

Cisco HyperFlex All-NVMe Systems for Oracle Real Application Clusters: Reference Architecture

Contents

Executive summary	3
Cisco HyperFlex HX Data Platform All-NVMe storage	4
Why use Cisco HyperFlex all-NVMe systems for Oracle RAC deployments	5
Oracle RAC 19c Database on Cisco HyperFlex systems	6
Oracle Database scalable architecture overview	8
Solution components	9
Engineering validation	17
Performance comparison between all-NVMe and all-flash clusters	18
Reliability and disaster recovery	21
Conclusion	24
For more information	25

Executive summary

Oracle Database is the choice for many enterprise database applications. Its power and scalability make it attractive for implementing business-critical applications. However, making those applications highly available can be extremely complicated and expensive.

Oracle Real Application Clusters (RAC) is the solution of choice for customers to provide high availability and scalability to Oracle Database. Originally focused on providing best-in-class database services, Oracle RAC has evolved over the years and now provides a comprehensive high-availability stack that also provides scalability, flexibility and agility for applications.

With the Cisco HyperFlex™ solution for Oracle RAC databases, organizations can implement RAC databases using a highly integrated solution that scales as business demand increases. RAC uses a shared-disk architecture that requires all instances of RAC to have access to the same storage elements. Cisco HyperFlex uses the [Multi-writer option](#) to enable virtual disks to be shared between different RAC virtual machines.

This reference architecture provides a configuration that is fully validated to help ensure that the entire hardware and software stack is suitable for a high-performance clustered workload. This configuration follows the industry best practices for Oracle Databases in a VMware virtualized environment. Additional details about deploying Oracle RAC on VMware can be found [here](#).

Cisco HyperFlex HX Data Platform All-NVMe storage

Cisco HyperFlex systems are designed with an end-to-end software-defined infrastructure that eliminates the compromises found in first-generation products. With all-Non-Volatile Memory Express (NVMe) memory storage configurations and a choice of management tools, Cisco HyperFlex systems deliver a tightly integrated cluster that is up and running in less than an hour and that scales resources independently to closely match your Oracle Database requirements. For an in-depth look at the Cisco HyperFlex architecture, see the Cisco® white paper [Deliver Hyperconvergence with a Next-Generation Platform](#).

An all-NVMe storage solution delivers more of what you need to propel mission-critical workloads. For a simulated Oracle online transaction processing (OLTP) workload, it provides 71 percent more I/O operations per second (IOPS) and 37 percent lower latency than our previous-generation all-flash node. The behavior mentioned here was tested on a Cisco HyperFlex system with NVMe configurations, and the results are provided in the “[Engineering validation](#)” section of this document. A holistic system approach is used to integrate Cisco HyperFlex HX Data Platform software with Cisco HyperFlex HX220c M5 All NVMe Nodes. The result is the first fully engineered hyperconverged appliance based on NVMe storage.

- **Capacity storage:** The data platform’s capacity layer is supported by Intel 3D NAND NVMe solid-state disks (SSDs). These drives currently provide up to 32 TB of raw capacity per node. Integrated directly into the CPU through the PCI Express (PCIe) bus, they eliminate the latency of disk controllers and the CPU cycles needed to process SAS and SATA protocols. Without a disk controller to insulate the CPU from the drives, we have implemented reliability, availability, and serviceability (RAS) features by integrating the Intel Volume Management Device (VMD) into the data platform software. This engineered solution handles surprise drive removal, hot pluggability, locator LEDs, and status lights.
- **Cache:** A cache must be even faster than the capacity storage. For the cache and the write log, we use Intel® Optane™ DC P4800X SSDs for greater IOPS and more consistency than standard NAND SSDs, even in the event of high-write bursts.
- **Compression:** The optional Cisco HyperFlex Acceleration Engine offloads compression operations from the Intel® Xeon® Scalable CPUs, freeing more cores to improve virtual machine density, lowering latency, and reducing storage needs. This helps you get even more value from your investment in an all-NVMe platform.
- **High-performance networking:** Most hyperconverged solutions consider networking as an afterthought. We consider it essential for achieving consistent workload performance. That’s why we fully integrate a 40-Gbps unified fabric into each cluster using Cisco Unified Computing System™ (Cisco UCS®) fabric interconnects for high-bandwidth, low-latency, and consistent-latency connectivity between nodes.
- **Automated deployment and management:** Automation is provided through Cisco Intersight™, a software-as-a-service (SaaS) management platform that can support all your clusters from the cloud to wherever they reside in the data center to the edge. If you prefer local management, you can host the Cisco Intersight Virtual Appliance, or you can use Cisco HyperFlex Connect management software.

All-NVMe solutions support most latency-sensitive applications with the simplicity of hyperconvergence. Our solutions provide the first fully integrated platform designed to support NVMe technology with increased performance and RAS.

Why use Cisco HyperFlex all-NVMe systems for Oracle RAC deployments

Oracle Database acts as the back end for many critical and performance-intensive applications. Organizations must be sure that it delivers consistent performance with predictable latency throughout the system. Cisco HyperFlex all-NVMe hyperconverged systems offer the following advantages:

- High performance: NVMe nodes deliver the highest performance for mission-critical data center workloads. They provide architectural performance to the edge with NVMe drives connected directly to the CPU rather than through a latency-inducing PCIe switch.
- Ultra-low latency with consistent performance: Cisco HyperFlex all-NVMe systems, when used to host the virtual database instances, deliver extremely low latency and consistent database performance.
- Data protection (fast clones, snapshots, and replication factor): Cisco HyperFlex systems are engineered with robust data protection techniques that enable quick backup and recovery of applications to protect against failures.
- Storage optimization (always-active inline deduplication and compression): All data that comes into Cisco HyperFlex systems is by default optimized using inline deduplication and data compression techniques.
- Dynamic online scaling of performance and capacity: The flexible and independent scalability of the capacity and computing tiers of Cisco HyperFlex systems allow you to adapt to growing performance demands without any application disruption.
- No performance hotspots: The distributed architecture of the Cisco HyperFlex HX Data Platform helps ensure that every virtual machine can achieve the storage IOPS capability and make use of the capacity of the entire cluster, regardless of the physical node on which it resides. This feature is especially important for Oracle Database virtual machines because they frequently need higher performance to handle bursts of application and user activity.
- Nondisruptive system maintenance: Cisco HyperFlex systems support a distributed computing and storage environment that helps enable you to perform system maintenance tasks without disruption.

Several of these features and attributes are particularly applicable to Oracle RAC implementations, including consistent low-latency performance, storage optimization using always-on inline compression, dynamic and seamless performance and capacity scaling, and nondisruptive system maintenance.

Oracle RAC 19c Database on Cisco HyperFlex systems

This reference architecture guide describes how Cisco HyperFlex systems can provide intelligent end-to-end automation with network-integrated hyperconvergence for an Oracle RAC database deployment. Cisco HyperFlex systems provide a high-performance, easy-to-use, integrated solution for an Oracle Database environment.

The Cisco HyperFlex data distribution architecture allows concurrent access to data by reading and writing to all nodes at the same time. This approach provides data reliability and fast database performance. Figure 1 shows the data distribution architecture.

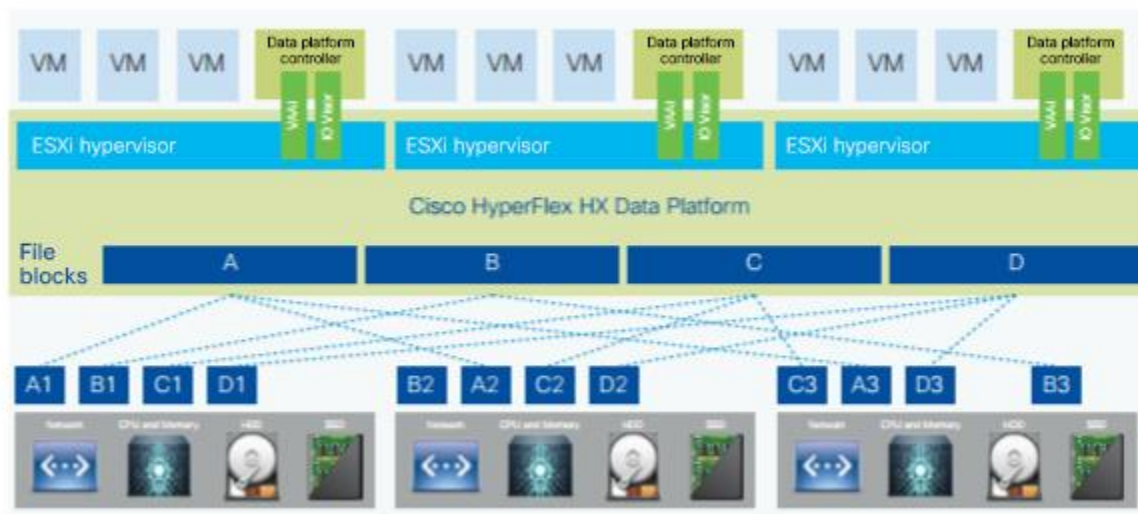


Figure 1.
Data distribution architecture

This reference architecture uses a cluster of four Cisco HyperFlex HX220c M5 All NVMe Nodes to provide fast data access. Use this document to design an Oracle RAC database solution that meets your organization's requirements and budget

This hyperconverged solution integrates servers, storage systems, network resources, and storage software to provide an enterprise-scale environment for Oracle Database deployments. This highly integrated environment provides reliability, high availability, scalability, and performance for Oracle virtual machines to handle large-scale transactional workloads. The solution uses four virtual machines to create a single four-node Oracle RAC database for performance, scalability, and reliability. The RAC node uses the Oracle Enterprise Linux operating system for the best interoperability with Oracle databases.

Cisco HyperFlex systems also support other enterprise Linux platforms such as SUSE and Red Hat Enterprise Linux (RHEL). For a complete list of virtual machine guest operating systems supported for VMware virtualized environments, see the [VMware Compatibility Guide](#).

Oracle RAC with VMware virtualized environment

This reference architecture uses VMware virtual machines to create Oracle RAC with four nodes. Although this solution guide describes a four-node configuration, this architecture can support scalable all-flash Cisco HyperFlex configurations, as well as scalable RAC nodes and scalable virtual machine counts and sizes as needed to meet your deployment requirements.

Note: For best availability, Oracle RAC virtual machines should be hosted on different VMware ESX servers. With this setup, the failure of any single ESX server will not take down more than a single RAC virtual machine and node with it.

Figure 2 shows the Oracle RAC configuration used in the solution described in this document.

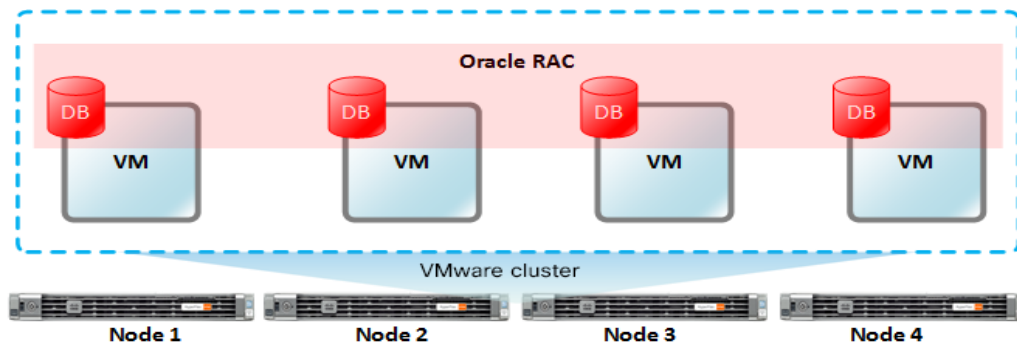


Figure 2.
Oracle Real Application Cluster configuration

Oracle RAC allows multiple virtual machines to access a single database to provide database redundancy while providing more processing resources for application access. The distributed architecture of the Cisco HyperFlex system allows a single RAC node to consume and properly use resources across the Cisco HyperFlex cluster.

The Cisco HyperFlex shared infrastructure enables the Oracle RAC environment to evenly distribute the workload among all RAC nodes running concurrently. These characteristics are critical for any multitenant database environment in which resource allocation may fluctuate.

The Cisco HyperFlex all-NVMe cluster supports large cluster sizes, with the capability to add compute-only nodes to independently scale the computing capacity of the cluster. This approach allows any deployment to start with a small environment and grow as needed, using a pay-as-you-grow model.

This reference architecture document is written for the following audience:

- Database administrators
- Storage administrators
- IT professionals responsible for planning and deploying an Oracle Database solution

To benefit from this reference architecture guide, familiarity with the following is required:

- Hyperconvergence technology
- Virtualized environments
- SSD and flash storage
- Oracle Database 19c
- Oracle Automatic Storage Management (ASM)
- Oracle Enterprise Linux

Oracle Database scalable architecture overview

This section describes how to implement Oracle RAC database on a Cisco HyperFlex system using a four-node cluster. This reference configuration helps ensure proper sizing and configuration when you deploy a RAC database on a Cisco HyperFlex system. This solution enables customers to rapidly deploy Oracle databases by eliminating engineering and validation processes that are usually associated with deployment of enterprise solutions.

Table 1. Oracle virtual machine configuration

Resource	Details for Oracle virtual machine
Virtual machine specifications	24 virtual CPUs (vCPUs) 128 GB of vRAM
Virtual machine controllers	4 × Paravirtual SCSI (PVSCSI) controller
Virtual machine disks	1 × 500-GB VMDK for virtual machine OS 4 × 500-GB VMDK for Oracle data 3 × 70-GB VMDK for Oracle redo log 2 × 80-GB VMDK for Oracle Fast Recovery Area 3 × 40-GB VMDK for Oracle Cluster Ready Services and voting disk

This solution uses virtual machines for Oracle RAC nodes. Table 1 summarizes the configuration of the virtual machines with VMware.

Figure 3 provides a high-level view of the environment.

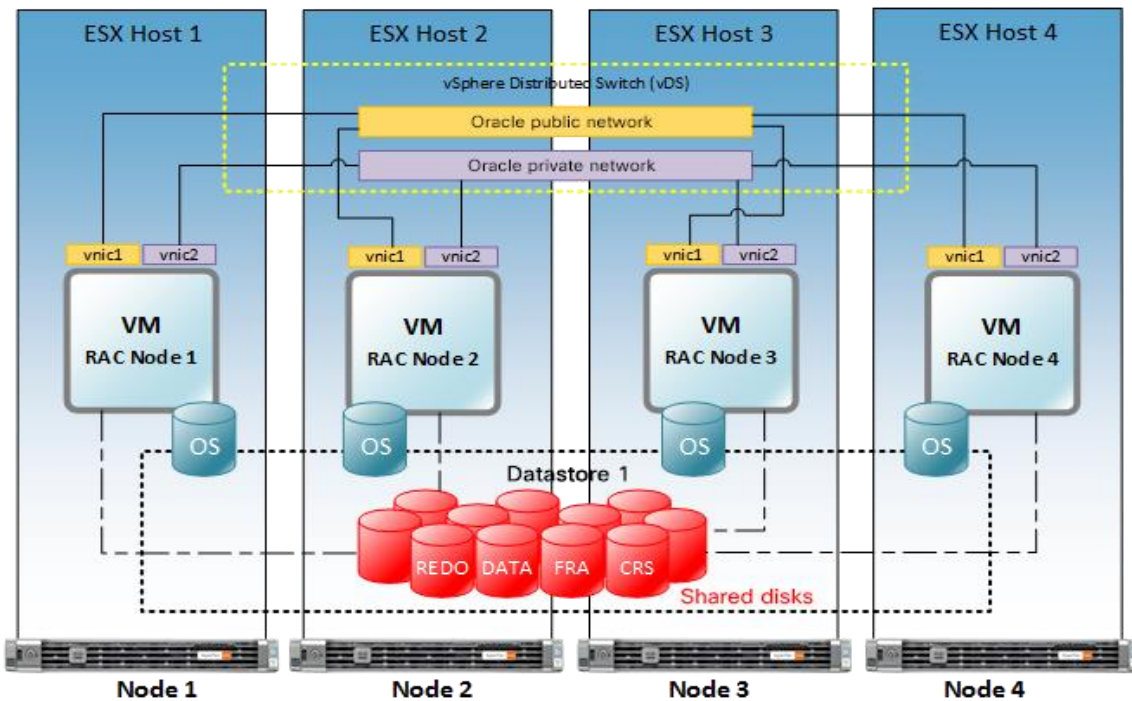


Figure 3.
High-level solution design

Solution components

This section describes the components of this solution. Table 2 summarizes the main components of the solution. Table 3 summarizes the HX220c M5 Node configuration for the cluster.

Hardware components

This section describes the hardware components used for this solution.

Cisco HyperFlex system

The Cisco HyperFlex system provides next-generation hyperconvergence with intelligent end-to-end automation and network integration by unifying computing, storage, and networking resources. The Cisco HyperFlex HX Data Platform is a high performance, flash-optimized distributed file system that delivers a wide range of enterprise-class data management and optimization services. HX Data Platform is optimized for flash memory, reducing SSD wear while delivering high performance and low latency without compromising data management or storage efficiency.

The main features of the Cisco HyperFlex system include:

- Simplified data management
- Continuous data optimization
- Optimization for flash memory
- Independent scaling
- Dynamic data distribution

Visit [Cisco's website](#) for more details about the Cisco HyperFlex HX-Series.

Cisco HyperFlex HX220c M5 All NVMe Nodes

Nodes with all-NVMe storage are integrated into a single system by a pair of Cisco UCS 6200 or 6300 Series Fabric Interconnects. Each node includes two Cisco Flexible Flash (Flex Flash) Secure Digital (SD) cards, a single 120-GB SSD data-logging drive, a single SSD write-log drive and up to six 4-TB NVMe SSD drives, for a contribution of up to 24-TB of raw storage capacity to the cluster. The nodes use Intel Xeon processor 6140 family CPUs and next-generation DDR4 memory and offer 12-Gbps SAS throughput. They deliver significant performance and efficiency gains and outstanding levels of adaptability in a 1-rack-unit (1RU) form factor.

This solution uses four Cisco HyperFlex HX220c M5 All NVMe Nodes for a four-node server cluster to provide two-node failure reliability.

See the [Cisco HyperFlex HX220c M5 All NVMe Node data sheet](#) for more information.

Cisco UCS 6200 Series Fabric Interconnects

The Cisco UCS 6200 Series Fabric Interconnects are a core part of Cisco UCS, providing both network connectivity and management capabilities for the system. The 6200 Series offers line-rate, low-latency, lossless 10 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), and Fibre Channel functions.

The Cisco UCS 6200 Series provides the management and communication backbone for the Cisco UCS B-Series Blade Servers and 5100 Series Blade Server Chassis. All chassis, and therefore all blades, attached to the 6200 Series Fabric Interconnects become part of a single, highly available management domain. In addition, by

supporting unified fabric, the 6200 Series provides both LAN and SAN connectivity for all blades within the domain.

The Cisco UCS 6200 Series uses a cut-through networking architecture, supporting deterministic, low-latency, line-rate 10 Gigabit Ethernet on all ports, switching capacity of 2 terabits (Tb), and bandwidth of 320 Gbps per chassis, independent of packet size and enabled services. The product family supports Cisco low-latency, lossless 10 Gigabit Ethernet unified network fabric capabilities, which increase the reliability, efficiency, and scalability of Ethernet networks. The fabric interconnect supports multiple traffic classes over a lossless Ethernet fabric from the blade through the interconnect. Significant savings in total cost of ownership (TCO) come from an FCoE optimized server design in which network interface cards (NICs), host bus adapters (HBAs), cables, and switches can be consolidated.

Note: Although the testing described here was performed using Cisco UCS 6200 Series Fabric Interconnects, the Cisco HyperFlex HX Data Platform does include support for Cisco UCS 6300 Series Fabric Interconnects, which provide higher performance with 40 Gigabit Ethernet.

Table 2. Reference architecture components

Hardware	Description	Quantity
Cisco HyperFlex HX220c M5 All NVMe Node servers	Cisco 1-rack-unit (1RU) hyperconverged node that allows for cluster scaling with minimal footprint requirements	4
Cisco UCS 6200 Series Fabric Interconnects	Fabric interconnects	2

Table 3. Cisco HyperFlex HX220c M5 Node configuration

Description	Specification	Notes
CPU	2 Intel Xeon Gold 6140 CPUs at 2.30 GHz	
Memory	24 × 32-GB DIMMs	
Cisco Flexible Flash (FlexFlash) Secure Digital (SD) card	240-GB SSD	Boot drives
SSD	500-GB SSD	Configured for housekeeping tasks
	375-GB SSD	Configured as cache
	6 × 4-TB SSD	Capacity disks for each node
Hypervisor	VMware vSphere, 6.5.0	Virtual Platform for HX Data Platform software
Cisco HyperFlex HX Data Platform software	Cisco HyperFlex HX Data Platform Release 4.0(1b)	

Software components

This section describes the software components used for this solution.

VMware vSphere

VMware vSphere helps you get performance, availability, and efficiency from your infrastructure while reducing the hardware footprint and your capital expenditures (CapEx) through server consolidation. Using VMware products and features such as VMware ESX, vCenter Server, High Availability (HA), Distributed Resource Scheduler (DRS), and Fault Tolerance (FT), vSphere provides a robust environment with centralized management and gives administrators control over critical capabilities.

VMware provides product features that can help manage the entire infrastructure:

- vMotion: vMotion allows nondisruptive migration of both virtual machines and storage. Its performance graphs allow you to monitor resources, virtual machines, resource pools, and server utilization.
- Distributed Resource Scheduler: DRS monitors resource utilization and intelligently allocates system resources as needed.
- High Availability: HA monitors hardware and OS failures and automatically restarts the virtual machine, providing cost-effective failover.
- Fault Tolerance: FT provides continuous availability for applications by creating a live shadow instance of the virtual machine that stays synchronized with the primary instance. If a hardware failure occurs, the shadow instance instantly takes over and eliminates even the smallest data loss.

For more information, visit the [VMware website](#).

Oracle Database 19c

Oracle Database 19c now provides customers with a high-performance, reliable, and secure platform to easily and cost-effectively modernize their transactional and analytical workloads on-premises. It offers the same familiar database software running on-premises that enables customers to use the Oracle applications they have developed in-house. Customers can therefore continue to use all their existing IT skills and resources and get the same support for their Oracle databases on their premises.

For more information, visit the [Oracle website](#).

Note: The validated solution discussed here uses Oracle Database 19c Release 3. Limited testing shows no issues with Oracle Database 19c Release 3 or 12c Release 2 for this solution.

Table 4. Reference architecture software components

Software	Version	Function
Cisco HyperFlex HX Data Platform	Release 4.0(1b)	Data platform
Oracle Enterprise Linux Oracle UEK Kernel	Version 7.6 4.14.35-1902.3.1.el7uek.x86_64 x86_64	OS for Oracle RAC Kernel version in Oracle Linux
Oracle Grid and (ASM)	Version 19c Release 3	Automatic storage management
Oracle Database	Version 19c Release 3	Oracle Database system

Software	Version	Function
Oracle Swingbench, Order Entry Workload	Version 2.5	Workload suite
Recovery Manager (RMAN)		Backup and recovery manager for Oracle Database
Oracle Data Guard	Version 19c Release 3	High availability, data protection and disaster recovery for Oracle Database

Storage architecture

This reference architecture uses an all-NVMe configuration. The HX220c M5 All NVMe Nodes allow eight NVMe SSDs. However, two per node are reserved for cluster use. NVMe SSDs from all four nodes in the cluster are striped to form a single physical disk pool. (For an in-depth look at the Cisco HyperFlex architecture, see the Cisco white paper [Deliver Hyperconvergence with a Next-Generation Platform](#). A logical datastore is then created for placement of Virtual Machine Disk (VMDK) disks. The storage architecture for this environment is shown in Figure 4. This reference architecture uses 4-TB NVMe SSDs.

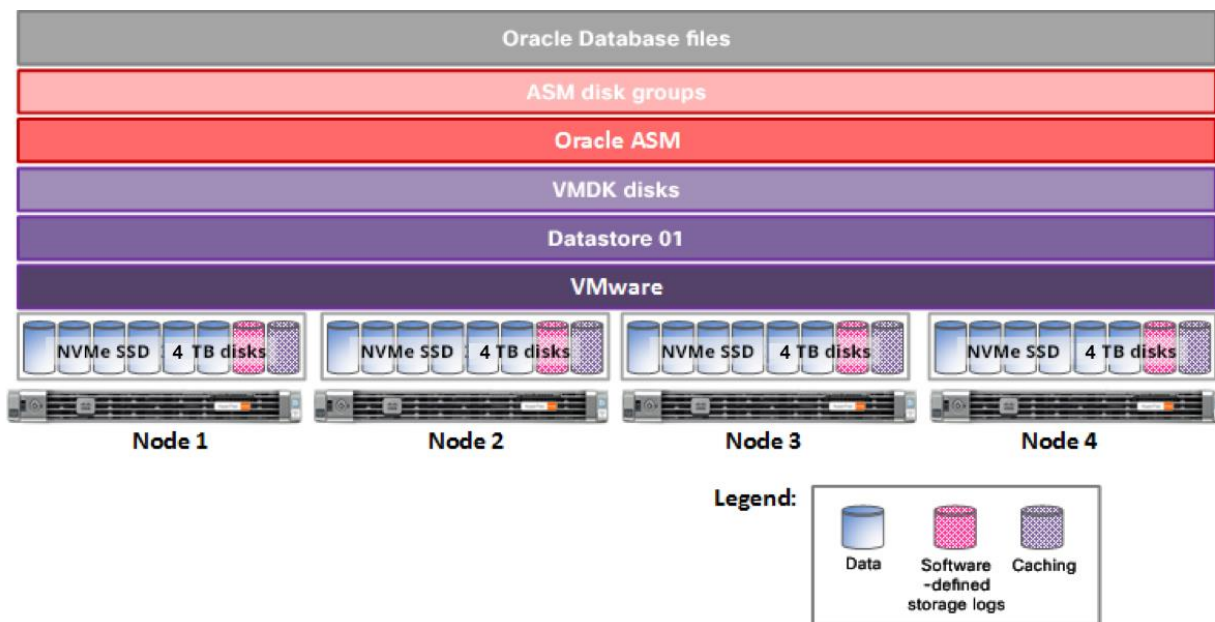


Figure 4.
Storage architecture

Storage configuration

This solution uses VMDK disks to create shared storage that is configured as an Oracle Automatic Storage Management, or ASM, disk group. Because all Oracle RAC nodes must be able to access the VMDK disks concurrently, you should configure the Multi-writer option for sharing in the virtual machine disk configuration. For optimal performance, distribute the VMDK disks to the virtual controller using Table 5 for guidance.

Note: In general both HX and Oracle ASM provides RF factor. But in our test environment, we are only using the RF factor provided by HX not by the Oracle ASM. The capacities vary depending on the RF factor being set (If RF is set to 2, actual capacities are one-half of raw capacity and If RF is set to 3, actual capacities are one-third of raw capacity).

Table 5. Assignment of VMDK disks to SCSI controllers; Storage layout for each virtual machine (all disks are shared with all four Oracle RAC nodes)

SCSI 0 (Paravirtual)	SCSI 1 (Paravirtual)	SCSI 2 (Paravirtual)	SCSI 3 (Paravirtual)
500 GB, OS disk	500 GB, Data1	500 GB, Data2	70 GB, Log1
	500 GB, Data3	500 GB, Data4	70 GB, Log2
	80 GB, FRA1	80 GB, FRA2	70 GB, Log3
	40 GB, CRS1	40 GB, CRS2	
	40 GB, CRS3	2000 GB, RMAN	

Configure the following settings on all VMDK disks shared by Oracle RAC nodes (Figure 5):

The screenshot shows the configuration for a VMDK disk. The following settings are highlighted with red boxes:

- Type:** Thick provision eager zeroed
- Sharing:** Multi-writer
- Disk File:** [RAC] Rac1_01/Rac1_01_5.vmdk
- Shares:** Normal, 1000
- Limit - IOPs:** Unlimited
- Virtual flash read cache:** 0 GB
- Virtual Device Node:** SCSI controller 1, SCSI(1:2)
- Disk Mode:** Independent - Per...

Figure 5. Settings for VMDK disks shared by Oracle RAC nodes

For additional information about the Multi-writer option and the configuration of shared storage, see this [VMWare knowledgebase article](#).

Table 6 summarizes the Cisco ASM disk groups for this solution that are shared by all Oracle RAC nodes.

Table 6. Oracle ASM disk groups

Oracle ASM disk group	Purpose	Stripe size	Capacity
DATA-DG	Oracle database disk group	4 MB	2000 GB
REDO-DG	Oracle database redo group	4 MB	210 GB
CRS-DG	Oracle RAC Cluster Ready Service disk group	4 MB	120 GB
FRA-DG	Oracle Fast Recovery Area disk group	4 MB	160 GB

Oracle Database configuration

This section describes the Oracle Database configuration for this solution. Table 7 summarizes the configuration details.

Table 7. Oracle Database configuration

Settings	Configuration
SGA_TARGET	96 GB
PGA_AGGREGATE_TARGET	30 GB
Data files placement	ASM and DATA DG
Log files placement	ASM and REDO DG
Redo log size	30 GB
Redo log block size	4 KB
Database block	8 KB

Network configuration

The Cisco HyperFlex network topology consists of redundant Ethernet links for all components to provide the highly available network infrastructure that is required for an Oracle Database environment. No single point of failure exists at the network layer. The converged network interfaces provide high data throughput while reducing the number of network switch ports. Figure 6 shows the network topology for this environment.

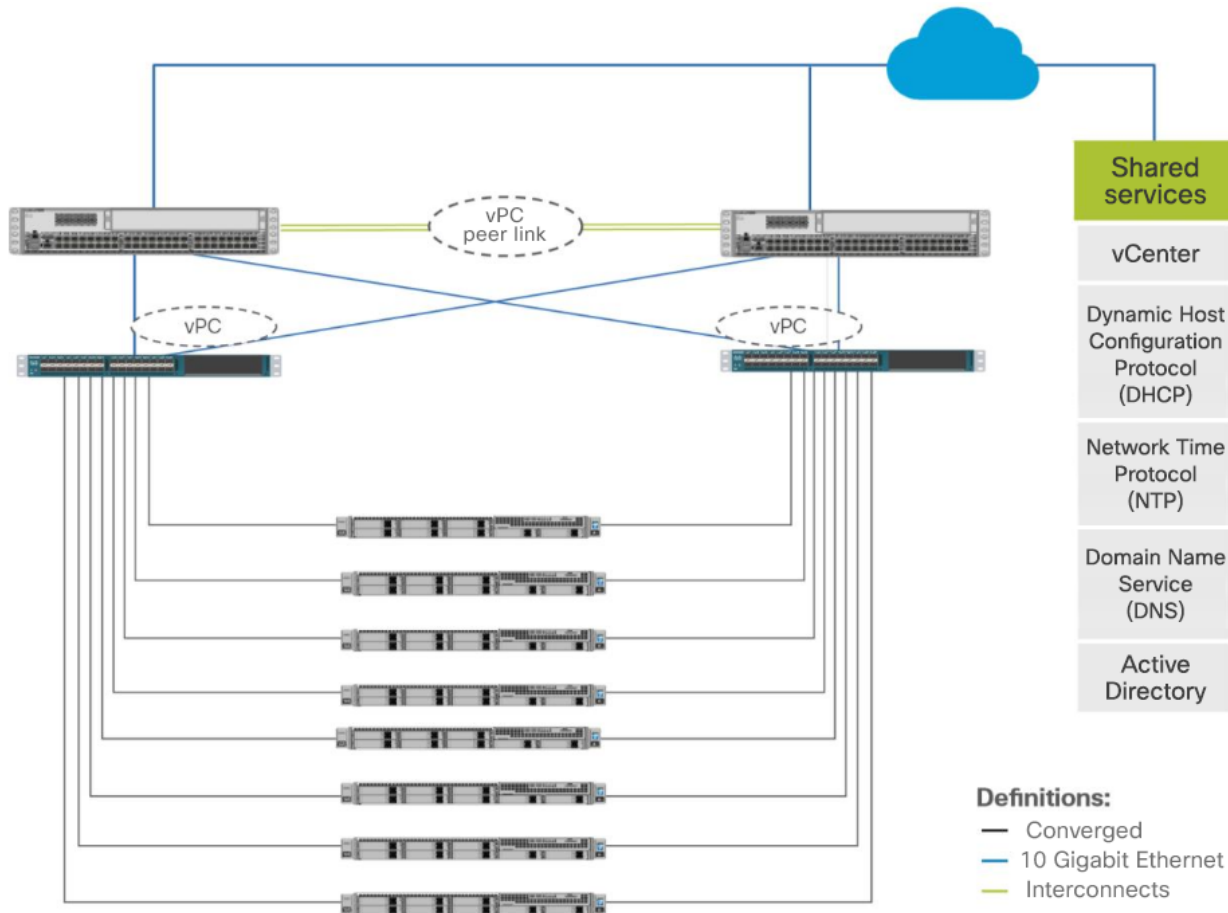


Figure 6.
Network topology

Storage configuration

For most deployments, a single datastore for the cluster is sufficient, resulting in fewer objects that need to be managed. The Cisco HyperFlex HX Data Platform is a distributed file system that is not vulnerable to many of the problems that face traditional systems that require data locality. A VMDK disk does not have to fit within the available storage of the physical node that hosts it. If the cluster has enough space to hold the configured number of copies of the data, the VMDK disk will fit because the HX Data Platform presents a single pool of capacity that spans all the hyperconverged nodes in the cluster. Similarly, moving a virtual machine to a different node in the cluster is a host migration; the data itself is not moved.

In some cases, however, additional datastores may be beneficial. For example, an administrator may want to create an additional HX Data Platform datastore for logical separation. Because performance metrics can be filtered to the data-store level, isolation of workloads or virtual machines may be desired. The datastore is thinly provisioned on the cluster. However, the maximum datastore size is set during data-store creation and can be used to keep a workload, a set of virtual machines, or end users from running out of disk space on the entire cluster and thus affecting other virtual machines. In such scenarios, the recommended approach is to provision the entire virtual machine, including all its virtual disks, in the same datastore and to use multiple datastores to separate virtual machines instead of provisioning virtual machines with virtual disks spanning multiple datastores.

Another good use for additional datastores is to assist in throughput and latency in high-performance Oracle deployments. If the cumulative IOPS of all the virtual machines on a VMware ESX host surpasses 10,000 IOPS,

the system may begin to reach that queue depth. In ESXTOP, you should monitor the Active Commands and Commands counters, under Physical Disk NFS Volume. Dividing the virtual machines into multiple datastores can relieve the bottleneck. The default value for ESX queue depth per datastore on a Cisco HyperFlex system is 1024.

Another place at which insufficient queue depth may result in higher latency is the SCSI controller. Often the queue depth settings of virtual disks are overlooked, resulting in performance degradation, particularly in high-I/O workloads. Applications such as Oracle Database tend to perform a lot of simultaneous I/O operations, resulting in virtual machine driver queue depth settings insufficient to sustain the heavy I/O processing (the default setting is 64 for PVSCSI). Hence, the recommended approach is to change the default queue depth setting to a higher value (up to 254) as suggested in this [VMware knowledgebase article](#).

For large-scale and high-I/O databases, you always should use multiple virtual disks and distribute those virtual disks across multiple SCSI controller adapters rather than assigning all of them to a single SCSI controller. This approach helps ensure that the guest virtual machine accesses multiple virtual SCSI controllers (four SCSI controllers maximum per guest virtual machine), thus enabling greater concurrency using the multiple queues available for the SCSI controllers.

Paravirtual SCSI queue depths settings

Large-scale workloads with intensive I/O patterns require adapter queue depths greater than the PVSCSI default values. Current PVSCSI queue depth default values are 64 (for devices) and 254 (for adapters). You can increase the PVSCSI queue depths to 254 (for devices) and 1024 (for adapters) in a Microsoft Windows or Linux virtual machine.

The following parameters were configured in the design discussed in this document:

- `vmw_pvscsi.cmd_per_lun=254`
- `vmw_pvscsi.ring_pages=32`

For additional information about PSCSI queue depth settings, see this [VMware knowledgebase article](#).

Engineering validation

The performance, functions, and reliability of this solution were validated while running Oracle Database in a Cisco HyperFlex environment. The Oracle Swingbench test kit was used to create and test the Order Entry workload, an OLTP-equivalent database workload.

Performance testing

This section describes the results that were observed during the testing of this solution. To better understand the performance of each area and component of this architecture, each component was evaluated separately to help ensure that optimal performance was achieved when the solution was under stress. The transactions-per-minute (TPM) metric in the Swingbench benchmark kit was used to measure the performance of Oracle Database.

These results are presented to provide some data points for the performance observed during the testing. They are not meant to provide comprehensive sizing guidance. For proper sizing of Oracle or other workloads, please use the Cisco HyperFlex Sizer available at <https://hyperflexsizer.cloudapps.cisco.com/>.

Oracle RAC node scale test

The node scale test validates the scalability of the Oracle RAC cluster when running the Swingbench test with 200 users. The scale testing consists of four tests using four RAC nodes. The test was run for 60 minutes. The TPM average for the test is reported in Table 8.

Table 8. Oracle node scale test

	Average TPM
4 nodes	989150

Oracle user scale test

The user scale test validates the capability of Oracle RAC to scale with user count. The test starts with 160 users and increases to 800 users. This test shows TPM gains as the user count increases, with the four-node Cisco HyperFlex HX220c All Flash cluster starting to saturate toward the higher end of the user count (Figure 7).

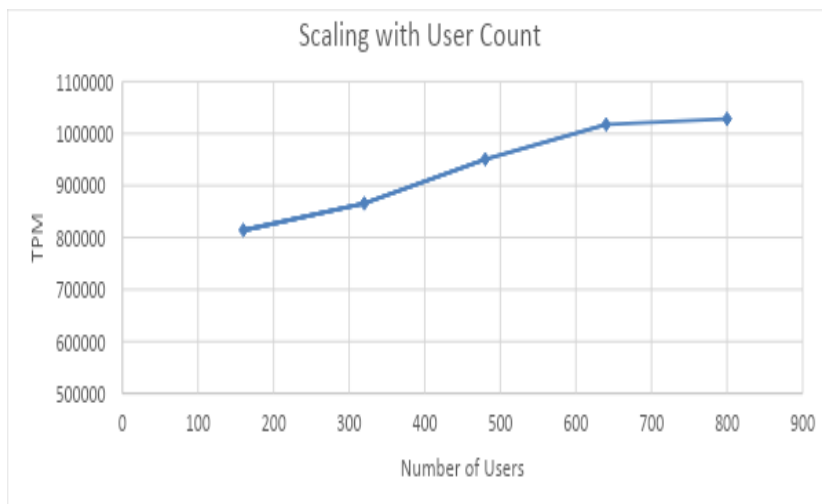


Figure 7.
Oracle user scale test

Performance comparison between all-NVMe and all-flash clusters

Increasing performance at the end-user level requires a holistic approach in designing your solutions. Simply adding low-latency storage is not enough. Therefore, Cisco HyperFlex systems have been designed to provide balanced high performance with very low latency. Configurations using all-flash and all-NVMe nodes, together with our standard high-throughput network and fast computing, supports consistently high performance even for large databases. The distributed architecture provides every virtual machine with access to high IOPS regardless of the physical location of the virtual machine. This capability is important for virtual machines running Oracle Database because they frequently need higher performance to handle bursts of application or user activity. All-NVMe storage performs even faster (as shown by the testing reported here) and is excellent for databases that require ultra-low latency.

Table 9 compares the performance of the two cluster types.

Table 9. Comparison matrix

	All-NVMe test setup	All-flash test setup
Cisco HyperFlex HX Data Platform	Release 4.0(1b)	Release 2.1(1b)
Cisco HyperFlex servers	HX220c M5 All NVMe Node servers	HX220c M4 All Flash Node servers
Oracle Database	Version 19c Release 3	Version 12c Release 2
Oracle SLOB	Version 2.4	Version 2.3
Oracle Grid and ASM	Version 19c Release 3	Version 12c Release 2
Replication factor	3	2
Hypervisor	Version 6.5.0	Version 6.0

See the [Cisco HyperFlex and Oracle RAC white paper about all-flash setup](#) for detailed information.

Oracle RAC node scale test

By comparing the results of the all-NVMe and all-flash clusters in the graph in Figure 8, you can see that NVMe storage delivers higher performance than the all-flash cluster. To handle larger databases with consistently high performance at a very low latency, NVMe storage is always preferable to an all-flash cluster. The node scale test validates the capability of Oracle RAC to scale with user count. The test starts with 160 users, then increases to 480 users, and then increases to 800 users. The graph in Figure 8 shows marginal gains in TPM with the all-NVMe cluster compared to the all-flash cluster.

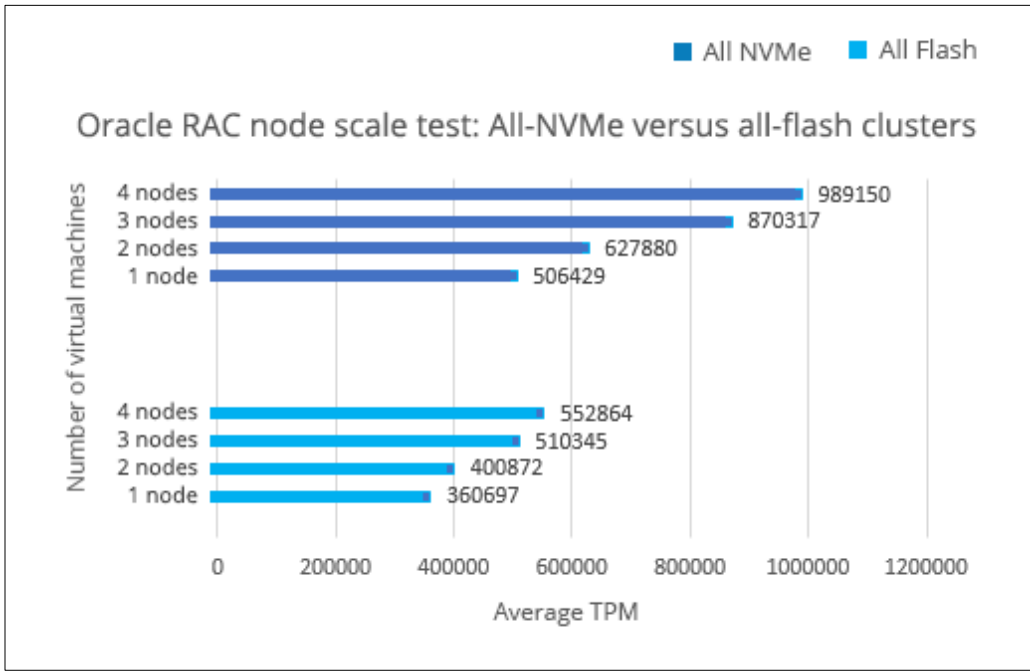


Figure 8.
Oracle RAC node scale test results

Oracle RAC user scale test

The user scale test validates the capability of Oracle RAC to scale with user count. The test starts with 160 users, then increases to 480 users, and then increases to 800 users. As in the node scale testing, NVMe storage delivers higher performance than the all-flash cluster. The graph in Figure 9 shows marginal gains in TPM with the all-NVMe cluster compared to the all-flash cluster.

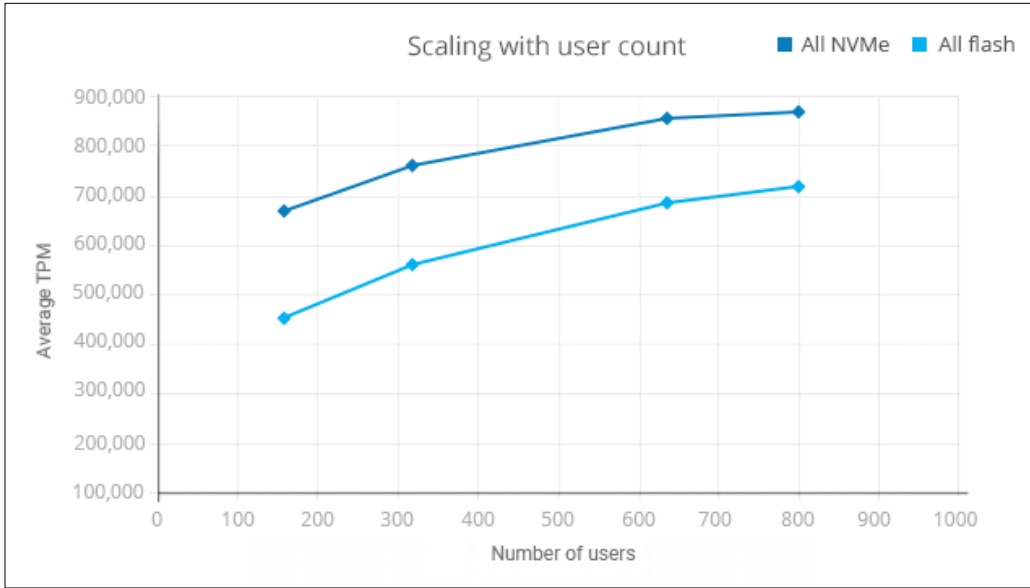


Figure 9.
Oracle RAC user scale test results

Reliability and disaster recovery

This section describes some additional reliability and disaster-recovery options available for use with Oracle RAC databases.

Disaster recovery

Oracle RAC is usually used to host mission-critical solutions that require continuous data availability to prevent planned or unplanned outages in the data center. Oracle Data Guard is an application-level disaster-recovery feature that can be used with Oracle RAC.

The Cisco HyperFlex HX Data Platform also provides disaster recovery. Note, though, that HX Data Platform cannot use this feature to protect Oracle RAC virtual machines as RAC uses multiwriter option which isn't supported by this feature.

Figure 10 shows a typical disaster-recovery setup. It uses two HX Data Platform clusters: one for the local site and one for the remote site over a WAN connection

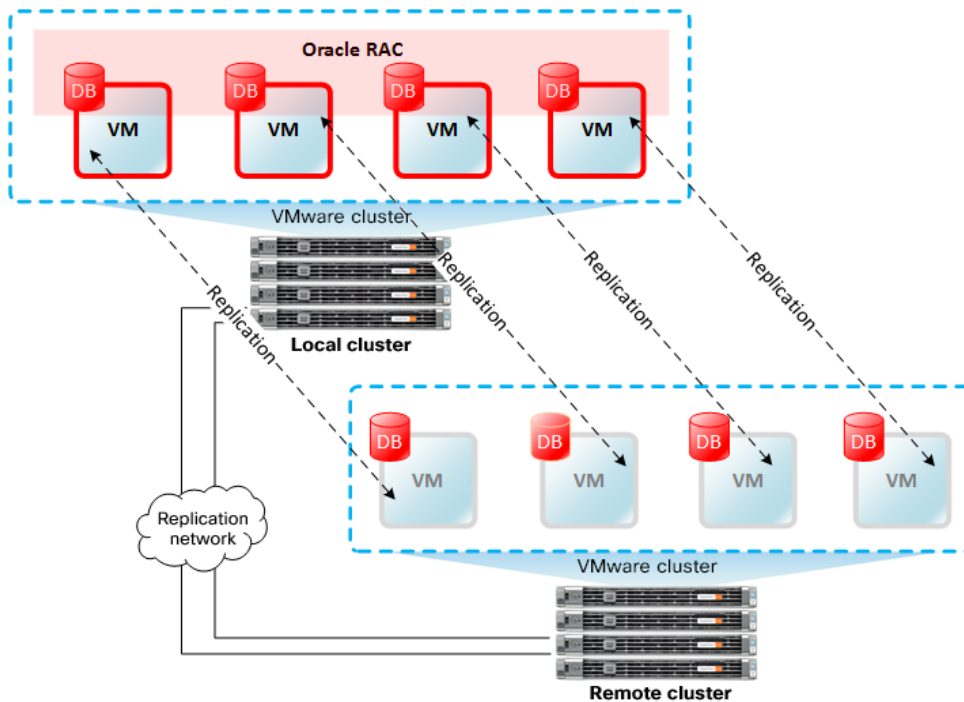


Figure 10.
Disaster-recovery configuration

Cisco HyperFlex HX Data Platform disaster recovery

The Cisco HyperFlex HX Data Platform disaster-recovery feature allows you to protect virtual machines from a disaster by setting up replication by running virtual machines between a pair of network-connected clusters. Protected virtual machines running on one cluster replicate to the other cluster in the pair, and vice versa. The two paired clusters typically are located at a distance from each other, with each cluster serving as the disaster-recovery site for virtual machines running on the other cluster.

Oracle Data Guard

Oracle Data Guard helps ensure high availability, data protection, and disaster recovery for enterprise data. It provides a comprehensive set of services that create, maintain, manage, and monitor one or more standby databases to enable production Oracle databases to survive disasters and data corruptions. Data Guard maintains these standby databases as copies of the production database. Then, if the production database becomes unavailable because of a planned or an unplanned outage, Data Guard can switch any standby database to the production role, reducing the downtime associated with the outage.

Data Guard can be used with traditional backup, restoration, and cluster techniques to provide a high level of data protection and data availability. Data Guard transport services are also used by other Oracle features such as Oracle Streams and Oracle GoldenGate for efficient and reliable transmission of redo logs from a source database to one or more remote destinations

Oracle RAC database backup and recovery: Oracle Recovery Manager

Oracle Recovery Manager (RMAN) is an application-level backup and recovery feature that can be used with Oracle RAC. It is a built-in utility of Oracle Database. It automates the backup and recovery processes. Database administrators (DBAs) can use RMAN to protect data in Oracle databases. RMAN includes block-level corruption detection during both backup and restore operations to help ensure data integrity.

In the lab environment, RMAN was configured on one RAC node for testing. A dedicated 2000-GB VMDK disk was configured as the backup repository under the `/backup` mount point. By default, RMAN creates backups on disk and generates backup sets rather than image copies. Backup sets can be written to disk or tape.

The RMAN test was performed running a 200-user Swingbench test workload on the Oracle RAC database. The backup operation did not affect the TPM count in the test. However, CPU use on Node 1 increased as a result of the RMAN processes.

Figure 11 shows the RMAN environment for the backup testing. The figure shows a typical setup. In this case, the backup target is the mount point `/backup`.

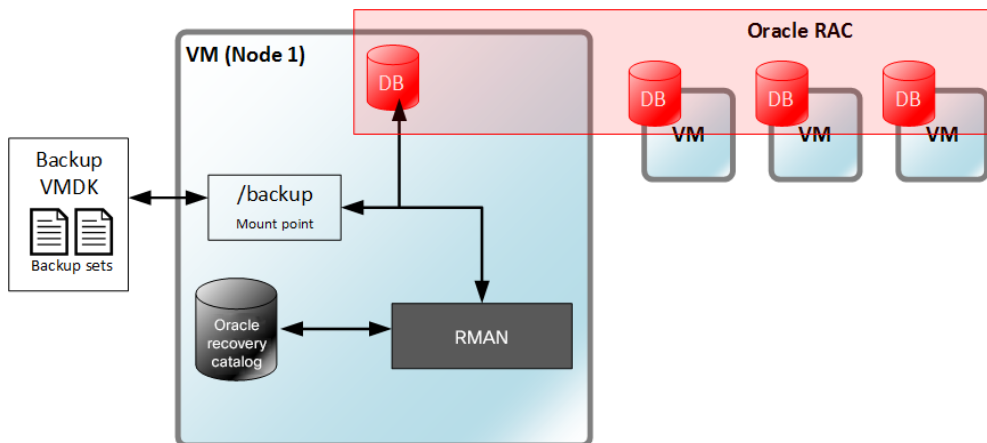


Figure 11.
RMAN test environment

Two separate RMAN backup tests were run: one while the database was idle with no user activity, and one while the database was under active testing with 200 user sessions. Table 10 shows the test results.

Table 10. RMAN backup results

	During idle database	During active Swingbench testing
Backup type	Full backup	Full backup
Backup size	204 GB	204 GB
Backup elapsed time	00:07:45	
Backup size	204 GB	
Backup throughput	130 MBps	336 MBps

Table 11 shows the average TPM for two separate Swingbench tests: one when RMAN was used to perform a hot backup of the database, and one for a Swingbench test with no RMAN activity. Test results show little impact on the Oracle RAC database during the backup operation.

Table 11. Impact of RMAN backup operation

	Average TPM	Average TPS
RMAN hot backup during Swingbench test	671901	11499
Swingbench only	707613	12076

Conclusion

Cisco is a leader in the data center industry. Our extensive experience with enterprise solutions and data center technologies enables us to design and build an Oracle RAC database reference architecture on hyperconverged solutions that is fully tested, protecting our customer's investment and offering a high-level ROI. The Cisco HyperFlex architecture helps enable databases to achieve optimal performance with very low latency—features that are critical for enterprise-scale applications.

Cisco HyperFlex systems provide the design flexibility needed to engineer a highly integrated Oracle Database system to run enterprise applications that use industry best practices for a virtualized database environment. The balanced, distributed data-access architecture of Cisco HyperFlex systems supports Oracle scale-out and scale-up models, reducing the hardware footprint and increasing data center efficiency.

As the amount and types of data increase, flexible systems with predictable performance are needed to address database sprawl. By deploying a Cisco HyperFlex all-NVMe configuration, you can run your database deployments on an agile platform that delivers insight in less time and at less cost.

The comparison graphs provided in the [“Engineering validation”](#) section of this document show that the all-NVMe storage configuration has higher performance capabilities at ultra-low latency compared to the all-flash cluster configuration. Thus, to handle Oracle OLTP workloads at very low latency, the all-NVMe solution is preferred over the all-flash configuration.

This solution delivers many business benefits, including the following:

- Increased scalability
- High availability
- Reduced deployment time with a validated reference configuration
- Cost-based optimization
- Data optimization
- Cloud-scale operation

Cisco HyperFlex systems provide the following benefits for this reference architecture:

- Optimized performance for transactional applications with little latency
- Balanced and distributed architecture that increases performance and IT efficiency with automated resource management
- Capability to start with a smaller investment and grow as business demand increases
- Enterprise application-ready solution
- Efficient data storage infrastructure
- Scalability

For more information

For additional information, consult the following resources:

- Cisco HyperFlex white paper: Deliver Hyperconvergence with a Next-Generation Data Platform:
<https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/white-paper-c11-736814.pdf>
- Cisco HyperFlex systems solution overview:
<https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/solution-overview-c22-736815.pdf>
- Oracle Databases on VMware Best Practices Guide, Version 1.0, May 2016:
<http://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/solutions/vmware-oracle-databases-on-vmware-best-practices-guide.pdf>
- Cisco HyperFlex All NVMe at-a-glance:
<https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/le-69503-aag-all-nvme.pdf>
- Hyperconvergence for Oracle: Oracle Database and Real Application Clusters:
<https://www.cisco.com/c/dam/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/le-60303-hxsql-aag.pdf>

Americas Headquarters

Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters

Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters

Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at <https://www.cisco.com/go/offices>.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. To view a list of Cisco trademarks, go to this URL: <https://www.cisco.com/go/trademarks>. Third-party trademarks mentioned are the property of their respective owners. The use of the word partner does not imply a partnership relationship between Cisco and any other company. (1110R)