Reference Architecture: Microsoft SQL Server and VMware vSAN with ThinkSystem Servers

Last update: 07 August 2017
Version 1.0
Configuration Reference Number: DBSMS04XX73

Describes reference architecture for Microsoft SQL Server using VMware vSAN for hyperconverged storage

Contains performance data for sizing recommendations

Includes deployment details and best practices

Contains detailed bill of materials for servers and network switches

Mike Perks

Pawan Sharma

Table of Contents

1 Introduction 1

2 Business problem and business value 2

2.1 Business problem 2

2.2 Business value 2

3 Requirements 3

3.1 Functional requirements 3

3.2 Non-functional requirements 3

4 Architectural overview 4

5 Component model 5

5.1 Microsoft SQL Server 5

5.2 VMware vSAN 6

6 Operational model 8

6.1 Hardware components 8

6.2 Hyperconverged servers 10

6.3 Networking 12

6.4 Systems management 13

7 Deployment considerations 15

7.1 Sizing considerations 15

7.2 Best practices and limitations 15

8 Appendix: Bill of Materials 18

8.1 Server BOM 18

8.2 Networking BOM 20

8.3 Rack BOM 20

Resources 21

1 Introduction

This document describes the reference architecture for the Microsoft SQL Server 2016 using Lenovo® ThinkSystem servers and networking with VMware vSAN for hyperconverged storage. The intended audience is IT professionals, technical architects, sales engineers, and consultants to assist in planning, designing, and implementing virtualized Microsoft SQL Server on VMware ESXi and vSAN.

This document provides an overview of the business problem and business value that is addressed by Microsoft SQL Server. A description of customer requirements is followed by an architectural overview of the solution and a description of the logical components. The operational model describes the architecture for deploying into small to large Enterprises. Performance and sizing information is provided with the best practices and networking considerations for implementing Microsoft SQL Server and VMware vSAN. The last section features detailed Bill of Materials configurations for Lenovo ThinkSystem servers and Lenovo network switches that are used in the solution.

2 Business problem and business value

The following section provides a summary of the business problems that this reference architecture is intended to help address, and the value that this solution can provide.

2.1 Business problem

As one of the fastest growing database platforms, Microsoft SQL Server deployments are becoming increasingly critical to organizations. They are used in everything from departmental databases to business-critical workloads, including enterprise resource planning, customer relationship management and business intelligence. At the same time, enterprises are virtualizing SQL Server to consolidate their datacenter footprint, control costs and accelerate provisioning. These trends of delivering SQL Server databases as dynamic, virtualized services make it essential to select the right server and storage architecture.

2.2 Business value

Database performance has long been the primary criteria for selecting infrastructure. Multicore processors and large system memory capacity have now moved the performance conversation away from compute to the storage system. Storage solutions that support virtualized SQL Server VMs need to handle a dynamic mix of transactional (OLTP) and analytical (OLAP) databases, along with their unique storage I/O profiles and active datasets. This requires efficiently delivering random and sequential read/write at high performance, across sizable amounts of active or hot data.

Virtualization of databases using VMware ESXi and VMware vSAN hyperconverged capabilities allows more database instances thus reducing the number of servers required for smaller deployments. This solution employs the latest generation of solid state drives to provide high IOPs, low latency, and high storage capacity for virtualized SQL Server databases.

This solution validated design utilizes VMware vSAN to configure physical servers in a hyperconverged cluster. VMware vSAN unifies VM and storage deployment while utilizing on-node storage resources to greatly simplify physical host deployment with linear scalability. Additionally, this hyperconverged platform provides:

- Higher performance and scalability: Start small and scale databases as your needs grow, but without the concessions of traditional infrastructure
- Improved availability: Keep key applications protected and running with frequent, easy-to-restore backups and affordable, simple disaster recovery
- Reduced operational complexity: Leverage simple, consumer-grade management, VM centric operations and unprecedented insight into application and storage performance
- Reduced Cost: A VMware vSAN storage platform reduces CAPEX and OPEX over traditional storage

3 Requirements

This section descirbes the functional and non-functional requirements for this reference architecture.

3.1 Functional requirements

Table 1 lists the functional requirements of a database management system (database) such as Microsoft SQL Server.

Table 1: Functional requirements

Requirement name	**Description**
Stores any kind of data	A database management system should be able to store any kind of data including binary such as images and video.
Support ACID Properties	Database must support ACID (Accuracy, Completeness, Isolation, and Durability) properties.
Represents complex relationships between data	Database must represent the complex relationships between data to make the efficient and accurate use of data.
Database schema	Database must provide a method to create and maintain the database schema using both GUI and command line.
Database operations	Database must provide a method to submit SQL queries and return results using GUI, command line, and other interfaces.
Reporting	Database should provide a method to generate formatted reports in various file formats, on-screen or printed
Multiple views	Depending on role, users may see different views of the data.
Concurrent use	Database must be able to respond to multiple requests at a time from multiple sources.

3.2 Non-functional requirements

Table 2 lists the non-functional requirements that are needed for deployment.

Table 2: Non-functional requirements

Requirement name	**Description**
Data integrity	Integrity ensures the quality and reliability of database system
High availability	This critical part of IT infrastructure must always be available
Disaster recovery	Provide ability for secondary data center to take over if the primary data center suffers a catastrophic failure and all components fail
Scalability	Solution components such as compute and storage capacity and performance scale with an increase in number of concurrent users or transactions
Security	Solution provides ways to secure data based on authorized role
Ease of installation	Reduced complexity of database deployment
Ease of management/operations	Simple management of infrastructure including support for rolling upgrades of hardware and software
Backup/Recovery	Solution support for integrated backup

4 Architectural overview

Microsoft SQL Server is a database platform for large-scale online transaction processing (OLTP), data warehousing, and a business intelligence platform for data integration, analysis, and reporting solutions. It uses a common set of tools to deploy and manage databases for in-house and cloud environments.

Figure 1 shows an architectural overview of Microsoft SQL Server on Lenovo ThinkSystem SR650 with VMware vSAN. It shows how other VMs such as virtual desktop infrastructure (VDI) or Microsoft Exchange can use virtualized Microsoft SQL Server in a cluster of VMware ESXi hypervisor and VMware vSAN based servers.

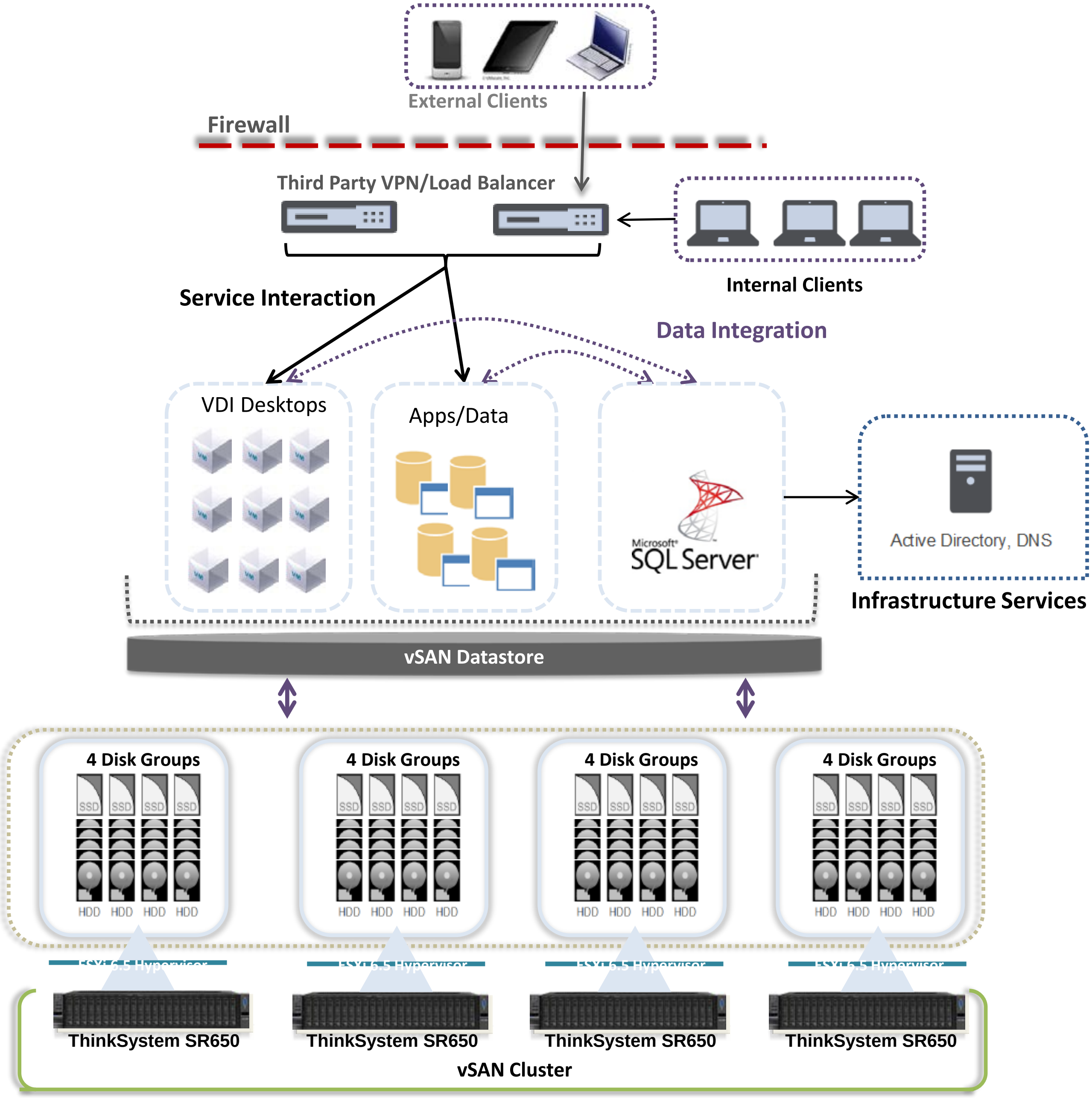


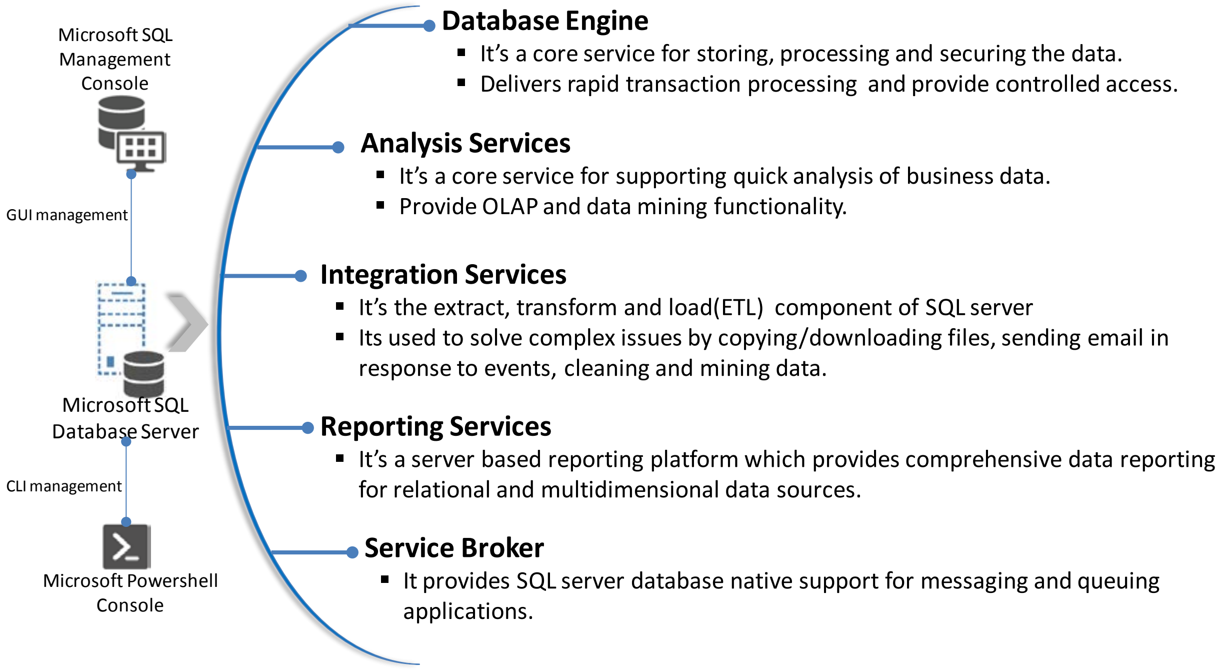
Figure 1: Lenovo ThinkSystem SR650 4 node hybrid vSAN cluster with Microsoft SQL Server

5 Component model

This chapter describes the main software components for the solution.

5.1 Microsoft SQL Server

Figure 2 is a layered component view for Microsoft SQL Server 2016.


Microsoft SQL Management Console
GUI management
Microsoft SQL Database Server
CLI management
Microsoft Powershell Console

Database Engine

- It’s a core service for storing, processing and securing the data.
- Delivers rapid transaction processing and provide controlled access.

Analysis Services

- It’s a core service for supporting quick analysis of business data.
- Provide OLAP and data mining functionality.

Integration Services

- It’s the extract, transform and load(ETL) component of SQL server
- Its used to solve complex issues by copying/downloading files, sending email in response to events, cleaning and mining data.

Reporting Services

- It’s a server based reporting platform which provides comprehensive data reporting for relational and multidimensional data sources.

Service Broker

- It provides SQL server database native support for messaging and queuing applications.

Figure 2: Component model with Microsoft SQL Server

Microsoft SQL Server features the following main components:

Database Engine This part of SQL Server actually creates and drives relational databases.

Analysis Services SQL Server Analysis Services (SSAS) is the data analysis component of SQL Server. It can create OLAP (OnLine Analytical Processing) cubes — sophisticated programming objects for organizing data inside a relational database — and do data mining (pulling relevant data out of a database in response to an ad-hoc question).

Integration Services SQL Server Integration Services (SSIS) performs the extract-transform-load (ETL) process that cleans up and formats raw data from source systems for inclusion in the database as ready-to-use information.

Reporting Services SQL Server Reporting Services (SSRS) provides reporting regardless of a database’s operating system.

Service Broker SQL Server Service Broker provides native support for messaging and queuing applications which makes it easier to build distributed and reliable applications that use the Database Engine components.

Other software components such as Lenovo XClarity Administrator are not shown. As well as providing management of Lenovo hardware, XClarity Administrator also has a plugin for VMware vCenter, which is further described in “Systems management” on page 10.

5.2 VMware vSAN

VMware vSAN is a Software Defined Storage (SDS) solution embedded in the ESXi hypervisor. VMware vSAN pools flash caching devices and magnetic disks across three or more 10 GbE connected servers into a single shared datastore that is resilient and simple to manage.

VMware vSAN can be scaled to 64 servers, with each server supporting up to five disk groups, with each disk group consisting of a solid-state drives (SSDs) and up to seven hard disk drives (HDDs). Performance and capacity can be easily increased by adding components, such as disks, flash, or servers.

The flash cache is used to accelerate reads and writes. Frequently read data is kept in read cache; writes are coalesced in cache and destaged to disk efficiently, which greatly improves application performance.

VMware vSAN manages data in the form of flexible data containers that are called *objects*. The following types of objects for VMs are available:

- VM Home
- VM swap (`.vswp`)
- VMDK (`.vmdk`)
- Snapshots (`.vmsn`)

Internally, VM objects are split into multiple components that are based on performance and availability requirements that are defined in the VM storage profile. These components are distributed across multiple hosts in a cluster to tolerate simultaneous failures and meet performance requirements. VMware vSAN uses a distributed RAID architecture to distribute data across the cluster. Components are distributed with the use of the following two storage policies:

- Number of stripes per object. It uses RAID 0 method.
- Number of failures to tolerate. It uses either RAID-1 or RAID-5/6 method. RAID-5/6 is currently supported for an all flash configuration only.

VMware vSAN uses the Storage Policy-based Management (SPBM) function in vSphere to enable policy driven VM provisioning, and uses vSphere APIs for Storage Awareness (VASA) to make available vSAN storage capabilities to vCenter. This approach means that storage resources are dynamically provisioned based on requested policy, and not pre-allocated as with many traditional storage solutions. Storage services are precisely aligned to VM boundaries; change the policy, and vSAN implements the changes for the selected VMs. Table 3 lists the vSAN storage policies.

Table 3: vSAN storage policies

Storage Policy	Description	Default	Maximum
Failure Tolerance Method	Defines a method used to tolerate failures. RAID-1 uses mirroring and RAID 5/6 uses parity blocks (erasure encoding) to provide space efficiency. RAID-5/6 is supported only for All Flash configurations. RAID 5 requires minimum 4 hosts and RAID 6 requires minimum 6 hosts. When RAID 5/6 is chosen, RAID 5 is used when FTT=1 and RAID 6 is used when FTT=2.	RAID-1	N/A
Number of failures to tolerate	Defines the number of host, disk, or network failures a VM object can tolerate. For *n* failures tolerated, n+1 copies of the VM object are created and 2n+1 hosts with storage are required. For example with a FTT=1, RAID-1 uses 2x the storage and RAID-5/6 uses 1.33x the storage. When FTT=2, RAID-1 uses 3x the storage and RAID-5/6 uses 1.5x the storage.	1	3
Number of disk stripes per object	The number of HDDs across which each replica of a VM object is striped. A value higher than 1 might result in better performance, but can result in higher use of resources.	1	12
Object space reservation	Percentage of the logical size of the object that should be reserved (or thick provisioned) during VM creation. The rest of the storage object is thin provisioned. If your disk is thick provisioned, 100% is reserved automatically. When deduplication and compression is enabled, this should be set to either 0% (do not apply) or 100%.	0%	100%
Flash read cache reservation	SSD capacity reserved as read cache for the VM object. Specified as a percentage of the logical size of the object. Should be used only to address read performance issues. Reserved flash capacity cannot be used by other objects. Unreserved flash is shared fairly among all objects.	0%	100%
Force provisioning	If the option is set to Yes, the object is provisioned, even if the storage policy cannot be satisfied by the data store. Use this parameter in bootstrapping scenarios and during an outage when standard provisioning is no longer possible. The default of No is acceptable for most production environments.	No	N/A
IOPS limit for object	Defines IOPS limit for a disk and assumes a default block size of 32 KB. Read, write and cache operations are all considered equivalent. When the IOPS exceeds the limit, then IO is throttled.	0	User Defined
Disable object checksum	Detects corruption caused by hardware/software components including memory, drives, etc. during the read or write operations. Object checksums carry a small disk IO, memory and compute overhead and can be disabled on a per object basis.	No	Yes

6 Operational model

This section describes the mapping of the logical components of Microsoft SQL Server and VMware vLAN onto Lenovo servers, storage, and networking. The BOM configurations for the hardware are described in Section 8 on page 18.

Figure 3 shows the overall operational model. In this example four Lenovo ThinkSystem SR650 servers are clustered using vSAN to provide a single distributed datastore across the four servers. Each server uses the ESXi 6.5 hypervisor and has two VMs for Microsoft SQL Server. The servers are networked together using Lenovo RackSwitch G8272 top of rack switches.

The cluster of servers is not restricted to just running Microsoft SQL Server VMs and and can operate in combination with other VMs to provide a single scalable platform for applications.

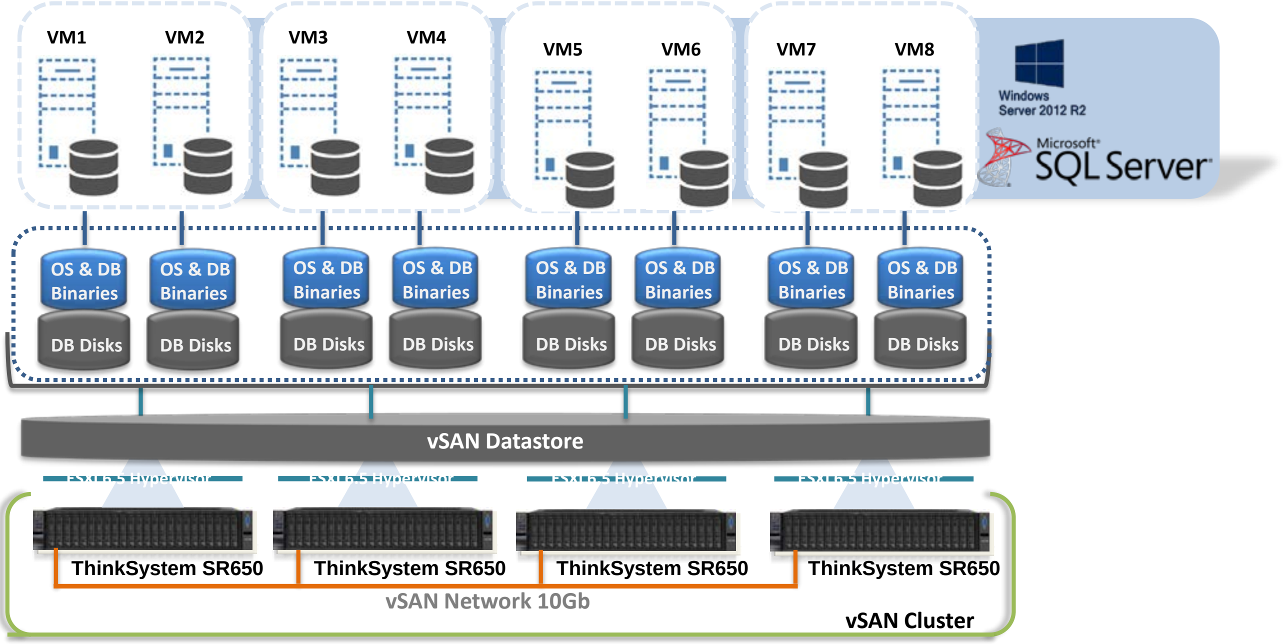


Figure 3: Operational Model for Microsoft SQL Server with VMware vSAN.

6.1 Hardware components

The following section describes the hardware components in a Microsoft SQL Server deployment.

6.1.1 Servers

You can use Lenovo ThinkSystem SR630 and ThinkSystem SR650 server platforms to compute clusters using vSAN for hyperconverged storage. The ThinkSystem SR650 offers more drive bays than the ThinkSystem SR630.

Lenovo ThinkSystem SR630

Lenovo ThinkSystem SR630 (as shown in Figure 4) is an ideal 2-socket 1U rack server for small businesses up to large enterprises that need industry-leading reliability, management, and security, as well as maximizing performance and flexibility for future growth. The SR630 server is designed to handle a wide range of workloads, such as databases, virtualization and cloud computing, virtual desktop infrastructure (VDI), infrastructure security, systems management, enterprise applications, collaboration/email, streaming media, web, and HPC. It offers up to twelve 2.5-inch hot-swappable SAS/SATA HDDs or SSDs together with up to 4 on-board NVMe PCIe ports that allow direct connections to the U.2 NVMe PCIe SSDs.

Figure 4: Lenovo ThinkSystem SR630

For more information, see this website: lenovopress.com/lp0643

Lenovo ThinkSystem SR650

Lenovo ThinkSystem SR650 (as shown in Figure 5) is similar to the SR630 but in a 2U form factor.

Figure 5: Lenovo ThinkSystem SR650

The key differences compared to the SR630 server are more expansion slots and chassis to support up to twenty-four 2.5-inch or fourteen 3.5-inch hot-swappable SAS/SATA HDDs or SSDs together with up to 8 on-board NVMe PCIe ports that allow direct connections to the U.2 NVMe PCIe SSDs. The SR650 server also supports up to two NVIDIA GRID cards for graphics acceleration.

For more information, see this website: lenovopress.com/lp0644

6.1.2 10 GbE networking

The following Lenovo 10 GbE top of rack (ToR) switch is recommended for use with VMware vSAN:

- Lenovo RackSwitch G8272

For more information about network switches, see this website: shop.lenovo.com/us/en/systems/networking/ethernet-rackswitch

Lenovo RackSwitch G8272

The Lenovo RackSwitch G8272 that uses 10 Gb SFP+ and 40 Gb QSFP+ Ethernet technology is specifically designed for the data center. It is ideal for today's big data, cloud, and optimized workload solutions. It is an enterprise class Layer 2 and Layer 3 full featured switch that delivers line-rate, high-bandwidth switching, filtering, and traffic queuing without delaying data. Large data center-grade buffers help keep traffic moving, while the hot-swap redundant power supplies and fans (along with numerous high-availability features) help provide high availability for business sensitive traffic.

The RackSwitch G8272 (shown in Figure 6), is ideal for latency sensitive applications, such as high-performance computing clusters and financial applications. In addition to the 10 Gb Ethernet (GbE) and 40 GbE connections, the G8272 can use 1 GbE connections. The G8272 supports the newest protocols, including Data Center Bridging/Converged Enhanced Ethernet (DCB/CEE) for Fibre Channel over Ethernet (FCoE), iSCSI and network-attached storage (NAS).

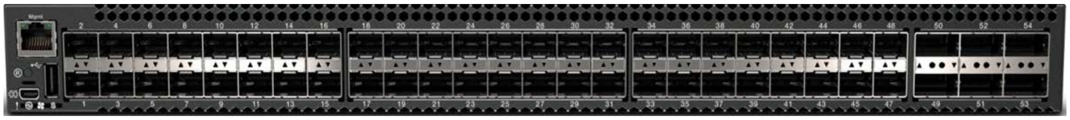


Figure 6: Lenovo RackSwitch G8272

For more information, see this website: lenovopress.com/tips1267

6.1.3 1 GbE network

The following Lenovo 1GbE ToR switch is recommended for use with VMware vSAN:

- Lenovo RackSwitch G8052

Lenovo RackSwitch G8052

The Lenovo System Networking RackSwitch G8052 (as shown in Figure 7) is an Ethernet switch that is designed for the data center and provides a virtualized, cooler, and simpler network solution. The Lenovo RackSwitch G8052 offers up to 48 1 GbE ports and up to 4 10 GbE ports in a 1U footprint. The G8052 switch is always available for business-sensitive traffic by using redundant power supplies, fans, and numerous high-availability features.

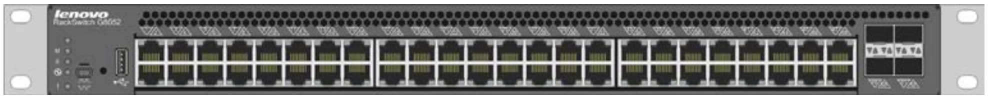


Figure 7: Lenovo RackSwitch G8052

For more information, see this website: lenovopress.com/tips0813

6.2 Hyperconverged servers

This section describes the recommended configuration for Microsoft SQL Server 2016 using a vSAN cluster of four Lenovo ThinkSystem SR650 servers. Each host has ESXi 6.5b hypervisor and is configured with VMware vSAN, vSphere high availability (HA), and vSphere Distributed Resource Scheduler (DRS).

The cluster of 4 servers has 2 Microsoft SQL Server VMs per host, each configured as follows:

- 20 vCPUs
- 200 GB RAM
- Windows Server 2012 R2
- Microsoft SQL 2016 Enterprise

VMware vSAN storage policies are configured on each host as follows:

- Number of FTT = 1
- Number of disk stripes per object = 1
- Force Provisioning = No
- Object space reservation = 100%
- Flash read cache reservation = 0%

The remainder of this section provides a high-level summary of the performance results of executing the HammerDB test suite on the 4 node vSAN cluster using 8 Microsoft SQL Server VMs.

HammerDB is a graphical open source database load testing and benchmarking tool for Linux and Windows to test databases running on any operating system. HammerDB is automated, multi-threaded and extensible with dynamic scripting support. See this website for more details: hammerdb.com.

A 5,000 scale OLTP database was used for each VM, which equates to about 1.4 TB of database records on the cluster per VM and a total storage of 11.2 TB in the cluster.

6.2.1 Hybrid performance test results

Each server has the following configuration:

- 2 x Intel Scalable 8176 (28 cores @ 2.1 GHz) processors
- 768 GB RAM
- 4 x 800 GB SAS SSDs
- 20 x 1.8 TB SAS HDD

The 4 node cluster has a raw storage of 144 TB of capacity HDDs which equates to approximately 67 TB of actual storage using a vSAN FTT value of 1.

Separate load servers were used to simulate the user load. Each OLTP database workload was simulated by running 500 users per SQL VM simultaneously for a total of 1000 users per node. Each test run had a 10 minute ramp up phase and executed for 20 minutes to simulate 60,000 transactions per user.

Table 4 shows the results of the HammerDB test during the 20 minute steady-state period.

Table 4: Microsoft SQL Server results with hybrid cluster

Node	**VM Name**	**HammerDB TPM**	**HammerDB NOPM**
Node 1	SQL Instance - 1	1775979	386276
	SQL Instance - 2	1730726	376235
Node 2	SQL Instance - 3	1177832	256765
	SQL Instance - 4	1425509	309960
Node 3	SQL Instance - 5	1368647	297768
	SQL Instance - 6	1708404	371334
Node 4	SQL Instance - 7	1550474	337235
	SQL Instance - 8	1636702	355699
	Total	12374273	2691272

The 4 node hybrid cluster can execute over 12 million transactions per minute and almost 2.7 million new orders per minute. The CPU utilization is 65% to 70%.

6.2.2 All flash performance test results

Each server has the following configuration:

- 2 x Intel Scalable 8176 (28 cores @ 2.1 GHz) processors
- 768 GB RAM
- 4 x 800 GB SAS SSDs
- 8 x 7.68 GB SAS SSDs

The 4 node cluster has a raw storage of 246 TB of capacity SSDs which equates to approximately 114 TB of actual storage using a vSAN FTT value of 1. VMware vSAN compression and de-duplication are disabled and the failure tolerance method for all disks was set to RAID-1.

Separate load servers were used to simulate the user load. Each OLTP database workload was simulated by running 675 users per SQL VM simultaneously for a total of 1350 users per node. Each test run had a 10 minute ramp up phase and executed for 20 minutes to simulate 60,000 transactions per user.

Table 5 shows the results of the HammerDB test during the 20 minute steady-state period.

Table 5: Microsoft SQL Server results with all flash cluster

Node	VM Name	HammerDB TPM	HammerDB NOPM
Node 1	SQL Instance - 1	1926525	419961
	SQL Instance - 2	2239894	506725
Node 2	SQL Instance - 3	1904585	414548
	SQL Instance - 4	1331949	289474
Node 3	SQL Instance - 5	1329051	289044
	SQL Instance - 6	1324114	288118
Node 4	SQL Instance - 7	1586640	345030
	SQL Instance - 8	1533505	333106
	Total	**13176263**	**2886006**

The 4 node cluster can execute over 13 million transactions per minute and almost 2.9 million new orders per minute. The CPU utilization is 75% to 80%.

6.3 Networking

The 10GbE data network is the fabric that carries all inter-node storage I/O traffic for the vSAN distributed file system, in addition to the user data traffic via the virtual Network Interface Cards (NICs) exposed through the hypervisor to the virtual machines.

It is a networking best practice to use VLANs to logically separate different kinds of network traffic. The following standard VLANs are recommended:

- Management Used for all management traffic for the hypervisor
- Storage network Used for vSAN storage traffic
- vSphere vMotion Used to move VMs from one server to another.
- Fault Tolerance Used to support the fault tolerance (FT) feature of vSphere.

It is recommended that two top of rack (ToR) switches are used for redundancy. In order to support the logical pairing of the network adapter ports and to provide automatic failover of the switches, the Lenovo RackSwitch G8272 supports virtual link aggregation groups (VLAGs). When VLAG is enabled over the inter-switch link (ISL) trunk, it enables logical grouping of these switches. When one of the switches is lost, or the uplink from the host to the switch is lost, the connectivity is automatically maintained over the other switch. In addition, the Lenovo Cloud Network Operating System (CNOS) should be used on the G8272 switches.

Figure 8 shows the scenario of two dual-port or one quad-port NIC connectivity into two ToR Lenovo RackSwitch G8272 switches with VLAG.

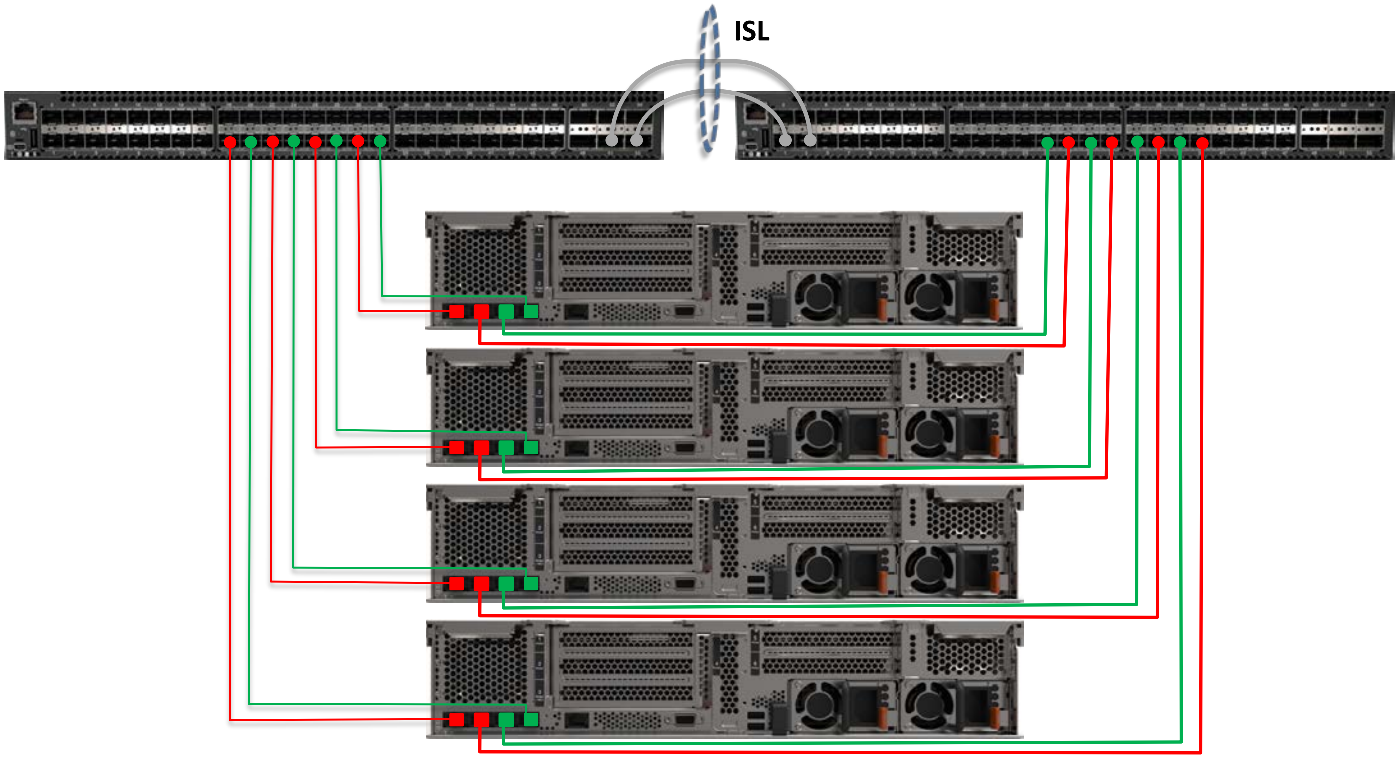
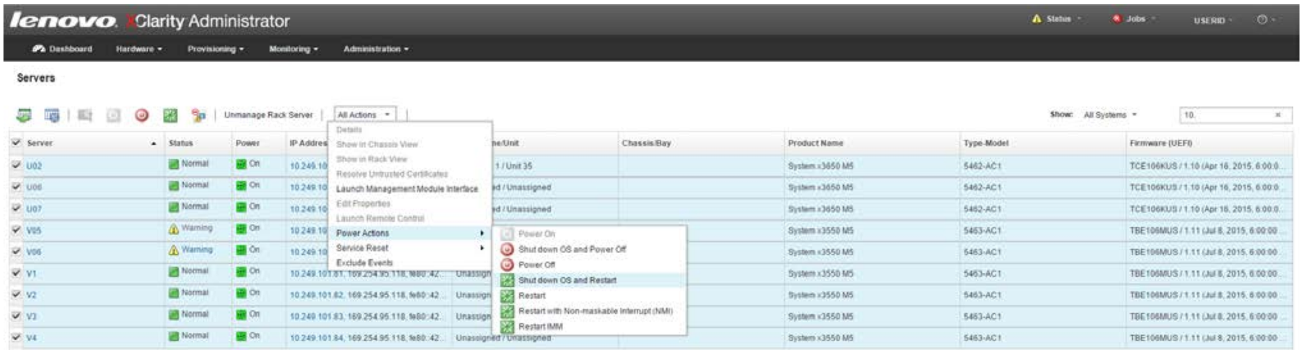


Figure 8: Redundancy with 10GbE ToR switches

6.4 Systems management

Lenovo XClarity™ Administrator is a centralized resource management solution that reduces complexity, speeds up response, and enhances the availability of Lenovo® server systems and solutions.

The Lenovo XClarity Administrator provides agent-free hardware management for Lenovo's ThinkSystem® rack servers, System x® rack servers, and Flex System™ compute nodes and components, including the Chassis Management Module (CMM) and Flex System I/O modules. Figure 9 shows the Lenovo XClarity administrator interface, in which Flex System components and rack servers are managed and are seen on the dashboard. Lenovo XClarity Administrator is a virtual appliance that is quickly imported into a virtualized environment server configuration.



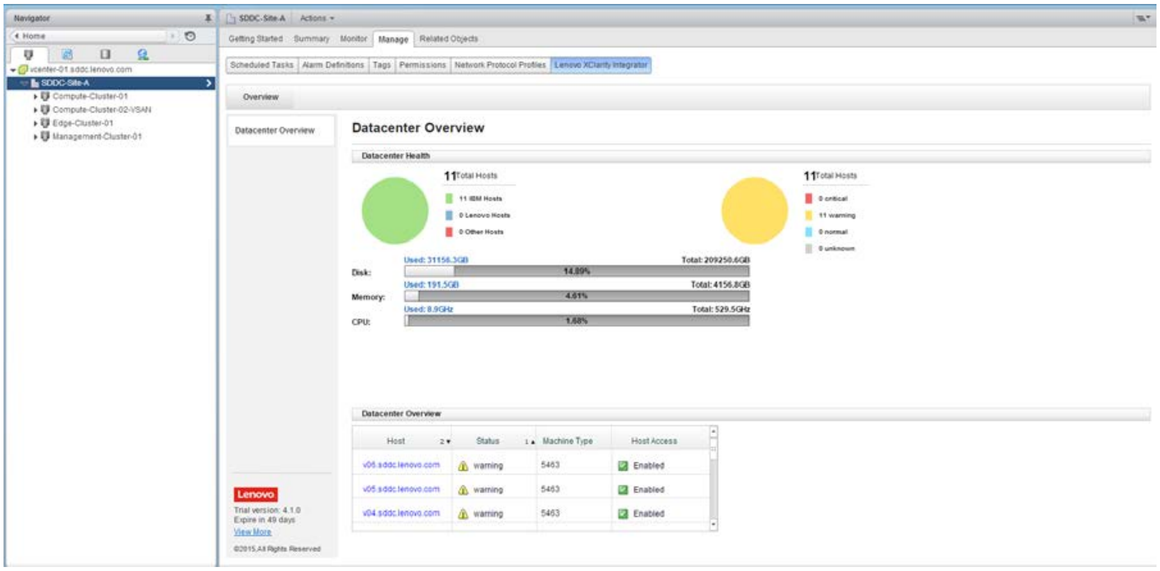
☑ Server	Status	Power	IP Addres	ne/Unit	Chassis/Bay	Product Name	Type-Model	Firmware (UEFI)
☑ U02	Normal	On	10.249.10	1 / Unit 35		System x3650 M5	5462-AC1	TCE106KUS / 1.10 (Apr 16, 2015, 6:00:0...
☑ U06	Normal	On	10.249.10	ed / Unassigned		System x3650 M5	5462-AC1	TCE106KUS / 1.10 (Apr 16, 2015, 6:00:0...
☑ U07	Normal	On	10.249.10	ed / Unassigned		System x3650 M5	5462-AC1	TCE106KUS / 1.10 (Apr 16, 2015, 6:00:0...
☑ V05	⚠ Warning	On	10.249.10			System x3550 M5	5463-AC1	TBE106MUS / 1.11 (Jul 8, 2015, 6:00:00 ...
☑ V06	⚠ Warning	On	10.249.10			System x3550 M5	5463-AC1	TBE106MUS / 1.11 (Jul 8, 2015, 6:00:00 ...
☑ V1	Normal	On	10.249.101.81, 169.254.95.118, fe80::42...	Unassign		System x3550 M5	5463-AC1	TBE106MUS / 1.11 (Jul 8, 2015, 6:00:00 ...
☑ V2	Normal	On	10.249.101.82, 169.254.95.118, fe80::42...	Unassign		System x3550 M5	5463-AC1	TBE106MUS / 1.11 (Jul 8, 2015, 6:00:00 ...
☑ V3	Normal	On	10.249.101.83, 169.254.95.118, fe80::42...	Unassign		System x3550 M5	5463-AC1	TBE106MUS / 1.11 (Jul 8, 2015, 6:00:00 ...
☑ V4	Normal	On	10.249.101.84, 169.254.95.118, fe80::42...	Unassigned / Unassigned		System x3550 M5	5463-AC1	TBE106MUS / 1.11 (Jul 8, 2015, 6:00:00 ...

Figure 9: XClarity Administrator interface

6.4.1 Lenovo XClarity integration with VMware

Lenovo also provides a Clarity integration modules for VMware vCenter. For more information, see this website: http://www3.lenovo.com/us/en/data-center/software/systems-management/c/systems-management.

By using the Lenovo XClarity Integrator for VMware vCenter, administrators can consolidate physical resource management in VMware vCenter, which reduces the time that is required for routine system administration.



Datacenter Overview

Datacenter Health

11 Total Hosts
🟩 11 IBM Hosts
🟦 0 Lenovo Hosts
🟥 0 Other Hosts

11 Total Hosts
🟥 0 critical
🟨 11 warning
🟦 0 normal
▨ 0 unknown

Disk: Used: 31156.3GB 14.89% Total: 209250.6GB
Memory: Used: 191.5GB 4.61% Total: 4156.8GB
CPU: Used: 8.9GHz 1.68% Total: 529.5GHz

Datacenter Overview

Host 2▼	Status 1▲	Machine Type	Host Access
v06.sddc.lenovo.com	⚠ warning	5463	☑ Enabled
v05.sddc.lenovo.com	⚠ warning	5463	☑ Enabled
v04.sddc.lenovo.com	⚠ warning	5463	☑ Enabled

Figure 10: Lenovo XClarity Integrator for VMware vCenter

The Lenovo XClarity Integrator for VMware vCenter provides the following features and benefits:

- Extends Lenovo XClarity Administrator features to the virtualization management console
- Enables management of legacy infrastructure from the virtualization management console
- Reduces workload downtime by dynamically triggering workload migration in clustered environments during rolling server reboots or firmware updates, and predicted hardware failures

7 Deployment considerations

This chapter contains additional considerations for deploying Microsoft SQL Server 2016 on a VMware vSAN cluster.

7.1 Sizing considerations

There are several considerations for sizing Microsoft SQL Server VMs in a VMware vSAN cluster.

The processor cores and clock speed depend on the number of transactions per minute required from the database. The performance results in section 6.2 on page 10 show the capabilities of a server using two platinum high-end Intel Xeon Scalable processors. If more processing power is required, then the 4 and 8 processor capable Lenovo ThinkSystem SR850 or ThinkSystem SR950 servers could be used. Lenovo recommends leaving 20-30% spare processor capacity to allow for failover when a server fails in the cluster.

VMware vSAN requires 32GB of system memory and two vCPUs. The remainder of the system memory should be sized to the VMs that will run on the server. Lenovo recommends leaving 20-30% spare memory capacity to allow for failover when a server fails in the cluster. Because Microsoft SQL Server is designed to utilize all available memory, Lenovo recommends not to over-provision memory as it leads to paging and thus significantly slowing performance.

The amount of storage to allocate for the database depends on a number of factors including the type of application that is being served by SQL Server. There are many different sizing tools available from Microsoft and other sources. Additional storage capacity should be built in both for growth in the database size and in case of a server failure when the data needs to be rebalanced across the remaining servers in the cluster. Lenovo recommends at least 30-40% extra capacity in the storage. The vSAN failures to tolerate (FTT) policy of 1 requires a raw capacity of twice the storage capacity. For example if the expected maximum size of the database files is 20TB, then Lenovo recommends a raw capacity of at least 52TB (i.e. fourteen 3.84 TB SSDs).

A related factor is the number of transaction per second provided by the vSAN storage. The transaction rate is directly related the achievable IOPS. The cache SSDs should be sized to the working set of the database. The capacity HDDs or SSDs should be sized to the amount of storage required. SSDs provide better overall performance and are available in 3.84TB and 7.68TB sizes which also allows for much larger database capacities that 2.5" HDDs.

7.2 Best practices and limitations

This section describes recommended best practices to provide data optimization and high availability of Microsoft SQL Server on VMware vSAN. For more information, see the VMware best practices document: "*Architecting Microsoft SQL Server on VMware vSphere*" vmware.com/content/dam/digitalmarketing/vmware/en/pdf/solutions/sql-server-on-vmware-best-practices-guide.pdf.

7.2.1 Cluster high availability

The minimum number of nodes in each cluster is 3 and should be at least 4 to provide failover. The following high availability features are recommended for an ESXi-based cluster:

- VMware vSphere high availability (HA) for failover
- VMware vSphere distributed resource scheduler (DRS) for load balancing
- Microsoft AlwaysOn availability groups (AAGs) for data redundancy

VMware vSphere HA pools VMs into a cluster to increase data resiliency. If a host fails, VMware HA moves the VMs to other hosts with spare capacity. Lenovo recommends enabling the “Admission Control Setting” and using the “Admission Control Policy” to set the percentage of cluster resources reserved as failover spare capacity.

VMware vSphere DRS can be used to group ESXi hosts into resource clusters to provide highly available resources and balance workloads. For example if the workload for one or more VMs changes significantly, then DRS will redistribute the VMs across the hosts. Decreases in workloads may result in consolidating the VMs and a temporary power down unneeded hosts. In order to keep the active working set for each SQL Server VM local to the node, Lenovo recommends creating a host group for each node and a “should” rule that keeps each SQL Server VM on a 1 to 1 ratio with the ESXi host. The hosts should be configured with a minimum of N+1 availability.

The Microsoft AlwaysOn availability groups (AAGs) is a recommended high availability mechanism. It uses a shared-nothing approach where transactions are replicated to other nodes so each node has a full copy of the database and transaction logs. This provides a very fast failover mechanism. The DRS anti-affinity rules need to be used to ensure that the SQL Server VMs are placed on different physical hosts.

DR across datacenters can also be done using AAGs assuming there is sufficient band-width between the sites. The scenarios for active-active and active-passive DR sites using AAGs are outside the scope of this document.

7.2.2 Data optimization

Lenovo recommends to use “thick provision eager zeroed” provisioning for all database storage as thin provisioning or lazy zeroed requires additional I/O for each new page written to disk.

The default failures to tolerate (FTT) policy is 1 and is satisfactory for most mission critical Microsoft SQL Server databases with AlwaysOn enabled. For all flash configurations, RAID 5 should be used for data virtual disks and RAID 1 should be use for transaction logs.

For all flash configurations, data compression can be used to further increase data capacity especially for data that is less frequently accessed. However data de-duplication is not recommended and should be disabled for SQL Server because of the frequency of changes. Note that de-duplication may be beneficial for backup volumes which are not changed very often.

7.2.3 Virtual Disks Configuration

Multiple virtual disks should be used to keep SQL binaries, database, and database logs files and achieve optimal SQL performance. All four SCSI controllers should be utilized and virtual disks should be distributed evenly across controllers as shown in Table 6.

Table 6: Mapping of virtual disks for ESXi and cluster size

Virtual Disk	vSCSI Controller Type	Controller #	Cluster size
Operating System	LSI Logic SAS	0	4 KB
SQL Binaries	LSI Logic SAS	0	4 KB
Backup/Restore	LSI Logic SAS	0	64 KB
SQL Database – 1	VMware Paravirtual	1	64 KB
SQL Database – 2	VMware Paravirtual	1	64 KB
TempDB – 1	VMware Paravirtual	2	64 KB
TempDB – 2	VMware Paravirtual	2	64 KB
TempDB log files	VMware Paravirtual	3	64 KB
Database log files	VMware Paravirtual	3	64 KB

All SQL database and log drives should be formatted with 64KB NTFS cluster size as it enhances the I/O performance without adding any overhead. The OS and SQL binary drives should be formatted with the standard 4KB NTFS cluster size. Drives space utilization should not be above 80% to achieve optimal performance.

To maximize the storage performance of SQL Server VMs, Lenovo recommends using the ESXi Paravirtual SCSI (PVSCSI) adapters. Each PVSCSI adapter can support up to 15 VMDKs.

7.2.4 SQL Server Files

To achieve high performance, the database should be split into multiple files across multiple virtual disks. In general, one database file per vCPU is ideal. For example a VM with 4 vCPUs hosting 400GB database could be split into four 100GB database files and should spread evenly across the two virtual disks.

For write intensive databases, it is recommended to distribute the database files over four or more virtual disks as it improves the write performance on the back-end and delivers consistent performance.

SQL log files (DB and TempDB) are written sequentially, so using multiple log files wouldn’t improve the performance. Using single log file per database is recommended.

TempDB is used as scratch space by the applications and is one of the most important factors of SQL performance. The number of TempDB files to be used is based on the vCPU count. If the vCPUs are less than 8, then configure the same number of TempDB files. If the number of vCPUs is higher than 8, then start with 8 TempDB files and monitor the contention for in-memory allocation (PAGELATCH_XX). The number of TempDB file should be increased in increments of four until contention is eliminated.

It is recommended to create all TempDBs with the same size and not allow for autogrowth. The TempDB file sizing should be done based on required application and is usually 1-10% of the database size.

8 Appendix: Bill of Materials

This appendix features the Bill of Materials (BOMs) for different configurations of hardware for virtualized Microsoft SQL Server deployments on VMware vSAN. There are sections for servers and networking switches that are orderable from Lenovo.

The BOM lists in this appendix are not meant to be exhaustive and must always be confirmed with the configuration tools. Note that vSAN server configurations must follow the vSAN rules and match the VMware Compatibility Guide (VCG): vmware.com/resources/compatibility. Any description of pricing, support, and maintenance options is outside the scope of this document.

For connections between ToR switches and devices (servers, storage, and chassis), the connector cables are configured with the device. The ToR switch configuration includes only transceivers or other cabling that is needed for failover or redundancy.

8.1 Server BOM

Table 7 and Table 8 list the vSAN hybrid and all flash BOMs server to be used with Microsoft SQL Server VMs.

Table 7: ThinkSystem SR650 (hybrid)

Code	Description	Quantity
7X06CTO1WW	ThinkSystem SR650 - 3yr Warranty	1
AUQB	Lenovo ThinkSystem Mainstream MB - 2U	1
AUVV	ThinkSystem MS 2U 24x2.5" Chassis	1
AWDH	Intel Xeon Platinum 8176 28C 165W 2.1GHz Processor	2
AUND	ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	24
AURA	ThinkSystem 2U/Twr 2.5" SATA/SAS 8-Bay Backplane	3
AUNL	ThinkSystem 430-8i SAS/SATA 12Gb HBA	3
AUMV	ThinkSystem M.2 with Mirroring Enablement Kit	1
AUUV	ThinkSystem M.2 CV3 128GB SATA 6Gbps Non-Hot Swap SSD	2
AUKJ	ThinkSystem 10Gb 2-port SFP+ LOM	1
AUPW	ThinkSystem XClarity Controller Standard to Enterprise Upgrade	1
AXCH	ThinkSystem Toolless Slide Rail Kit with 2U CMA	1
AUS8	ThinkSystem SR550/SR590/SR650 EIA Latch w/ VGA Upgrade Kit	1
AVWF	ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	2
6311	2.8m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2
5977	Select Storage devices - no configured RAID required	1
AXFT	VMware ESXi 6.5 (factory installed)	1
AUMG	ThinkSystem 2.5" HUSMM32 400GB Performance SAS 12Gb Hot Swap SSD	4
AUM1	ThinkSystem 2.5" 1.2TB 10K SAS 12Gb Hot Swap 512n HDD	20

Table 8: ThinkSystem SR650 (all flash)

Code	Description	Quantity
7X06CTO1WW	ThinkSystem SR650 - 3yr Warranty	1
AUQB	Lenovo ThinkSystem Mainstream MB - 2U	1
AUVV	ThinkSystem MS 2U 24x2.5" Chassis	1
AWDH	Intel Xeon Platinum 8176 28C 165W 2.1GHz Processor	2
AUND	ThinkSystem 32GB TruDDR4 2666 MHz (2Rx4 1.2V) RDIMM	24
AURA	ThinkSystem 2U/Twr 2.5" SATA/SAS 8-Bay Backplane	2
AUNL	ThinkSystem 430-8i SAS/SATA 12Gb HBA	2
AUMV	ThinkSystem M.2 with Mirroring Enablement Kit	1
AUUV	ThinkSystem M.2 CV3 128GB SATA 6Gbps Non-Hot Swap SSD	2
AUKJ	ThinkSystem 10Gb 2-port SFP+ LOM	1
AUPW	ThinkSystem XClarity Controller Standard to Enterprise Upgrade	1
AXCH	ThinkSystem Toolless Slide Rail Kit with 2U CMA	1
AUS8	ThinkSystem SR550/SR590/SR650 EIA Latch w/ VGA Upgrade Kit	1
AVWF	ThinkSystem 1100W (230V/115V) Platinum Hot-Swap Power Supply	2
6311	2.8m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2
5977	Select Storage devices - no configured RAID required	1
AXFT	VMware ESXi 6.5 (factory installed)	1
AUMH	ThinkSystem 2.5" HUSMM32 800GB Performance SAS 12Gb Hot Swap SSD	4
AUML	ThinkSystem 2.5” PM1633a 7.68TB Capacity SAS 12Gb Hot Swap SSD	8

8.2 Networking BOM

Table 9 and Table 10 list the BOMs for the network switches for 1 GbE and 10 GbE connectivity respectively

Table 9: RackSwitch G8052

Code	Description	Quantity
7159G52	Lenovo System Networking RackSwitch G8052 (Rear to Front)	1
6201	1.5m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2
3802	1.5m Blue Cat5e Cable	3
A3KP	Lenovo System Networking Adjustable 19" 4 Post Rail Kit	1

Table 10: RackSwitch G8272

Code	Description	Quantity
7159CRW	Lenovo System Networking RackSwitch G8272 (Rear to Front)	1
6201	1.5m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2
A3KP	Lenovo System Networking Adjustable 19" 4 Post Rail Kit	1
A1DP	1m QSFP+-to-QSFP+ cable	1
A1DM	3m QSFP+ DAC Break Out Cable	0

8.3 Rack BOM

Table 11 lists the BOM for a rack.

Table 11: Rack BOM

Code	Description	Quantity
93634PX	42U 1100mm Enterprise V2 Dynamic Rack	1
39Y8941	DPI Single Phase C19 Enterprise PDU (without line cord)	6
40K9614	DPI 30a Line Cord (NEMA L6-30P)	6

Resources

For more information about the topics that are described in this document, see the following resources:

- Microsoft SQL Server:

  docs.microsoft.com/en-us/sql/index

- VMware vSphere Hypervisor (ESXi):

  vmware.com/products/vsphere-hypervisor

- VMware vCenter Server:

  vmware.com/products/vcenter-server

- VMware vSAN:

  vmware.com/products/virtual-san

- VMware Compatibility Guide (VCG)

  vmware.com/resources/compatibility

Trademarks and special notices

References in this document to Lenovo products or services do not imply that Lenovo intends to make them available in every country.

Lenovo, the Lenovo logo, ThinkSystem, ThinkAgile, ThinkCentre, ThinkVision, ThinkVantage, ThinkPlus and Rescue and Recovery are trademarks of Lenovo.

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel Inside (logos), MMX, and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used Lenovo products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-Lenovo products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by Lenovo. Sources for non-Lenovo list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. Lenovo has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-Lenovo products. Questions on the capability of non-Lenovo products should be addressed to the supplier of those products.

All statements regarding Lenovo future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. Contact your local Lenovo office or Lenovo authorized reseller for the full text of the specific Statement of Direction.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in Lenovo product announcements. The information is presented here to communicate Lenovo’s current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard Lenovo benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.

Any references in this information to non-Lenovo websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this Lenovo product and use of those websites is at your own risk.