Implementing SQL Server 2016 with Microsoft Storage Spaces Direct on Dell EMC PowerEdge R730xd

Design Guide

Dell EMC Engineering
October 2017

Table of contents

1 Introduction 5
1.1 Scope 5
1.2 Audience 5
2 Solution components 6
2.1 Dell EMC hardware components 6
2.1.1 Dell EMC S2D Ready Bundle 6
2.1.2 Dell EMC Networking components 7
2.2 Microsoft software components 8
2.2.1 Windows Server 2016 Datacenter Edition 8
2.2.2 Storage Spaces Direct 8
2.2.3 SQL Server 2016 8
3 Solution architecture 9
3.1 Server and storage architecture 9
3.2 Network architecture 10
4 Design principles 11
4.1 High availability 11
4.1.1 Server high availability 11
4.1.2 Networking high availability 11
4.1.3 Storage resiliency 11
5 Solution design 12
6 Solution configuration and best practices 13
6.1 Server configuration 13
6.2 Storage configuration 13
6.2.1 Drive selection 13
6.2.2 Volume configuration options 13
6.2.3 Managing cache for SQL Server 14
6.3 Windows Server, Hyper-V, and SQL Server configuration 14
6.3.1 Windows OS configuration 14
6.3.2 Microsoft Hyper-V 2016 15
6.3.3 Microsoft SQL Server 2016 15
7 Sample implementation 17
7.1 OLTP workload solution 17

8 References 19

1 Introduction

Dell EMC Microsoft Storage Spaces Direct Ready Nodes (S2D RN) are pre-configured with certified components. At the base of their deployment is the versatile Dell EMC PowerEdge R730xd server. The architecture follows a modular approach treating each R730xd server as a building block. This helps deploy scalable and high-performance infrastructure.

Windows Storage Spaces Direct (S2D) is a software-defined storage offering, part of Windows Server 2016. S2D can have either hyper-converged or converged implementations. This document focuses on SQL Server performance and scaling on S2D hyper-converged implementations.

1.1 Scope

The scope of this guide is to outline the best practices for designing and implementing Microsoft SQL Server 2016 on a hyper-converged infrastructure based on PowerEdge R730xd servers.

1.2 Audience

The target audiences for this document are database administrators, system administrators, storage administrators and architects who design and maintain database infrastructures. Readers should have some knowledge of Microsoft Windows Server, Storage Spaces Direct, Microsoft SQL Server, Dell EMC PowerEdge servers, and Dell EMC Networking products.

2 Solution components

The components used in this Dell EMC Ready Bundle for Microsoft SQL Server are categorized into:

- Dell EMC hardware components
- Microsoft software components

2.1 Dell EMC hardware components

Dell EMC offers a wide range of enterprise products and solutions, including servers, storage, networking, software, and services. The products and solutions are designed with reliability and scalability. They are engineered to handle the most demanding business applications and workloads such as ERP, database, BI, and HPC.

2.1.1 Dell EMC S2D Ready Bundle

Dell EMC offers Ready Nodes for S2D, which are tested and validated configurations, to provide the best possible performance. The Ready Node described in this guide is a validated, pre-configured building block for a software-defined storage solution with S2D.

The Dell EMC Ready Node for S2D consists of pre-configured R730xd servers. PowerEdge R730xd is a versatile server supporting a broad range of deployment options in terms of NICs, drives, and processor capabilities.

2.1.1.1 PowerEdge R730xd

PowerEdge R730xd, a Dell EMC 13th generation PowerEdge server, is optimized for HCI deployments. With two sockets and a wide range of processor options, the R730xd has the capability to fulfil the computing needs of the most demanding workloads.

This 2U rack mounted server provides high storage density with up to 16 x 3.5” drives: 12 x 3.5” (front bay) and 4 x 3.5” (internal bay) options. In addition, the 2 x 2.5” flex bay drive slots support OS deployment needs for HCI. PowerEdge R730xd can also be configured by using a 2.5’’ chassis form factor that supports up to 24 drives in the chassis.

The different HDD, SSD, and NVMe drive options available with the PowerEdge R730xd server enable multiple hybrid and all-flash configurations for S2D on Dell EMC Ready Nodes.

PowerEdge R730xd provides the storage density and compute power required to maximize the benefits of S2D and make the best use of the advanced features in Windows Server 2016.

The following table shows the technical specifications of Dell EMC PowerEdge R730xd in detail:

Table 1 Dell EMC PowerEdge server technical specifications

Component	Specification
Server	R730xd
Form factor	2U
Processor	Intel Xeon processor E5-2600 v4 product family
Processor sockets	2
Chipset	Intel C610 series chipset
Memory	Up to 1.5 TB (24 DIMM slots): 4 GB/8 GB/16 GB/32 GB/64 GB DDR4 up to 2400 MT/s
I/O Slots	• 7 PCIe slots and 1 dedicated PERC slot: • 3 x 8 half-length, low-profile • 1 x 16 full-length, full-height • 3 x 8 full-length, full-height
RAID Controller	• Internal controllers: PERC S130 (SW RAID), PERC H330, PERC H730, PERC H730P • External HBAs (RAID): PERC H830 • External HBAs (non-RAID): 12 Gbps SAS HBA
Network Interface	1 x QLogic 57800 2x10 Gb DA/SFP+ + 2 x 1 Gb BT Network Daughter Card (NDC)
	1 x QLogic 57810 Dual Port 10 Gb DA/SFP+ Converged Network Adapter

2.1.2 Dell EMC Networking components

Dell EMC provides networking solutions to enable future-ready enterprises. These solutions, based on open standards, free customers from outdated, proprietary approaches. The networking switches selected for the S2D RN have been carefully chosen from Dell EMC's wide networking portfolio.

2.1.2.1 Dell EMC Networking S4048-ON

Dell EMC Networking S4048-ON is a 10 GbE, L2 and L3 capable network switch for storage, cluster, and client traffic. Some of the features supported by S4048-ON are:

- Routable RDMA over Converged Ethernet (RoCE)
- Converged network support for Data Center Bridging (DCB), with Priority Flow Control (802.1Qbb) and Enhanced Transmission Selection (802.1Qaz)

These features make S4048-ON the preferred choice for networking in an S2D deployment.

2.1.2.2 Dell EMC Networking S3048-ON

Dell EMC Networking S3048-ON is a 1000 BASE-T, L2 and L3 capable switch that provides 48 ports supporting 10 MbE/100 MbE/1 GbE and four 10 GbE SFP+ uplinks. In this RN configuration, S3048-ON is deployed to support the Out Of Band (OOB) connectivity between the PowerEdge R730xd servers and the Dell Networking S4048-ON switches.

2.2 Microsoft software components

The validated solution discussed in this guide uses Microsoft's software offerings—Windows Server 2016 and SQL Server 2016.

2.2.1 Windows Server 2016 Datacenter Edition

Windows Server 2016 is Microsoft's new cloud-ready operating system that provides enhanced security, built-in containers and support for new software defined capabilities for modern data centers. The following are some of the key features of Windows Server 2016:

- ReFS (Resilient File System)—Enables faster placement of VMs on the file system
- Software defined networking—Includes enhanced policies to control both physical and virtual networks
- Hyper-V—Supports nested virtualization

For a complete list of new features of Windows Server 2016, refer to Microsoft article What's New in Windows Server 2016

2.2.2 Storage Spaces Direct

Storage Spaces Direct, which is a part of Windows Server 2016, offers software-defined storage capabilities for disk drives directly attached to the R730xd. It provides pooling and resiliency features for storage. Also, it delivers high performance which can be utilized for IO intensive workloads such as SQL Server.

2.2.3 SQL Server 2016

SQL Server 2016 brings industry leading OLTP capabilities, new encryption features, greater support for In- Memory databases, and an all new end-to-end BI solution. The following are some of the key features of SQL Server 2016:

- Memory optimized tables—Provide enhanced compute resource utilization, optimized query plans, and extended functionality for In-Memory OLAP tables
- Mobile BI platform—Captures insights from online or offline data
- MAXDOP, Parameter Sniffing, Hotfixes—These settings can now be configured at a database level
- R Services—This feature can now be integrated directly into the SQL Server database
- Temporal tables—Provide historical view of tables and their values
- Query Store—Provides comparison of different queries over time

For more details about Microsoft SQL Server 2016 and its complete list of features, see the Microsoft article, What's New in SQL Server 2016

3 Solution architecture

The following sections describe the design principles of Dell EMC Ready Bundle for Microsoft SQL Server 2016 on S2D, and explain how the infrastructure sub-systems are architected. To illustrate the design principles, an implementation example for OLTP workload is given in section 8. Figure 1 illustrates a solution architecture for VMs deployed on a HCI infrastructure.

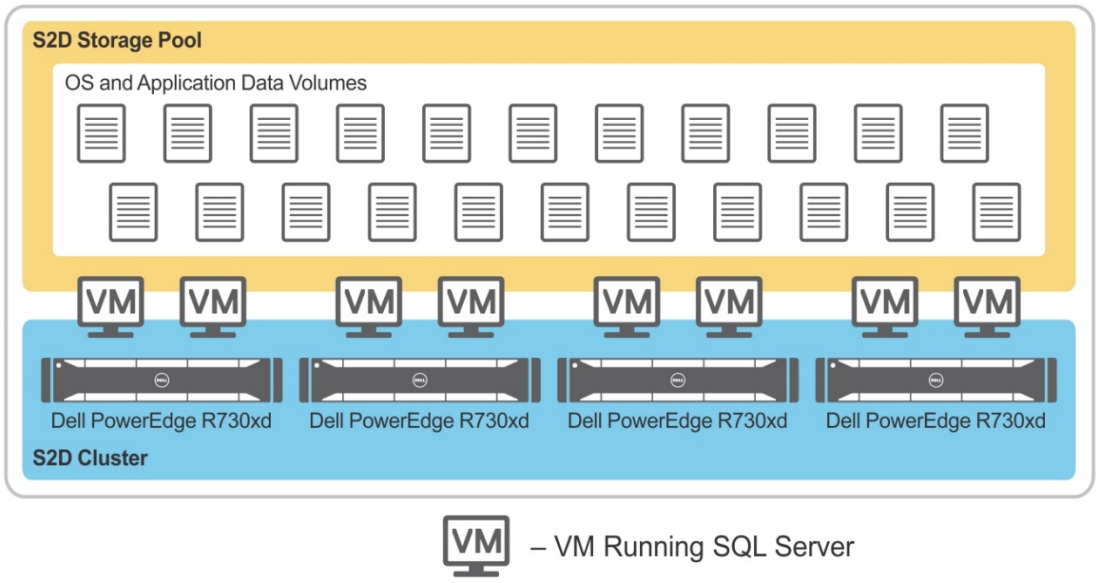

Figure 1 Solution architecture

3.1 Server and storage architecture

In an HCI infrastructure, the server is central for both the compute and storage resources. Hence, it is vital to carefully evaluate the specifications of the components within a server. The processor chosen for the server should be able to handle the storage overhead associated with HCI and the compute requirement for workloads that the customers intend to run on the servers. By having pre-designed, certified components that complement each other, customers can fully utilize the benefits of HCI with the lowest possible data center footprint.

The storage components utilized in S2D-RN include the DAS drives, storage controller, and the RDMA network. Each of the components is explored below:

- The DAS drives available within the R730xd offer a wide range for selection of drives for the S2D storage pool. These include SSD, HDD, and NVMe drives that are certified for hyper-converged workloads.
- The fabric to connect various storage drives across servers is the RDMA over Ethernet. Data-Center Bridging (DCB) concepts such as Priority-based Flow Control (PFC) and Enhanced Transmission Selection (ETS) are leveraged. These enable fast, optimized, and reliable transport for storage traffic.

3.2 Network architecture

The Dell EMC S2D Ready Node utilizes the Mellanox ConnectX-3 Pro and the Dell EMC Networking S4048-ON to provide a hyper-converged network for storage, application, and cluster traffic. DCB concepts such as PFC and ETS with appropriate parameters are used to achieve high performance. Figure 2 shows the network connections for an n-node S2D Ready Node setup.

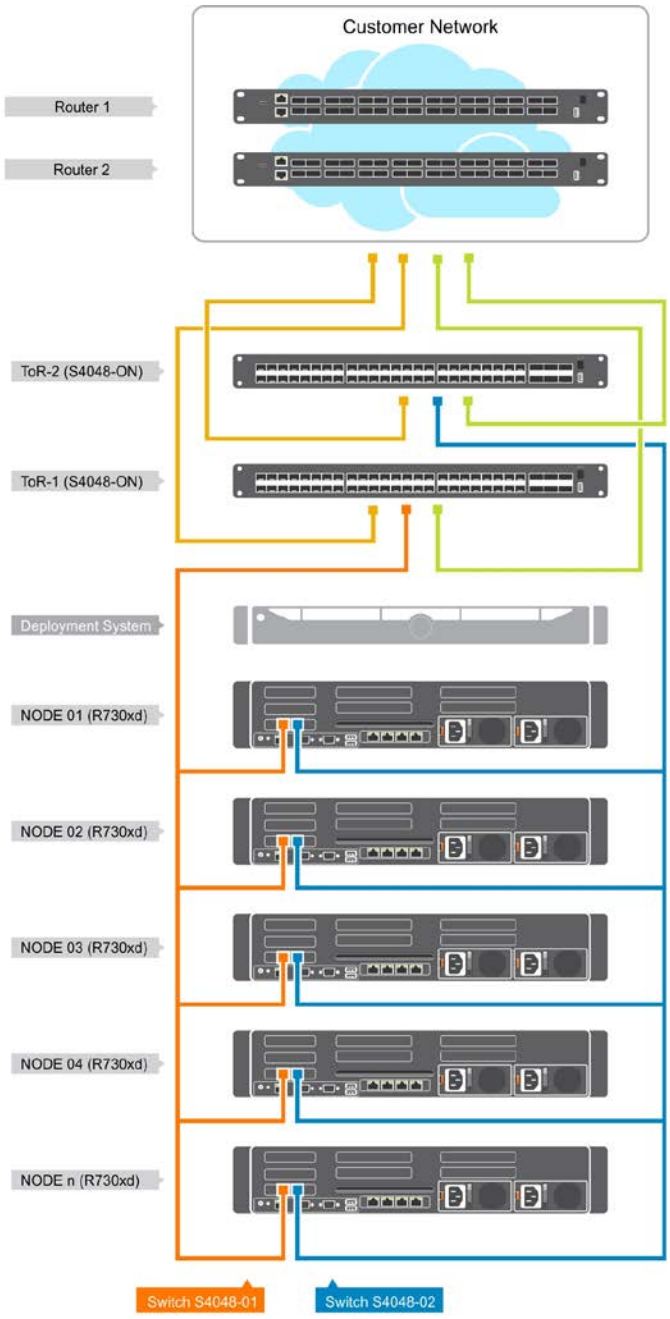

Figure 2 Network architecture

4 Design principles

Databases are mission critical and invaluable in terms of generating insights from the available data. Therefore, when designing an OLTP or OLAP database solution, high-availability (HA) and Service Level Agreement (SLA) are essential. Customers today also demand fast response time and optimized hardware resource utilization.

4.1 High availability

When configured properly, S2D has the capability to address the HA needs of today's demanding data centers. Dell EMC S2D Ready Bundle for SQL Server 2016 has HA through failover clustering. This provides HA and Disaster Recovery (DR) capabilities. The design is based on the cluster failover concept.

4.1.1 Server high availability

S2D relies on multiple servers within a cluster to provide HA in case of failures at the node level.

4.1.2 Networking high availability

The S2D Ready Bundle for SQL Server uses multiple switches, Switch Embedded Team (SET), and Virtual Link Trunk (VLT) to provide high availability for networking components on both logical and physical layers. For more information, see Dell EMC Microsoft Storage Spaces Direct Ready Nodes Deployment Guide.

4.1.3 Storage resilience

S2D provides mirroring and parity capabilities for volumes deployed on the S2D storage pool. These provide resilience based on the type of volume selected, and also provide varying levels of performance. The Dell EMC S2D Ready Node provides four fault domains that allow the creation of all four types of volumes—two-way mirror, three-way mirror, parity, or mixed. For more information about the types of volumes available in S2D, see Fault tolerance and storage efficiency in Storage Spaces Direct.

5 Solution design

Dell Ready Bundle for SQL Server 2016 offers a pre-architected and validated solution for a given size and use case. The hardware components supported as part of this solution offering are Dell EMC PowerEdge R730xd, Dell EMC Networking 4048-ON, and Dell EMC Networking 3038-ON.

The following diagram illustrates the solution design.

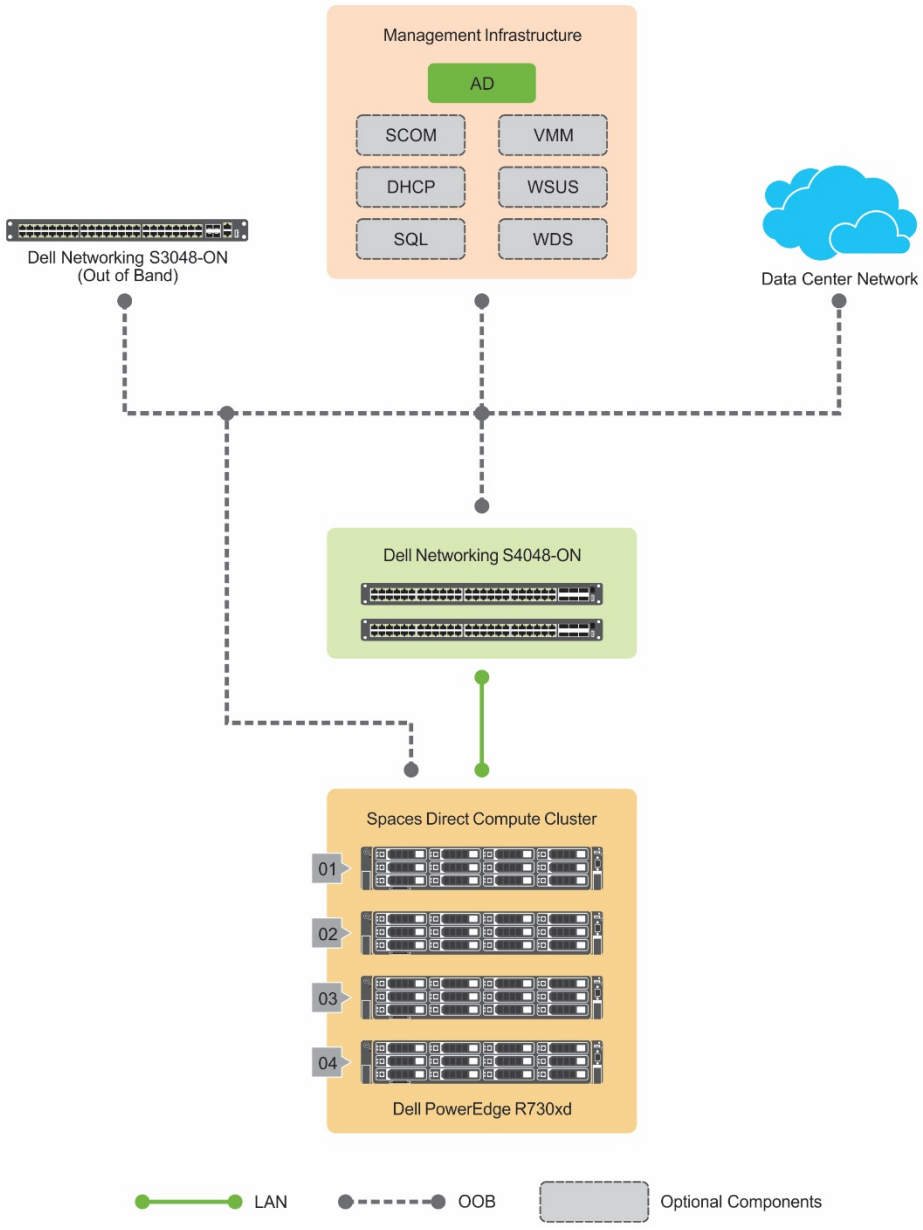

Figure 3 Solution Design

6 Solution configuration and best practices

This section covers the solution configuration and Dell EMC recommended best practices.

6.1 Server configuration

S2D RN servers are shipped with various settings preset as selected while ordering. These settings include Performance Plan, Hyper-Threading, and so on. The best practices recommended by Dell EMC for SQL Server workloads are listed below:

- Use High Performance Power Plan
- Enable Hyper-Threading

6.2 Storage configuration

Multiple factors need to be addressed while designing storage for a hyper-converged infrastructure. They include drive selection, volume configuration, and selecting cache options.

6.2.1 Drive selection

In an S2D setup, up to three types of drives can be utilized simultaneously. The fastest media is selected for caching. Dell EMC S2D Ready Node provides various options for performance and sizes of drives. The fastest media available is NVMe. The least expensive media that can be used to maximize capacity is HDDs. Multi-use SSDs are in the middle of the spectrum, balancing cost and capacity. Using volume sizing guidelines, approximate sizes of databases and expected growth, the capacity and number of drives required can be determined. The caching media should be selected to supply the required amount of IOPs for the database. To summarize, the following factors should be kept in mind while selecting drives:

- Type of the workload
- Expected IOPS, latency, and throughput
- Database size and expected growth

6.2.2 Volume configuration options

Volumes are data stores used to hold files for workloads running on the server. They are used to store Hyper-V files including VHD and VHDx files. All volumes are accessible to all servers within a cluster. While configuring volumes, pay attention to the following recommendations:

- The number of volumes created should be a multiple of the number of servers within the cluster.
- Create mirrored volumes for hot SQL Server data—because they provide the maximum performance for mixed read-write workloads.
- For SQL Server data warehouse workload, use dual parity to maximize the available capacity.
- Microsoft recommends limiting the number of volumes to 32 per cluster.

For more information about volumes and their planning, see Planning Volumes in Storage Spaces Direct.

6.2.3 Managing cache for SQL Server

The default settings for the write-back cache in S2D automatically size the cache based on certain preset factors. However, certain volumes may have high amounts of IOPs requirements. For such volumes, S2D allows a fixed amount of cache to be assigned. The PowerShell parameter –`WriteCacheSize` in the New-Volume command can be used for this purpose. Dell EMC recommends setting these parameters when creating the tempDB volume. Also, in OLTP scenarios, the volume that stores the log files can also be assigned by using this parameter.

Note: This parameter should be used with caution, as misconfiguration can cause performance degradation.

6.3 Windows Server, Hyper-V, and SQL Server configuration

This section outlines the best practices to configure Windows Server 2016, Hyper-V, and SQL Server 2016.

6.3.1 Windows OS configuration

The best practices for configuring the Windows OS are as follows:

- Use an **allocation unit size** of 64 KB to format the volume that stores the database files.
- Enable Windows LPIM policy by adding an account with privileges to run `sqlservr.exe`. The **Lock Pages in Memory (LPIM)** policy option determines which accounts can use a process to keep data in physical memory. This prevents the Windows operating system from paging out a significant amount of data from physical memory to virtual memory on disk.

6.3.2 Microsoft Hyper-V 2016

The following outlines the best practices for configuring the Hyper-V host and deploying VMs:

- **Do not over commit resources** – Hyper-V allows overcommitting resources like CPU and Network. However overcommitting leads to performance degradation when the resource usage exceeds the available resources.
- **Provision memory** – Memory provisioning can be either static or dynamic. Memory configuration and provisioning are explained in detail in section 6.4.3.
- **Provision compute resource** – A virtual CPU (vCPU) is a representation of the physical core of a processor or threads/logical processors in the core. Hyper-V allows adjusting the relative weights and reserves for a particular VM in **Processor → Settings**. If CPU resources are overcommitted, setting the weights and reserves optimizes the way these resources are used. You can prioritize or deprioritize specific VMs based on your needs.
- **Use fixed-size virtual hard disks** – Dell EMC recommends using fixed-size virtual hard disks (VHD/VHDX) for production workloads. Using dynamic virtual hard disks can result in occasional pauses when the dynamic disks need to be resized. Use dynamic disks for non-critical test environments or non-production environments.
- **Use separate VHDs/VHDXs for OS, data, and log files** – To rule out disk contention, Dell EMC recommends using separate VHDs/VHDXs. For best performance, create distinct VHDs/VHDXs for operating system, SQL Server data files, and SQL Server log files.

6.3.3 Microsoft SQL Server 2016

This section provides guidelines and best practices to optimize a virtualized SQL Server environment.

- **Memory settings**
  For SQL Server 2016, memory can be assigned for a VM either dynamically or statically. To decide whether memory for a VM should be assigned dynamically or statically, consider the following factors:
  - How frequently you monitor your database
  - The VM size (size of the processors, memory and other resources) as compared to the size of a single NUMA node on the host's physical architecture
  - Preference between performance and levels of scalability

  In general, dynamic memory is preferred when VMs are unmonitored, and are relatively smaller and when scalability is preferred over performance. On the other hand, for larger production VMs that are reasonably monitored, static memory is preferred for better and more consistent performance.
  For dynamically assigned memory, see the Microsoft article Running SQL Server with Hyper-V Dynamic Memory.

The best practices for dynamic memory are outlined below:

- Determine and implement the VM **Startup RAM** and minimum memory based on your needs. Microsoft recommends that **max server memory** be left at its default setting, which allows SQL Server to manage memory dynamically. However, if it is a VM on which you are running multiple applications, and/or you can reasonably ascertain the maximum amount of memory that you want to assign to SQL Server, Dell EMC recommends that you change this value based on your needs.
- Set **min server memory** (default value is zero) based on usage and performance considerations because dynamic memory is enabled for the VM.
- When using dynamic memory, Dell EMC recommends that you set reserves for this VM by using the **Memory Buffer** option. The amount of memory reserved should be based on the **min server memory** set for SQL Server and memory required for any other applications and the OS.

The recommendations for static memory allocation are outlined below:

- To maximize performance, Dell EMC recommends that you assign a particular amount of memory to a VM based on its virtual and physical NUMA architecture.
- To ensure optimal performance, disable the **NUMA Spanning** option for the Hyper-V host, so that each virtual NUMA node is backed by a single NUMA node.

Note: Use this option with caution as Hyper-V will not start, restore, or accept a live migration for a VM if the hypervisor is unable to map each virtual NUMA node to a physical NUMA node.

- Dell EMC recommends that you set **max server memory** and **min server memory** values based on the amount of memory you want to reserve for the operating system, typical requirements of your SQL Server, and other performance considerations. For more information, see Blitz Result: Memory Dangerously Low or Max Memory Too High.

- **Parallelism settings**
  - **Max Degree of Parallelism:** This SQL Server configuration option controls the number of processors used for the parallel execution of a query. If the SQL Server VM in question migrates from hosts with different core configurations, leave this setting at 0; however, if that is not the case, use a different value. For more information, see MAXDOP of Confusion.
  - **Cost threshold for parallelism:** This SQL Server configuration option specifies a threshold at which query plans run in parallel. Dell EMC recommends that you change the value of this setting from its default value of 5 to 50. You can adjust this number further depending on the workload requirements. For more information, see MAXDOP of Confusion.

7 Sample implementation

This guide provides a sample solution built for an OLTP solution. The architectural design, sizing, and configurations for the sample are outlined in the subsections below.

7.1 OLTP workload solution

This section outlines a SQL Server solution for an OLTP workload supporting up to 5,000 transactions on a database size of up to 20 TB. Figure 4 provides an overview of this solution.

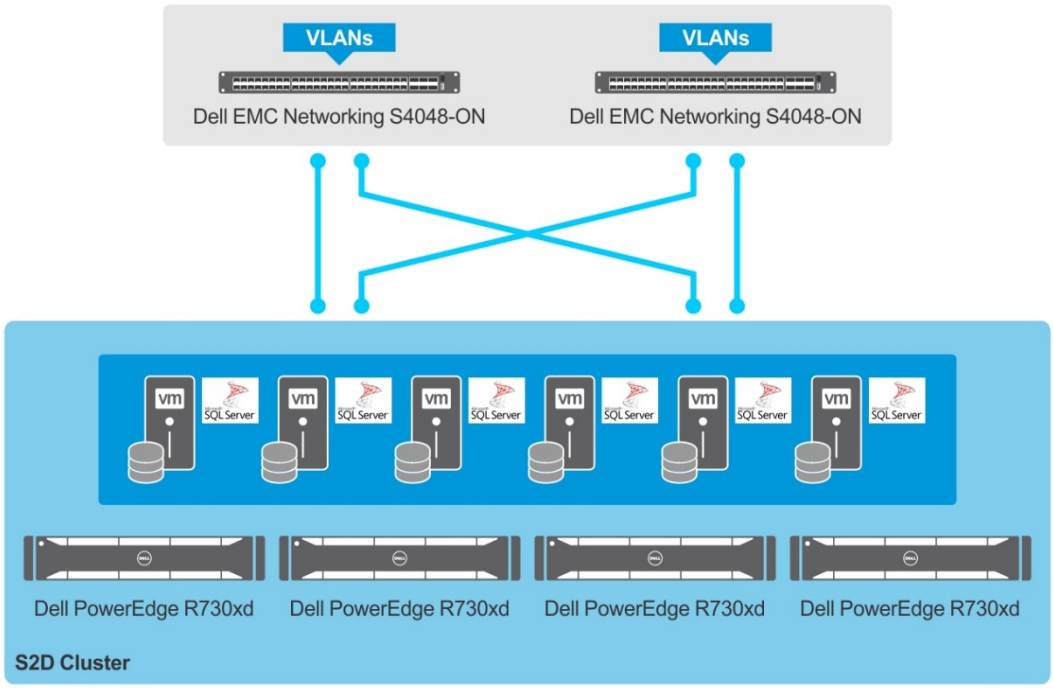

Figure 4 OLTP workload solution architecture

The components for this solution are listed in the below table:

Table 2 OLTP workload solution configuration

Component type	Details	Count
Server	Dell EMC PowerEdge R730xd server	4
Processor	Intel Xeon E5-2660 v4 2.0 GHz with 14 cores	8
Memory	512 GB (16 x 32 GB 2400 MT/s DIMM) – per server	NA
Storage controller	Internal HBA330	4
Storage – SSDs (cache)	800 GB Intel Write Intensive Solid State Drive, 2.5 inch, SATA	16
Storage – HDDs (capacity)	4 TB Toshiba 7.2K RPM, 6 Gbps, 3.5 inch, SATA	32
Network card	**NDC**: Intel i350 QP 1 GbE (default) **Add-On**: Mellanox Connect-X Pro 3 DP (10 GbE SFP+)	4+4
Network switch—backend	Dell EMC Networking S4048–ON	2
Network switch—management	Dell EMC Networking S3048–ON	2

The implementation details for this solution are listed in the table below:

Table 3 Workload implementation details

Component	Detail
Number of VMs	6
Number of database VHDs	6
Number of Log VHDs	6
Number of TempDB VHDs	6
Maximum database transaction rate per second	7,000
Per VM configuration	
vCPUs	24
Memory	192 GB
Memory for DB	128 GB
Total size of database VHD	6 TB
Total size of Log VHD	2 TB
Total size of TempDB VHDs	2 TB

8 References

- Dell EMC Microsoft Storage Spaces Direct Ready Nodes - Deployment Guide
- Deploying Storage Spaces Direct
- Fault tolerance and storage efficiency in Storage Spaces Direct
- Planning volumes in Storage Spaces Direct