 22.** Choose the data type

4. Click **add** and select the desired counters. We recommend the following counters:

- **Processor**
- **Cluster CSVFS**

Sample interval can be 5 seconds or longer. Click **Next** and select the location of report.

![Create new Data Collector Set](image)

**Figure 23.** Add performance counter for monitoring
5. The new collector set will be added. Select the data collector and edit its properties. You can set **Maximum samples**

![Figure 24](image)

**Figure 24.** Set maximum samples in DataCollector properties

6. When VMFleet starts running, allow five seconds for it to **warm up**. Then, click **Start** on your Data Collector Sets. If you set **Maximum Samples** already, the collector will stop automatically when the maximum is reached.

![Figure 25](image)

**Figure 25.** Content window of the Data Collector Sets
Technical white paper

Learn more at hpe.com/solutions/azurestackhci