DataON TracSystem S2D-5224 & Windows Server 2016 Storage Spaces Direct Solution with Mellanox Spectrum Switches

Rev 1.0
NOTE:
THIS HARDWARE, SOFTWARE OR TEST SUITE PRODUCT ("PRODUCT(S)"") AND ITS RELATED DOCUMENTATION ARE PROVIDED BY MELLANOX TECHNOLOGIES "AS-IS" WITH ALL FAULTS OF ANY KIND AND SOLELY FOR THE PURPOSE OF AIDING THE CUSTOMER IN TESTING APPLICATIONS THAT USE THE PRODUCTS IN DESIGNATED SOLUTIONS. THE CUSTOMER'S MANUFACTURING TEST ENVIRONMENT HAS NOT MET THE STANDARDS SET BY MELLANOX TECHNOLOGIES TO FULLY QUALIFY THE PRODUCT(S) AND/OR THE SYSTEM USING IT. THEREFORE, MELLANOX TECHNOLOGIES CANNOT AND DOES NOT GUARANTEE OR WARRANT THAT THE PRODUCTS WILL OPERATE WITH THE HIGHEST QUALITY. ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT ARE DISCLAIMED. IN NO EVENT SHALL MELLANOX BE LIABLE TO CUSTOMER OR ANY THIRD PARTIES FOR ANY DIRECT, INDIRECT, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES OF ANY KIND (INCLUDING, BUT NOT LIMITED TO, PAYMENT FOR PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY FROM THE USE OF THE PRODUCT(S) AND RELATED DOCUMENTATION EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.
# Table of Contents

Document Revision History .................................................................................................................. 4  
1  Setup ................................................................................................................................................ 5  
2  Overview .......................................................................................................................................... 6  
3  Configuration ................................................................................................................................... 6  
  3.1  IP Connectivity ........................................................................................................................ 6  
  3.1.1 VLANs ........................................................................................................................ 6  
  3.1.2 Interface .................................................................................................................... 6  
  3.1.3 L3 and VRRP ............................................................................................................ 7  
  3.1.4 IP Interfaces on the Servers ..................................................................................... 8  
  3.1.5 Verification .............................................................................................................. 10  
4  RDMA QoS Configuration ............................................................................................................ 10  
  4.1  Switch Configuration ............................................................................................................. 11  
  4.2  Server Configuration ............................................................................................................. 12  
  4.3  Other Related Commands .................................................................................................... 13  
  4.4  Script ..................................................................................................................................... 14  
  4.5  Verifying RDMA QoS Configuration ..................................................................................... 14  
  4.6  Benchmark Testing (Basic) .................................................................................................... 18  
  4.7  Congestion Control Verification ............................................................................................ 18  
  4.8  PFC Verification .................................................................................................................... 19  
  4.9  Packet Format Validation ..................................................................................................... 19  
5  DataON & Windows Server 2016 Storage Spaces Direct Configuration, Deployment and Testing .................................................................................................................................................. 22  
  5.1  DataON S2D Solution .............................................................................................................. 22  
  5.2  Hardware Configuration and Deployment Tips ........................................................................ 23  
  5.3  Benchmark & Testing Tips ...................................................................................................... 25  
  5.4  Management by MUST™ ...................................................................................................... 27  
  5.5  The DataON Difference ........................................................................................................ 27  
  5.6  About DataON ...................................................................................................................... 28
# Document Revision History

**Table 1: Document Revision History**

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1.0      | September 22, 2017 | Updated to DataON TracSystem S2D-5224L  
Added DataON & Windows Server 2016 Storage Spaces Direct content |
1 Setup

The setup includes two Mellanox Spectrum switches configured with VRRP and four DataON servers equipped with NVMe cards and Mellanox ConnectX®-4 network adapters dual ports for multi-path connectivity.

There are several models of Mellanox Spectrum™ switches, in this example we use the SN2700 32-port 100GbE switches.

- Servers: DataON TracSystem S2D-5224L (one node per S2D-5224 cluster, up to 16)
- Switches: Mellanox Spectrum™ SN2700
- Network adapters: Mellanox ConnectX®-4
- Mellanox Firmware Tools (WinMFT, version 4.7.0 or later) for utilities mentioned later
2 Overview

We configure this setup in 3 phases:

- IP connectivity
- RDMA QoS configuration (PFC/ECN, buffers) on the switches and servers
- Windows server configuration, core switch connectivity and other considerations

In this design, the traffic between the servers will be over RDMA (storage traffic), while traffic towards the core switches will be TCP/Management traffic and not RDMA traffic.

3 Configuration

Before you start, make sure you have the servers installed and powered up as well as the switches.

For first time MLNX-OS® installation, please refer to HowTo Get Started with Mellanox switches.

3.1 IP Connectivity

- Ports 1-28 downlinks to the servers (VLAN interface)
- Ports 29-30 uplinks towards the core switches (router ports)
- Ports 31-32 connected to the other ToR switch (port-channel)

3.1.1 VLANs

We will use the dual ports for the multi-path solution. Each server will be configured with different VLAN on each port.

In this example we will use VLANs 8 and 9:

```
switch (config) # vlan 8-9
switch (config) # vlan 8 name "Storage1"
switch (config) # vlan 9 name "Storage2"
```

3.1.2 Interface

1. Create LAG (port-channel) in trunk mode on ports 31 and 32. This link will be used for VRRP communication between switches.

```
switch (config) # interface port-channel 1
switch (config) # interface port-channel 1 description VRRP Link To other
switch (config) # interface ethernet 1/31 description VRRP Link To other
switch (config) # interface ethernet 1/32 description VRRP Link To other
switch (config) # interface ethernet 1/31 channel-group 1 mode on
switch (config) # interface ethernet 1/32 channel-group 1 mode on
switch (config) # interface port-channel 1 switchport mode trunk
```
NOTE: All VLANs are members of trunk ports by default.

2. Configure links 1-28 (downlinks) towards the servers as trunk.

```plaintext
switch (config) # interface ethernet 1/1 switchport mode trunk
switch (config) # interface ethernet 1/2 switchport mode trunk
...
switch (config) # interface ethernet 1/28 switchport mode trunk
```

NOTE: All VLANs are members of trunk ports by default. The trunk allow only tagged traffic, if untagged traffic is needed (e.g. PXE boot) as well on those ports, set the links to hybrid and configure it to allow all VLANs.

```plaintext
switch (config) # interface Ethernet 1/28 switchport mode hybrid
switch (config) # interface Ethernet 1/28 switchport hybrid allow-vlan all
```

Learn more about switchport on Mellanox switches in HowTo Configure Switch Port Types with MLNX-OS®.

3. Configure the uplink ports as router ports towards the core switches. Set the IP Address and subnet required on this interface.

```plaintext
switch (config) # interface ethernet 1/29 no switchport
switch (config) # interface ethernet 1/29 ip address 10.10.1.1 /24
switch (config) # interface ethernet 1/30 no switchport
switch (config) # interface ethernet 1/30 ip address 10.10.2.1 /24
```

NOTE: In this design, we assume that RDMA traffic will not pass via the core switches.

### 3.1.3 L3 and VRRP

We design the network to have two VLANs (multi-path). Each of the switches will be configured as VRRP master for a different VLAN, so both of the switches will be used (active-active).

1. Enable IP routing, and configure VLAN interface for each VLAN (8, 9).

```plaintext
switch (config) # ip routing vrf default
switch (config) # interface vlan 8
switch (config) # interface vlan 9
switch (config) # interface vlan 8 ip address 192.168.101.2 255.255.255.0
switch (config) # interface vlan 9 ip address 192.168.102.2 255.255.255.0
```

NOTE: Each ToR switch should be configured with different IP addresses, this IP address will be the local IP address of each switch.

ToR 1:

```plaintext
switch (config) # ip routing vrf default
switch (config) # interface vlan 8
switch (config) # interface vlan 9
switch (config) # interface vlan 8 ip address 192.168.101.2 255.255.255.0
switch (config) # interface vlan 9 ip address 192.168.102.2 255.255.255.0
```
ToR 2:

switch (config) # ip routing vrf default
switch (config) # interface vlan 8
switch (config) # interface vlan 9
switch (config) # interface vlan 8 ip address 192.168.101.3 255.255.255.0
switch (config) # interface vlan 9 ip address 192.168.102.3 255.255.255.0

2. Enable the VRRP protocol on the switch and configure virtual IP address for each VLAN. Make sure to design the VRRP master for each VLAN to be a different port (using the priority parameter, the master priority is 255).

ToR 1:

switch (config) # protocol vrrp
switch (config) # interface vlan 8 vrrp 8
switch (config) # interface vlan 8 vrrp 8 address 192.168.101.1
switch (config) # interface vlan 9 vrrp 9
switch (config) # interface vlan 9 vrrp 9 address 192.168.102.1
switch (config) # interface vlan 8 vrrp 8 priority 200
ToR 1 will be the VRRP Slave for this subnet

ToR 2:

switch (config) # protocol vrrp
switch (config) # interface vlan 8 vrrp 8
switch (config) # interface vlan 8 vrrp 8 address 192.168.101.1
switch (config) # interface vlan 9 vrrp 9
switch (config) # interface vlan 9 vrrp 9 address 192.168.102.1
switch (config) # interface vlan 9 vrrp 9 priority 200
ToR 2 will be the VRRP Slave for this subnet

To learn more about the VRRP configuration, see HowTo Configure VRRP on Mellanox Ethernet Switches.

3.1.4 IP Interfaces on the Servers

1. Make sure to install Windows Server 2016 with the latest WinOF-2 driver on the servers. The servers should be equipped with ConnectX®-4 or ConnectX®-5 dual port adapters.

2. Make sure that security features, like firewalls are disabled, so ping can pass.

3. Set the IP addresses on the interfaces:
   a. Configure port 1 on the server connected to ToR 1 with VLAN 8 with the IP range of that VLAN (192.168.101.X)
   b. Configure port 2 on the server connected to ToR 2 with VLAN 9 with the IP range of that VLAN (192.168.102.X)
c. Set the IP address to suit the VRRP virtual address.

![Internet Protocol Version 4 (TCP/IPv4) Properties]

- Obtain an IP address automatically
- Use the following IP address:
  - IP address: 192.168.101.11
  - Subnet mask: 255.255.255.0
  - Default gateway: 192.168.101.1

- Obtain DNS server address automatically
- Use the following DNS server addresses:
  - Preferred DNS server: 
  - Alternate DNS server: 

[Validate settings upon exit]

OK  Cancel

4. Add a route with lower metric from one network to the other network.

For example:

```
C:\> route add 192.168.101.0 192.168.102.1 METRIC 500
C:\> route add 192.168.102.0 192.168.101.1 METRIC 500
```

There are two ways now to reach the 101 network:
• via 192.168.101.11 (locally connected)
• via the other port 192.168.102.1 (the virtual router address)

So if one port is down, the traffic to that network will be sent from the second port.

**NOTE:** That the second route should have higher metric (510) in the example below. Higher metric will be lower priority to be used. We don't want that to be used regularly.

```
<table>
<thead>
<tr>
<th>Active Routes:</th>
<th>Network Destination</th>
<th>Netmask</th>
<th>Gateway</th>
<th>Interface</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0.0.0</td>
<td>0.0.0.0</td>
<td>0.0.0.0</td>
<td>192.168.101.1</td>
<td>192.168.101.11</td>
<td>266</td>
</tr>
<tr>
<td>127.0.0.0</td>
<td>127.0.0.0</td>
<td>255.0.0.0</td>
<td></td>
<td>127.0.0.1</td>
<td>32</td>
</tr>
<tr>
<td>192.168.101.0</td>
<td>192.168.101.0</td>
<td>255.255.255.255</td>
<td>192.168.101.0</td>
<td>192.168.101.1</td>
<td>266</td>
</tr>
<tr>
<td>192.168.102.0</td>
<td>192.168.102.0</td>
<td>255.255.255.255</td>
<td>192.168.102.0</td>
<td>192.168.102.1</td>
<td>266</td>
</tr>
<tr>
<td>224.0.0.0</td>
<td>224.0.0.0</td>
<td>240.0.0.0</td>
<td></td>
<td>224.0.0.1</td>
<td>32</td>
</tr>
<tr>
<td>255.255.255.255</td>
<td>255.255.255.255</td>
<td>255.255.255.255</td>
<td>192.168.101.1</td>
<td>192.168.101.1</td>
<td>266</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Persistent Routes:</th>
<th>Network Address</th>
<th>Netmask</th>
<th>Gateway Address</th>
<th>Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0.0.0</td>
<td>0.0.0.0</td>
<td>0.0.0.0</td>
<td>192.168.101.1</td>
<td>Default</td>
</tr>
</tbody>
</table>
```

### 3.1.5 Verification

Test the L3 connectivity. Make sure ping is running (all to all).

1. Ping between the servers (all to all) make sure that traffic reaches the virtual routers and the local router interfaces on the VLANs.
2. Ping to the core switches and to external servers.
3. Disable a port, verify that the traffic goes via the second port and reach the desired network (high availability).

### 4 RDMA QoS Configuration

There are different ways to Setup the RDMA layer required for the Windows S2D. To learn more about RDMA and RoCE, see RDMA/RoCE Solutions page.

The Recommended Network Configuration Examples for RoCE Deployment will give you a good start with the switch configuration for a few selected profiles.
To understand more about QoS requirement for RDMA, see [Understanding QoS Configuration for RoCE](#).

### 4.1 Switch Configuration

For this example, we will select Profile [Follow Lossless RoCE Configuration for MLNX-OS Switches in DSCP-Based QoS Mode](#) to configure the switches.

- Loss-less network, PFC is enabled on priority 3
- ECN is enabled on the switch for priority 3.
- Trust L3 is configured on the switch ports (classify the priority via DSCP field).
- Buffer pool configuration and priority mapping.
- CNP traffic will pass on DSCP 48.

### NOTE:

In the example below, we configure QoS on all ports. It is not needed to do so for the uplink ports, just for the ports that may carry RDMA traffic.

1. In order to make DCQCN congestion control to work, a user must enable ECN for RoCE traffic that run over traffic class 3:

   ```
   switch (config) # interface ethernet 1/1-1/32 traffic-class 3 congestion-control ecn minimum-absolute 150 maximum-absolute 1500
   ```

   For a fair sharing of switch buffer with other traffic classes it is recommended to configure ECN on all other traffic classes as well.

2. Buffer pool configuration.

   Allocating a buffer pool 0 for lossy traffic and pool 1 for lossless traffic.

   ```
   switch (config) # pool ePool1 direction egress-mc size 16777000 type dynamic
   switch (config) # pool ePool0 direction egress size 5242880 type dynamic
   switch (config) # pool iPool0 direction ingress size 5242880 type dynamic
   switch (config) # pool iPool1 direction ingress size 5242880 type dynamic
   ```

3. Bind interfaces to switch-priority.

   Binding switch priorities 3 and 6 to ingress PG group 3 and 6.

   ```
   switch (config) # interface ethernet 1/1-1/32 ingress-buffer iPort.pg6 bind switch-priority 6
   switch (config) # interface ethernet 1/1-1/32 ingress-buffer iPort.pg3 bind switch-priority 3
   ```

4. Mapping ingress/egress interface to pool configuration.

   Allocating buffer to priority 3 and mapping it to a lossless pool and allocating buffer to priority 6 and mapping it to a lossy pool:

   ```
   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.pg6 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 inf
   ```
5. Setting strict priority to CNPs over traffic class 6.

```
switch (config) # interface ethernet 1/1-1/32 traffic-class 6 dcb ets strict
```

6. Set trust mode L3 (DSCP)

```
switch (config) # interface ethernet 1/1-1/32 qos trust L3
```

7. Enable PFC on priority 3 on all ports:

```
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
```

### 4.2 Server Configuration

The servers should be configured with the following:

- PFC is enabled on DSCP 26
- Windows Storage Spaces Direct RDMA traffic is mapped to egress with priority 3
- ECN is enabled with priority 3
- PFC enabled with priority 3
- CNP traffic will be sent with DSCP 48.
- Trust L3 is used (priority to DSCP mapping)

1. Install Data Center Bridging Windows Feature.

```
PS C:\> Install-WindowsFeature data-center-bridging
```

<table>
<thead>
<tr>
<th>Success</th>
<th>Restart</th>
<th>Needed</th>
<th>Exit Code</th>
<th>Feature Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>True</td>
<td>No</td>
<td>Success</td>
<td>True</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>{Data Center Bridging}</td>
</tr>
</tbody>
</table>

2. Import the PowerShell modules that are required to configure DCB.

```
PS C:\> import-module netqos
PS C:\> import-module dcbqos
PS C:\> import-module netadapter
```

3. Enable QoS on the network adapter

```
PS C:\> Set-NetAdapterQos -Enabled 1 *
```

4. Enable Priority Flow Control (PFC) on the specific priority 3.

```
PS C:\> Enable-NetQosFlowControl -Priority 3
```

5. Locate the registry key for the Mellanox adapter, see HowTo Locate the Windows Registry key for Mellanox Adapters.

In this example, the registry key is:

```
{4d36e972-e325-11ce-bfc1-08002be10318}\0003
```

You will need that for the next configuration commands.
6. Map DSCP to priority for the RDMA traffic. In this example, we are using DSCP 26 to map into a priority 3 (PriorityToDscpMappingTable_3).

```
PS C:\> new-itemProperty -Path HKLM:\SYSTEM\CurrentControlSet\Control\Class"\{4d36e972-e325-11ce-bfc1-08002be10318\}"\0003\ -Name "PriorityToDscpMappingTable_3" -PropertyType "String" -Value "26" -Force
PriorityToDscpMappingTable_3 : 26
```

7. Create a Quality of Service (QoS) policy, and tag each type of traffic with the relevant priority.

In this example we used SMB port 445 with a CoS Value 3.

```
PS c:\> New-NetQosPolicy "SMBDirect" -NetDirectPortMatchCondition 445 -PriorityValue8021Action 3
Name           : SMBDirect
Owner          : Group Policy (Machine)
NetworkProfile : All
Precedence     : 127
JobObject      :
NetDirectPort  : 445
PriorityValue  : 3
```

For testing, you can add another port (e.g. 50000) that will be used later by performance tests (e.g. nd_write_bw).

```
PS c:\> New-NetQosPolicy "SMBDirect" -NetDirectPortMatchCondition 50000 -PriorityValue8021Action 3
Name           : SMBDirect_testRDMA
Owner          : Group Policy (Machine)
NetworkProfile : All
Precedence     : 127
JobObject      :
NetDirectPort  : 50000
PriorityValue  : 3
```

8. Enable ECN on priority 3, and set the DSCP value of the CNP traffic to 48.

```
PS c:\> Mlx5Cmd.exe -Qosconfig -Name RDMA1 -Dcqcn -Enable 3
The command was executed successfully
PS c:\> Mlx5Cmd.exe -Qosconfig -Name RDMA1 -Dcqcn -set -DcqcnCnpDscp 48
The command was executed successfully
PS c:\> Mlx5Cmd.exe -Qosconfig -Name RDMA2 -Dcqcn -Enable 3
The command was executed successfully
PS c:\> Mlx5Cmd.exe -Qosconfig -Name RDMA2 -Dcqcn -set -DcqcnCnpDscp 48
The command was executed successfully
```

9. In Device Manager, disable and re-enable RDMA1 and RDMA2 to make the settings active on the NICs.

### 4.3 Other Related Commands

The following commands are not needed in this procedure as there are VLANs, but in case of RDMA over untagged traffic, it should be used.
1. Do not add an 802.1Q tag to transmitted packets that are assigned an 802.1p priority. Note that they are not assigned a non-zero VLAN ID (for example priority-tagged). The default is 0x0 for DSCP-based PFC set to 0x1.

```
PS C:\> new-itemProperty -Path HKLM:\SYSTEM\CurrentControlSet\Control\Class"{4d36e972-e325-11ce-bfc1-08002be10318}"\0003\ -Name "TxUntagPriorityTag" -PropertyType "String" -Value "1" -Force
```

2. Map all untagged traffic to the lossless receive queue. The default is 0x0 for DSCP-based PFC set to 0x1.

```
PS C:\> new-itemProperty -Path HKLM:\SYSTEM\CurrentControlSet\Control\Class"{4d36e972-e325-11ce-bfc1-08002be10318}"\0003\ -Name "RxUntaggedMapToLossless" -PropertyType "String" -Value "1" -Force
```

### 4.4 Script

This script assumes a dual-port adapter (RDMA1 and RDMA2):

```
Install-WindowsFeature data-center-bridging
import-module netqos
import-module dcbqos
import-module netadapter
Set-NetAdapterQos -Enabled 1 *
Enable-NetQosFlowControl -Priority 3
new-itemProperty -Path HKLM:\SYSTEM\CurrentControlSet\Control\Class"{4d36e972-e325-11ce-bfc1-08002be10318}"\0003\ -Name "PriorityToDscpMappingTable_3" -PropertyType "String" -Value "26" -Force
new-itemProperty -Path HKLM:\SYSTEM\CurrentControlSet\Control\Class"{4d36e972-e325-11ce-bfc1-08002be10318}"\0002\ -Name "PriorityToDscpMappingTable_3" -PropertyType "String" -Value "26" -Force
New-NetQosPolicy "SMBDirect" -NetDirectPortMatchCondition 445 -PriorityValue8021Action 3
New-NetQosPolicy "SMBDirect_testRDMA" -NetDirectPortMatchCondition 50000 -PriorityValue8021Action 3
Mlx5Cmd.exe -Qosconfig -Name RDMA1 -Dcqcn -Enable 3 -set -DcqcnCnpDscp 48
Mlx5Cmd.exe -Qosconfig -Name RDMA2 -Dcqcn -Enable 3 -set -DcqcnCnpDscp 48
```

### 4.5 Verifying RDMA QoS Configuration

1. Verify that PFC is enabled on priority 3 and that NetDirect on the required port and priority (e.g. ports 445, 50000 are mapped to priority 3).

   - Get-NetAdapterQos

```
PS C:\> Get-NetAdapterQos
Name : RDMA1
Enabled : True
Capabilities :
          Hardware   Current
--------     -------
MacSecBypass : NotSupported
DcbxSupport : IEEE    IEEE
NumTCs(Max/ETS/PFC) : 8/8/8   8/8/8
OperationalTrafficClasses : TC TSA Bandwidth Priorities
                            -- ---     ----------
0 ETS      100%     0-7
```

2. Verify that PFC is enabled on priority 3.
   - Get-Net-QosFlowControl

   PS C:\> Get-NetQosFlowControl

<table>
<thead>
<tr>
<th>Priority</th>
<th>Enabled</th>
<th>PolicySet</th>
<th>IfIndex</th>
<th>IfAlias</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>False</td>
<td>Global</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>False</td>
<td>Global</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>False</td>
<td>Global</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>True</td>
<td>Global</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>False</td>
<td>Global</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>False</td>
<td>Global</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>False</td>
<td>Global</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>False</td>
<td>Global</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Verify that NetDirect on the required port and priority. For example, ports 445, 50000 are mapped to priority 3.
   - Get-NetQoSPolicy

   PS C:\Users\Administrator> Get-NetQoSPolicy

   Name : S2D Policy1
   Owner : Group Policy (Machine)
   NetworkProfile : All
   Precedence : 127
   JobObject :
   NetDirectPort : 50000
   PriorityValue : 3

   Name : SMB
   Owner : Group Policy (Machine)
   NetworkProfile : All
   Precedence : 127
   JobObject :
4. Check DCQCN/ECN Configuration via `Mlx5Cmd.exe` command.
   - Check the DCQCN is enabled on priority 3 for RP and NP
   - Verify that the DSCP CNP is mapped to DSCP 48

```
PS C:\Users\Administrator> Mlx5Cmd.exe -Qosconfig -Name RDMA1 -Dcqcn -get
DCQCN RP attributes for adapter "RDMA1":
  DcqcnRPEnablePrio0: 1
  DcqcnRPEnablePrio1: 1
  DcqcnRPEnablePrio2: 1
  DcqcnRPEnablePrio3: 1
  DcqcnRPEnablePrio4: 1
  DcqcnRPEnablePrio5: 1
  DcqcnRPEnablePrio6: 1
  DcqcnRPEnablePrio7: 1
  DcqcnClampTgtRateAfterTimeInc: 1
  DcqcnRpgTgtRate: 300
  DcqcnRpgByteReset: 32767
  DcqcnRpgThreshold: 5
  DcqcnRpgAiRate: 5
  DcqcnRpgHaiRate: 50
  DcqcnAlphaToRateShift: 11
  DcqcnRpgMinDecFac: 50
  DcqcnRpgMinRate: 1
  DcqcnRateToSetOnFirstCnp: 0
  DcqcnDceTcpG: 4
  DcqcnDceTcpRtt: 1
  DcqcnRateReduceMonitorPeriod: 4
  DcqcnInitialAlphaValue: 1023
DCQCN NP attributes for adapter "RDMA1":
  DcqcnNPEnablePrio0: 1
  DcqcnNPEnablePrio1: 1
  DcqcnNPEnablePrio2: 1
  DcqcnNPEnablePrio3: 1
  DcqcnNPEnablePrio4: 1
  DcqcnNPEnablePrio5: 1
  DcqcnNPEnablePrio6: 1
  DcqcnNPEnablePrio7: 1
  DcqcnCnpDscp: 48
  DcqcnCnpPrioMode: 1
  DcqcnCnpS02pPrio: 7
The command was executed successfully
```
5. Check the Priority to DSCP mapping.
   
   - Get the PCI location, for example 138.0.0

   ![Mellanox ConnectX-4 Adapter Properties](image)

6. Get the regKeys configured.
   
   - Verify that DSCP is mapped to 26

```
PS C:\Users\Administrator> Mlx5Cmd.exe -RegKeys -bdf 138.0.0
NIC 1:
  Adapter: Mellanox ConnectX-4 Adapter
  Location (PCI bus, device, function): (138,0,0)
  Registry Key             Value
  Default
    *IPChecksumOffloadIPv4   3     3
    *TCPUDPChecksumOffloadIPv4 3    3
    *TCPUDPChecksumOffloadIPv6 3    3
    *EncapsulatedPacketTaskOffload 1  1
    *EncapsulatedPacketTaskOffloadVngre 1  1
    *EncapsulatedPacketTaskOffloadVxlan 1  1
    *VxlanUDPPortNumber       4789   4789
    *LsoV2IPv4                1     1
    *LsoV2IPv6                1     1
    *TransmitBuffers          2048   2048
    TxIntModerationProfile    1     1
    *RSS                      1     1
    *ReceiveBuffers           512   512
    *NumRssQueues             8     128
    RecvCompletionMethod      1     1
    *RscIPv4                  1     1
    *RscIPv6                  1     1
    RxIntModerationProfile    1     1
    RxIntModeration           2     2
    *VMQ                     1     1
    *VMQVlanFiltering         1     1
    *Sriov                    1     0
    *RssOnHostVPorts          0     0
    *QOS                      1     0
    *FlowControl              3     3
    DcbxMode                  2     2
    PriorityToDscpMappingTable_3 26   3
    *PriorityVLANTag         3     3
```
4.6 Benchmark Testing (Basic)

1. Run RDMA traffic between two ports.

   For example:

   Server:
   ```plaintext
   PS C:\> nd_write_bw -D 10 -S 192.168.101.12 -p 50000
   ```

   Client
   ```plaintext
   PS C:\> nd_write_bw -D 10 -C 192.168.101.12 -p 50000
   ```

2. Open Performance Monitoring tool (perfmon) and add the following counter sets.
   - Mellanox WinOF-2 Congestion Control
   - Mellanox WinOF-2 Port QoS
   - RDMA Activity

3. Check performance.

4.7 Congestion Control Verification

1. Create a synthetic congestion in the network (for example, lower the speed of one port to 10G), open Performance Monitoring (perfmon) tool, and run the benchmark testing.

2. Check the Congestion Control counters are progressing on the Notification Point (NP)—receiver—and the Reaction Point (RP)—sender.

   RP example:

   ![RP Example](image)

   NP example:

   ![NP Example](image)

   If you see these counters, it means that DCQCN is working fine in the network (the switch upon congestion marks the IP ToS ECN bits.)
NOTE: PFC counters (pause counters) are not expected to advance.

### 4.8 PFC Verification

1. Disable ECN on one of the switch ports.

   ```
   switch (config) # no interface ethernet 1/1 traffic-class 3 congestion-control
   ```

2. Run the benchmark test, and verify that the PFC counters are progressing. The Congestion Control counters should not be progressing.

   ```
   Mellanox WinOF-2 Port QoS
   Bytes Received 597,635,394,054.0000
   Bytes Sent 977,339,740.000
   Bytes Total 598,533,981,062.0000
   KBytes Received/Sec 4,786,034.392
   KBytes Sent/Sec 7,313.125
   KBytes Total/Sec 4,792,916.795
   Packets Received 548,199,254.0000
   Packets Received/Sec 4,495,663.005
   Packets Sent 13,817,973.000
   Packets Sent/Sec 106,944.928
   Packets Total 561,944,442.0000
   Packets Total/Sec 4,601,759.696
   Rcv Pause Duration 0.000
   Rcv Pause Frames 0.000
   Sent Pause Duration 8,292.000
   Sent Pause Frames 2,754.000
   ```

3. Enable ECN back on the switch.

   ```
   switch (config) # interface ethernet 1/1 traffic-class 3 congestion-control ecn minimum-absolute 150 maximum-absolute 1500
   ```

### 4.9 Packet Format Validation

1. Capture RDMA traffic on one of the servers, use `Mlx5Cmd.exe` for that. For example:

   ```
   PS C:\> Mlx5Cmd.exe -Sniffer -name RDMA1 -start -filename testing_rdma.pcap
   ```

   See also [How To Capture RDMA traffic on mlx5 driver using mlx5cmd (Windows)](Link).

2. Run benchmark test.

3. Open the file in Wireshark.
4. Verify that the RDMA traffic is sent with DSCP 26 (as configured).

- DSCP 26
- ECN is not 00
5. Verify that the CNP traffic is send with DSCP 48 (as configured)

- DSCP 48
- RDMA OpCode is 0x81

Read more on CNP packet format in RoCEv2 CNP Packet Format Example.
5 DataON & Windows Server 2016 Storage Spaces Direct Configuration, Deployment and Testing

In this sample deployment, we use a DataON TracSystem S2D-5224L Hyper-Converged Infrastructure. The S2D-5224L is built to optimize the full stack of Microsoft Server 2016 Storage Spaces Direct in a hyper-converged platform. It is designed 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.

The TracSystem S2D-5224L utilizes Mellanox 40/100GbE switches that support low-latency RDMA networking for a loss-less network with no packet loss.

- **Hyper-V virtualization** – supports more than 40 Hyper-V virtual machines per node
- **Storage and networking with SMB3 over RDMA** – increases CPU efficiency while delivering the highest throughput and lowest latency.
- **Hyper-converged scalability** – delivers incremental compute, networking and storage resources while providing near-linear scalability. The HCI cluster can also be expanded via 12GB/s SAS JBODs.
- **Managed by MUST** – DataON’s exclusive MUST software provides visibility, monitoring and management for Windows Server 2016 environments.

5.1 DataON S2D Solution

The following is DataON TracSystem S2D-5224L HCI for Windows Server Storage Spaces Direct configuration with 4 nodes and Mellanox network fabric optimized for IOPS & performance for Windows Server 2016 Storage Spaces Direct.

<table>
<thead>
<tr>
<th>Form Factor</th>
<th>2U 24-bay 2.5”</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appliance Node</td>
<td>2U Server cluster - 4 Nodes</td>
</tr>
<tr>
<td>Windows Server</td>
<td>Intel® Xeon® Scalable Processor with Intel C620 Chipsets</td>
</tr>
<tr>
<td>2016</td>
<td>24x Intel SSD Data Center Family with NVMe Intel Optane SSDs</td>
</tr>
<tr>
<td>Hyper-Converged</td>
<td>Hyper-V Deployment</td>
</tr>
<tr>
<td>Infrastructure</td>
<td>100-300 VMs (scale to 224 physical cores and 24 DIMMs per CPU)</td>
</tr>
<tr>
<td>Storage Pool</td>
<td>20-150TB (3-way mirror [33%] or RS 2+2 [47% efficiency] MRV erasure coding).</td>
</tr>
<tr>
<td>Allocated Capacity</td>
<td>3M IOPS (100% read); 1.5M IOPS (70/30 read-write)</td>
</tr>
<tr>
<td>Performance</td>
<td>SMB3 over RDMA; 40/100G RDMA NIC</td>
</tr>
<tr>
<td>Networking Fabric</td>
<td>24x DDR4 DRIMMs per node</td>
</tr>
<tr>
<td>Memory Slot</td>
<td>7x PCIe 3.0 x8 per node</td>
</tr>
<tr>
<td>Expansion Slot</td>
<td>DataON MUST visibility, monitoring and management tool</td>
</tr>
</tbody>
</table>
DataON S2D Test Configuration and Result – (4) x DataON 5224 Node Cluster with Mellanox Network Fabric

**Hardware Details per Node**

<table>
<thead>
<tr>
<th>CPU</th>
<th>Intel Gold 6148 2.4GHz x 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td>384 GB</td>
</tr>
<tr>
<td>Cache</td>
<td>INTEL Optane P4800X 375GB NVMe SSD x 2</td>
</tr>
<tr>
<td>Data</td>
<td>INTEL DC P4500 4TB NVMe SSD x 16</td>
</tr>
</tbody>
</table>

*Figure 1 - NVME and SSD (3-way Mirror) Configuration*

<table>
<thead>
<tr>
<th>Volume</th>
<th>Filesystem Capacity used</th>
<th>Resiliency</th>
<th>Size (Mirror)</th>
<th>Size (Parity)</th>
<th>Footprint Efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>arizona-n1 RepS</td>
<td>18TB</td>
<td>3%</td>
<td>3-way Mirror</td>
<td>18TB</td>
<td>0</td>
</tr>
<tr>
<td>arizona-n2 RepS</td>
<td>18TB</td>
<td>3%</td>
<td>3-way Mirror</td>
<td>18TB</td>
<td>0</td>
</tr>
<tr>
<td>arizona-n3 RepS</td>
<td>18TB</td>
<td>3%</td>
<td>3-way Mirror</td>
<td>18TB</td>
<td>0</td>
</tr>
<tr>
<td>arizona-n4 RepS</td>
<td>18TB</td>
<td>3%</td>
<td>3-way Mirror</td>
<td>18TB</td>
<td>0</td>
</tr>
<tr>
<td>Collector</td>
<td>96Gb</td>
<td>18%</td>
<td>3-way Mirror</td>
<td>56Gb</td>
<td>0</td>
</tr>
</tbody>
</table>

36 Virtual Machines created on each (144 Virtual Machines in total).

VMFleet results:

Block size 4Kb, 8 Threads, 8 Outstanding I/O, (100% Read / 0% write), Random

<table>
<thead>
<tr>
<th>CSV FS</th>
<th>IOPS</th>
<th>Reads</th>
<th>Writes</th>
<th>Bn (MB/s)</th>
<th>Read</th>
<th>Write</th>
<th>Read Lat (ms)</th>
<th>Write Lat (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>arizona-n1</td>
<td>748,841</td>
<td>748,801</td>
<td>39</td>
<td>3,086</td>
<td>3,086</td>
<td>0.318</td>
<td>1.652</td>
<td></td>
</tr>
<tr>
<td>arizona-n2</td>
<td>774,786</td>
<td>774,735</td>
<td>41</td>
<td>3,176</td>
<td>3,174</td>
<td>0.137</td>
<td>1.260</td>
<td></td>
</tr>
<tr>
<td>arizona-n3</td>
<td>777,832</td>
<td>777,801</td>
<td>33</td>
<td>3,179</td>
<td>3,178</td>
<td>0.144</td>
<td>2.203</td>
<td></td>
</tr>
<tr>
<td>arizona-n4</td>
<td>777,834</td>
<td>777,801</td>
<td>33</td>
<td>3,179</td>
<td>3,178</td>
<td>0.144</td>
<td>2.203</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sys CPU</th>
<th>%</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>arizona-n1</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arizona-n2</td>
<td>93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arizona-n3</td>
<td>99</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>arizona-n4</td>
<td>77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**5.2 Hardware Configuration and Deployment Tips**

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

- **Make sure your system is certified for** [Microsoft Server](https://www.microsoft.com), [Microsoft Software-Defined Data Center](https://www.microsoft.com/software-defined-data-center), and [Windows Server Software-Defined](https://www.microsoft.com/windows-server/software-defined), for the most optimized infrastructure, operational visibility and reliability for their Windows Server 2016 based software-defined data center.

- **Check that all the S2D hardware components are certified for** [Microsoft Windows Server 2016](https://www.microsoft.com/windows-server/software-defined).

- **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.

- **Make sure you have enough memory for each node/system because it will impact application memory allocation to application workloads such as SQL**.

- **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).**

- **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).

- 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).
- Deploy SMB3 over RDMA networking with DCB enabled switches.

<table>
<thead>
<tr>
<th>SMB3 over RDMA Networking Fabric</th>
</tr>
</thead>
<tbody>
<tr>
<td>With S2D storage, use SMB3 multi-channel RDMA networking for consistent performance and low latency.</td>
</tr>
<tr>
<td>Make sure your DCB switch supports RDMA, as well as supports priority flow control (PFC).</td>
</tr>
<tr>
<td>Make sure you setup your DCB 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.</td>
</tr>
<tr>
<td>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</td>
</tr>
<tr>
<td>For best performance, use a 40G or 100G end-to-end network</td>
</tr>
<tr>
<td>Using a 40GbE to 10GbE splitter is not advised.</td>
</tr>
<tr>
<td>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).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DataON S2D Storage Setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
</tr>
<tr>
<td>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.</td>
</tr>
<tr>
<td>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.</td>
</tr>
<tr>
<td>Make sure you use the DataON S2D platform checklist to ensure proper system configuration.</td>
</tr>
<tr>
<td>Make sure you have our customized deployment guide and S2D installation PowerShell script for your S2D deployment with correct IP for the infrastructure.</td>
</tr>
</tbody>
</table>

With any DataON S2D appliance, we 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.

5.3 Benchmark & Testing Tips

<table>
<thead>
<tr>
<th>DataON S2D Testing &amp; Validation</th>
</tr>
</thead>
<tbody>
<tr>
<td>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).</td>
</tr>
<tr>
<td>Use VM Fleet for performance tuning</td>
</tr>
</tbody>
</table>
**DataON MUST** is the ideal tool for real-time diagnostics

**On the servers using Storage Spaces Direct, open the Resource Monitor**
(go to Task Manager and click on Resource Monitor).

**Setup steps**
- Open Task Manager
- Click Resource Monitor at the bottom of the screen
- Click Disk tab at the top of the screen
- Expand Disk Activity
- Click Total (B/sec) to sort from highest to lowest

**Testing**
- See latency in the “Response Time” column for the processes that are the highest total bytes/sec AND “normal” I/O priority.

**Real World Results**
- Before, Youth Villages was seeing 30-150ms consistently for latency on tempdb, normal log files and database.
- After implementing Storage Spaces Direct, Youth Villages is consistently seeing 1ms latency.
- 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.

<table>
<thead>
<tr>
<th>Measure Latency From the Application Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DataON MUST</strong> is the ideal tool for real-time diagnostics</td>
</tr>
<tr>
<td><strong>On the servers using Storage Spaces Direct, open the Resource Monitor</strong> (go to Task Manager and click on Resource Monitor).</td>
</tr>
<tr>
<td><strong>Setup steps</strong></td>
</tr>
<tr>
<td>- Open Task Manager</td>
</tr>
<tr>
<td>- Click Resource Monitor at the bottom of the screen</td>
</tr>
<tr>
<td>- Click Disk tab at the top of the screen</td>
</tr>
<tr>
<td>- Expand Disk Activity</td>
</tr>
<tr>
<td>- Click Total (B/sec) to sort from highest to lowest</td>
</tr>
<tr>
<td><strong>Testing</strong></td>
</tr>
<tr>
<td>- See latency in the “Response Time” column for the processes that are the highest total bytes/sec AND “normal” I/O priority.</td>
</tr>
<tr>
<td><strong>Real World Results</strong></td>
</tr>
<tr>
<td>- Before, Youth Villages was seeing 30-150ms consistently for latency on tempdb, normal log files and database.</td>
</tr>
<tr>
<td>- After implementing Storage Spaces Direct, Youth Villages is consistently seeing 1ms latency.</td>
</tr>
<tr>
<td>- 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.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Measure Performance Outside of the Application, from the Server Perspective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DataON MUST</strong> is the ideal tool for real-time diagnostics</td>
</tr>
<tr>
<td><strong>On the servers using Storage Spaces Direct, open the Resource Monitor</strong> (go to Task Manager and click on Resource Monitor).</td>
</tr>
<tr>
<td><strong>Setup steps</strong></td>
</tr>
<tr>
<td>- Open Task Manager</td>
</tr>
<tr>
<td>- Click Resource Monitor at the bottom of the screen</td>
</tr>
<tr>
<td>- Click Disk tab at the top of the screen</td>
</tr>
<tr>
<td>- Expand Disk Activity</td>
</tr>
<tr>
<td>- Click Total (B/sec) to sort from highest to lowest</td>
</tr>
<tr>
<td><strong>Testing</strong></td>
</tr>
<tr>
<td>- See latency in the “Response Time” column for the processes that are the highest total bytes/sec AND “normal” I/O priority.</td>
</tr>
<tr>
<td><strong>Real World Results</strong></td>
</tr>
<tr>
<td>- Before, Youth Villages was seeing 30-150ms consistently for latency on tempdb, normal log files and database.</td>
</tr>
<tr>
<td>- After implementing Storage Spaces Direct, Youth Villages is consistently seeing 1ms latency.</td>
</tr>
<tr>
<td>- 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.</td>
</tr>
</tbody>
</table>

**Download and extract diskspd (current version is 2.0.17)**

**Open an elevated command prompt:**
- run D:\Diskspd-v2.0.17\amd64fre\diskspd -b64k -d15 -h -L -o8 -t8 -r -w100 -c50M T:io.dat

**Where:**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>d:</td>
<td>Drive installed the application diskspd is installed.</td>
</tr>
<tr>
<td>-b64k</td>
<td>Testing with 64K blocks (use only numbers divisible by 4).</td>
</tr>
<tr>
<td>-d15</td>
<td>Test for 15 seconds.</td>
</tr>
<tr>
<td>-h</td>
<td>Disable software caching and set write-through I/O (to force the storage to use the disks and not just pull from cache).</td>
</tr>
<tr>
<td>-L</td>
<td>Measure latency statistics.</td>
</tr>
<tr>
<td>-o8</td>
<td>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.</td>
</tr>
<tr>
<td>-t8</td>
<td>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.</td>
</tr>
<tr>
<td>-r</td>
<td>Specifies random I/O.</td>
</tr>
</tbody>
</table>
-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:i.o.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.

### 5.4 Management by MUST™

All DataON TracSystem solutions are pre-configured with DataON MUST (Management Utility Software Tool) infrastructure visibility, monitoring and management software. Fully integrated with the Windows Storage Health Service API (SM-API), it provides advanced cluster monitoring, performance metrics, system health statistics, and automated system alerts for Windows Server 2016 Storage Spaces Direct.

MUST delivers SAN-like storage monitoring features through a single pane of glass, providing real-time dashboard level metrics for IOPS, latency, throughput on cluster nodes and volumes. With system alerts based on Windows Health Service faults and SAN-like call home services, systems administrators can be automatically notified of hardware failures, configuration issues and resource saturation.

DataON is the first to market with a tool that provides visibility, monitoring and management of your Windows Server 2016 deployments.

### 5.5 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 experts know how to design, deploy and support Windows Server storage and will work with you to performance tune your workloads with benchmarks. DataON solutions are:

- Customer-proven with over 600 enterprise deployments and greater than 100PB of DataON Storage Spaces Direct storage deployments.
- Optimized by our team of Microsoft experts to ensure successful deployments into your IT environment, tuned to your workloads.
- Certified for the Windows Server Software-Defined (WSSD) Program.
DataON has a proud history of supporting Windows Server environments, including:

- The FIRST certified enterprise JBODs for Windows Server 2012 R2.
- The FIRST Cluster-in-a-Box (CiB) appliances for small business and enterprise deployments with Hyper-V support

5.6 About DataON

DataON is the industry-leading provider of hyper-converged cluster appliances (HCCA) and storage systems optimized for Microsoft Windows Server environments. Our solutions are built with the single purpose of rapidly and seamlessly deploying Microsoft applications, virtualization, data protection and hybrid cloud services. Our company is exclusively focused on customers who have made the “Microsoft choice” and we provide the ultimate platform for the Microsoft software-defined data center (SDDC). DataON is a division of Area Electronics. For more information, go to www.dataonsstorage.com or call +1 (714) 441-8820.