White paper
Cisco public

The bridge to possible

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

Contents

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. In this reference paper, Cisco HyperFlex (Cisco HyperFlex HX Data Platform [HXDP], Release 5.0) uses the iSCSI Support to enable shared disk access across 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 industry best practices for Oracle Databases in a VMware virtualized environment. Get additional details about deploying Oracle RAC on VMware.

Cisco HyperFlex HX Data Platform Overview (All-NVMe storage) with iSCSI Support

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 white paper, Deliver Hyperconvergence with a Next-Generation Platform.

Cisco HyperFlex Data Platform (HX Data Platform, or HXDP) is a hyperconverged software appliance that transforms Cisco® servers into a single pool of compute and storage resources. It eliminates the need for network storage and enables seamless interoperability between computing and storage in virtual environments. The HX Data Platform provides a highly fault-tolerant distributed storage system that preserves data integrity and optimizes performance for Virtual Machine (VM) storage workloads. In addition, native compression and deduplication reduce storage space occupied by the VMs and VM workloads. HX Data Platform has many integrated components. These include Cisco Fabric Interconnects (FIs), Cisco UCS® Manager, Cisco HyperFlex–specific servers, and Cisco compute-only servers; VMware vSphere, ESXi servers, and vCenter; and the Cisco HX Data Platform Installer, controller VMs, HX Connect, vSphere HX Data Platform Plug-in, and stcli commands. HX Data Platform is installed on a virtualized platform such as VMware vSphere. During installation, after specifying the Cisco HyperFlex HX Storage Cluster name, the HX Data Platform creates a hyperconverged storage cluster on each of the nodes. As your storage needs to increase and you add nodes in the HX cluster, the HX Data Platform balances the storage across the additional resources. Compute-only nodes can be added to increase compute-only resources to the storage cluster. For a complete list of new features and the latest release document, see the Cisco HX Data Platform, Release 5.0.

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 HX240c M6All NVMe Nodes. The result is the first fully engineered hyperconverged appliance based on NVMe storage.

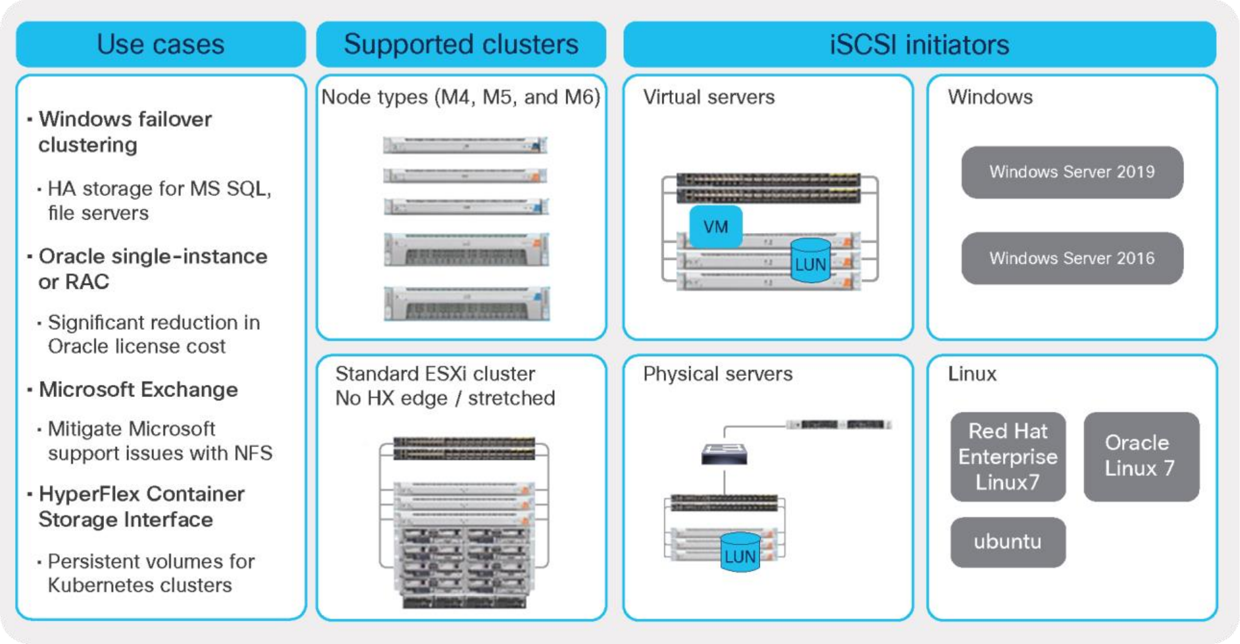
- **Capacity storage** – The data platform's capacity layer is supported by Intel® 3D NAND NVMe Solid-State Disks (SSDs). These drives currently provide up to 32 TB of raw capacity per node. Integrated directly into the CPU through the PCI Express (PCIe) bus, they eliminate the latency of disk controllers 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 Processors, 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 HyperFlex Connect management software.

All-NVMe solutions support most latency-sensitive applications with the simplicity of hyperconvergence. Our solutions provide the industry's first fully integrated platform designed to support NVMe technology with increased performance and RAS. This document uses a 4-node Ice Lake-based Cisco HyperFlex cluster.

Native iSCSI Protocol Support

Cisco HyperFlex HX Data Platform Release 5.0(1b) supports native iSCSI protocol for workloads that require block storage (for example, databases) or **shared disk access** (for example, failover clusters). For more details, see the Cisco HyperFlex HX Data Platform, Release 5.0 document.

Note: HyperFlex HX Data Platform Release 4.5(xx) also supports iSCSI protocol, but with latest M6 servers, it is mandatory to use the latest HXDP 5.x or higher release.



Use cases

- **Windows failover clustering**
  - HA storage for MS SQL, file servers
- **Oracle single-instance or RAC**
  - Significant reduction in Oracle license cost
- **Microsoft Exchange**
  - Mitigate Microsoft support issues with NFS
- **HyperFlex Container Storage Interface**
  - Persistent volumes for Kubernetes clusters

Supported clusters

Node types (M4, M5, and M6)

Standard ESXi cluster
No HX edge / stretched

iSCSI initiators

Virtual servers

Physical servers

Windows

Windows Server 2019

Windows Server 2016

Linux

Red Hat Enterprise Linux7

Oracle Linux 7

ubuntu

Figure 1.
iSCSI support in Cisco HyperFlex HX Data Platform Release 5.0(1b)

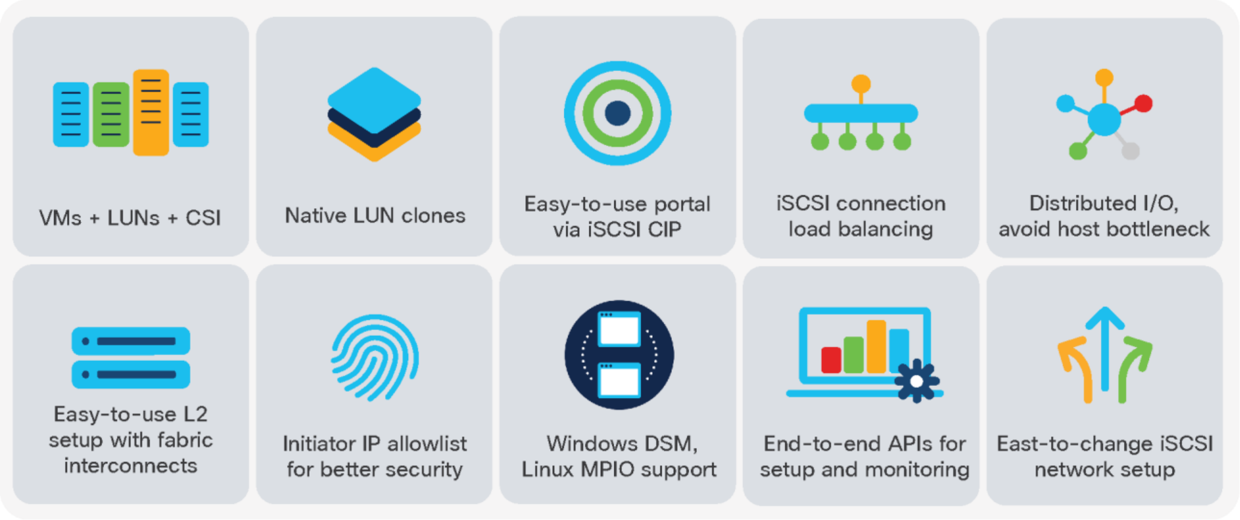

VMs + LUNs + CSI
Native LUN clones
Easy-to-use portal via iSCSI CIP
iSCSI connection load balancing
Distributed I/O, avoid host bottleneck
Easy-to-use L2 setup with fabric interconnects
Initiator IP allowlist for better security
Windows DSM, Linux MPIO support
End-to-end APIs for setup and monitoring
East-to-change iSCSI network setup

Figure 2.
Cisco HyperFlex iSCSI key features

Table 1. iSCSI support scale limits in Cisco HyperFlex HX Data Platform (HXDP) Release 5.0(1b)

Scale item	HXDP 4.5(1a) iSCSI support
iSCSI LUNs per HyperFlex cluster	32,768
iSCSI targets per HyperFlex cluster	128
iSCSI LUNs per target	256
Maximum iSCSI LUN size	64 TB
Maximum number of iSCSI sessions per controller VM	64
iSCSI I/O initiator queue depth per iSCSI Session	256
iSCSI I/O target side queue depth per controller	2048

For more details, refer to Cisco HyperFlex HX Data Platform release notes for latest information on scale limits.

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 optimized by default using inline-deduplication and data-compression.
- **Dynamic online scaling of performance and capacity** – The flexible and independent scalability of the capacity and computing tiers of Cisco HyperFlex systems allows 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 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. And with the **iSCSI support** included in the latest Cisco HyperFlex Data Platform release 5.0(1b), Oracle RAC clusters should be able to use the “**shared disk access**” feature, which allows all RAC instances to access the same storage by **eliminating the need to enable “multi-writer” option** on all the RAC virtual machines (an approach that was followed while testing with NFS storage).

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 3 shows the data distribution architecture.

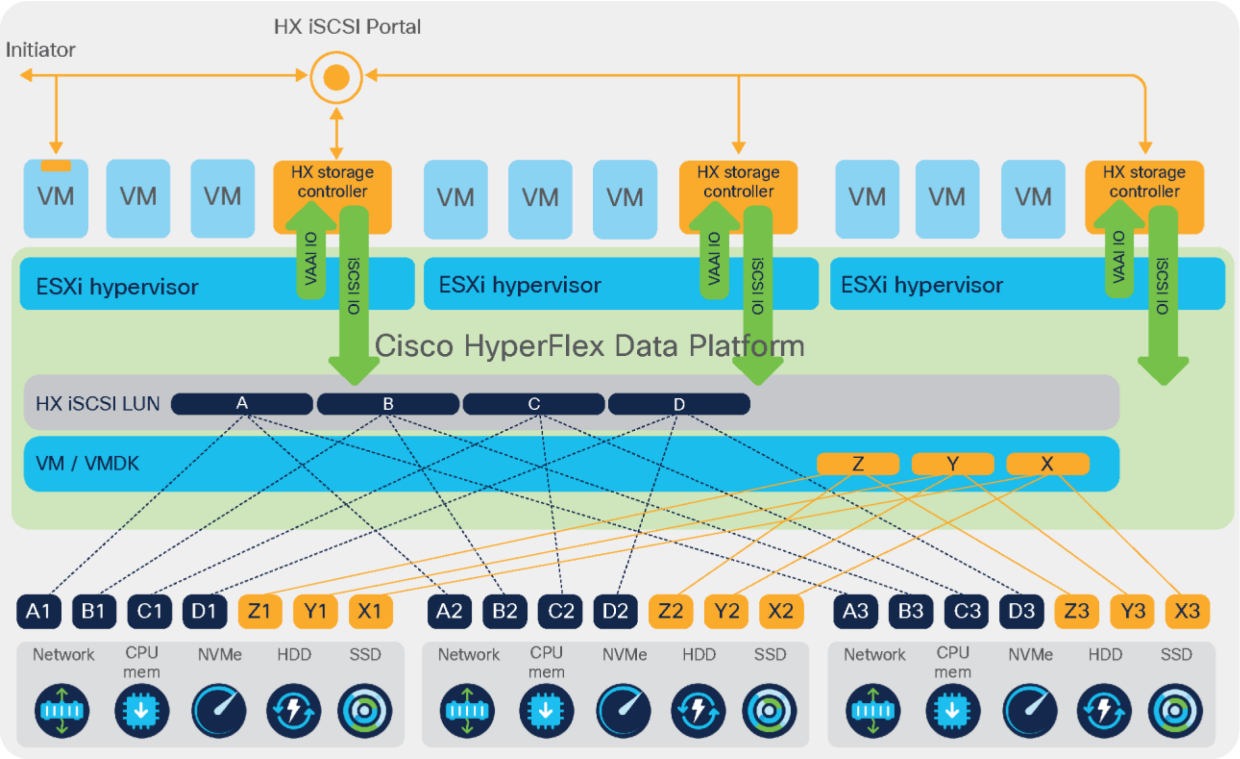


Figure 3.
Data distribution architecture

This reference architecture uses a cluster of four Cisco HyperFlex HX240C-M6SN All NVMe Nodes to provide fast data access. Use this document to design an Oracle RAC database 19c 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 an Oracle Database deployment. 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 a VMware virtualized environment

This reference architecture uses VMware virtual machines to create an Oracle RAC cluster with four nodes. Although this solution guide describes a 4-node configuration, this architecture can support scalable all-NVMe 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 4 shows the Oracle RAC configuration used in the solution described in this document.

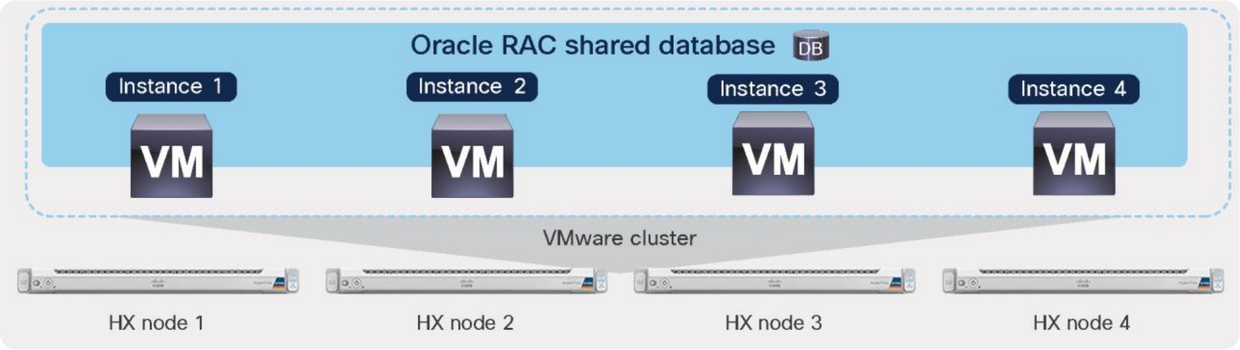


Figure 4.
Oracle Real Application Cluster configuration

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 4-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.

This solution uses virtual machines for Oracle RAC nodes. Table 2 summarizes the configuration of the virtual machines with VMware.

Table 2. Oracle virtual machine configuration

Resource	Details for Oracle virtual machine
Virtual machine specifications	24 virtual CPUs (vCPUs) 150 GB of vRAM
Virtual machine iSCSI drives	1 × 500-GB LUN for virtual machine OS 4 × 500-GB LUNs for Oracle data 3 × 70-GB LUNs for Oracle redo log 2 × 80-GB LUNs for Oracle Fast Recovery Area 3 × 40-GB LUNs for Oracle Cluster-Ready Services and voting disk

Figure 5 provides a high-level view of the environment.

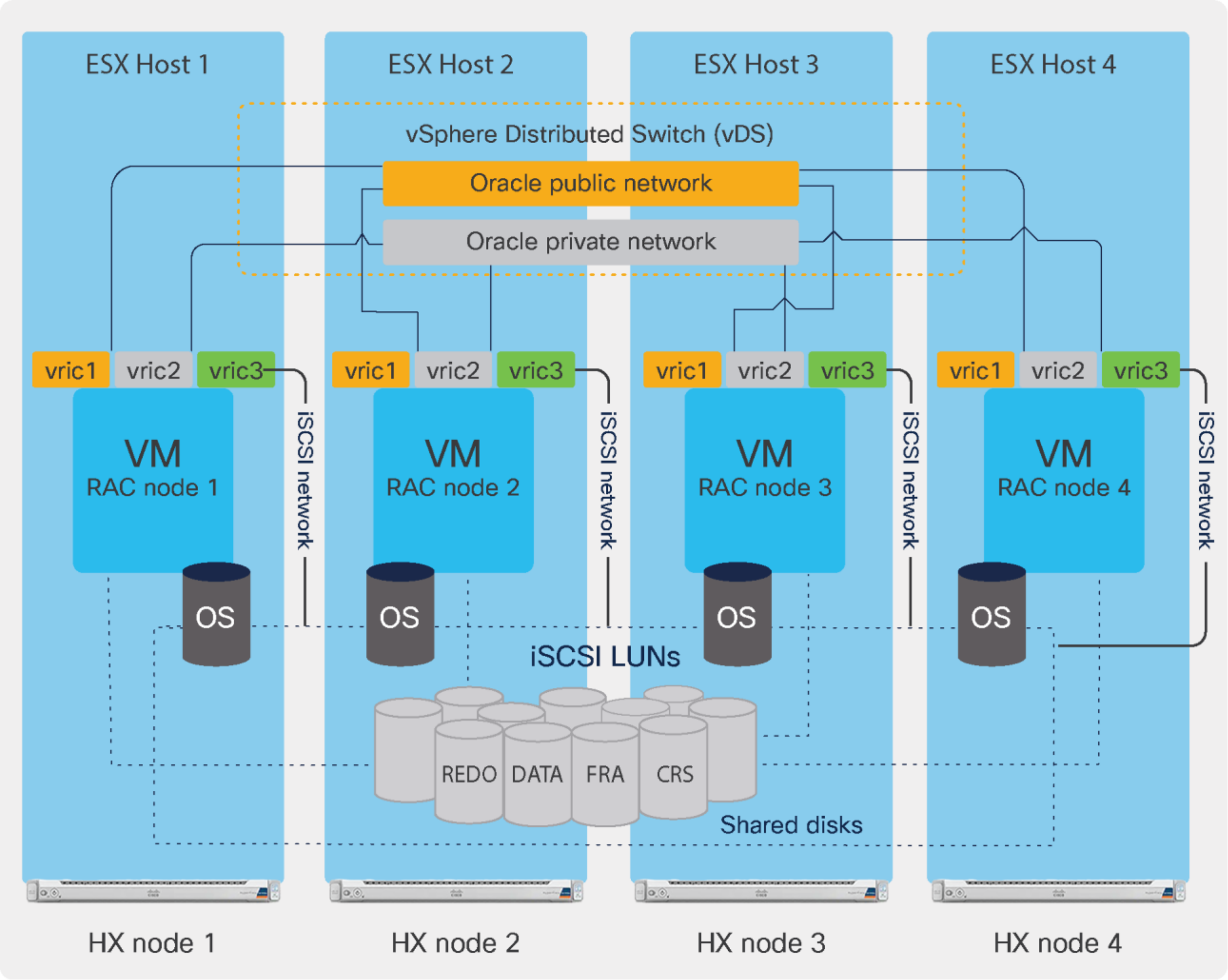


Figure 5.
High-level solution design

Solution components

This section describes the components of this solution. Table 3 summarizes the main components of the solution. Table 4 summarizes the Cisco HyperFlex HX240C-M6SN All NMVe 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. The 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 HX240C-M6SN All NVMe Nodes

Nodes with all-NVMe storage are integrated into a single system by a pair of Cisco UCS 6400 or 6300 series fabric interconnects. Each node includes an M2 boot drive (240GB), an NVMe drive (1TB) as a data-logging drive, a single OptaneNVMe SSD (375 GB) serving as a cachedrive, and up to six 3.8-TB NVMe SSD drives, for a contribution of up to 22.8TB of raw storage capacity. The nodes use the Intel Xeon Platinum 8368 processor family with Ice Lake-based CPUs and next-generation DDR4 memory and offer 12-Gbps SAS throughput. They deliver significant performance and efficiency gains as well as outstanding levels of adaptability in a 2-rack-unit (2RU) form factor.

This solution uses four Cisco HyperFlex HX240C-M6SN All NVMe Nodes for a four-node server cluster to provide two-node failure reliability when the Replication Factor (RF) is set to 3.

See the Cisco HyperFlex HX240c M6 All NVMe Node data sheet for more information.

Cisco UCS 6400 Series Fabric Interconnects

The Cisco UCS 6400 Series Fabric Interconnects are a core part of the Cisco Unified Computing System, providing both network connectivity and management capabilities for the system. The Cisco UCS 6400 Series offer line-rate, low-latency, lossless 10/25/40/100 Gigabit Ethernet, Fibre Channel over Ethernet (FCoE), and Fibre Channel functions. The Cisco UCS 6400 Series Fabric Interconnects provide the management and communication backbone for the Cisco UCS B-Series Blade Servers, Cisco UCS 5108 B-Series Server Chassis, Cisco UCS C-Series Rack Servers, and Cisco UCS S-Series Storage Servers. All servers attached to a Cisco UCS 6400 Series Fabric Interconnect become part of a single, highly available management domain. In addition, by supporting a unified fabric, Cisco UCS 6400 Series Fabric Interconnects provide both LAN and SAN connectivity for all servers within their domain. From a networking perspective, the Cisco UCS 6400 Series Fabric Interconnects use a cut-through architecture, supporting deterministic, low-latency, line-rate 10/25/40/100 Gigabit Ethernet ports, switching capacity of 3.82 Tbps for the 6454, 7.42 Tbps for the 64108, and 200 Gbps bandwidth between the 6400 Series fabric interconnect and the Cisco UCS I/O Module 2408 (IOM 2408) per UCS 5108 blade chassis, independent of packet size and enabled services. The product family supports Cisco low-latency, lossless 10/25/40/100 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 server through the fabric interconnect. Significant TCO savings come from an FCoE-optimized server design in which Network Interface Cards (NICs), Host Bus Adapters (HBAs), cables, and switches can be consolidated.

Table 3. Reference architecture components

Hardware	Description	Quantity
Cisco HyperFlex HX240C-M6SN All NVMe Nodes	Two-rack-unit (2RU) hyperconverged node that allows for cluster scaling with minimal footprint requirements	4
Cisco UCS 6400 Series Fabric Interconnects	Fabric interconnects	2

Table 4. Cisco HyperFlex HX240C-M6SN Node configuration

Description	Specification	Notes
CPU	2 Intel Xeon Platinum 8368 CPU @ 2.40GHz Cores per socket: 38	
Memory	4 × 128-GB DIMMs	
Cisco Flexible Flash (Cisco FlexFlash) Secure Digital (SD) card	240-GB M2	Boot drives
SSD	1-TB NVMe	Configured for housekeeping tasks
	375-GB Optane SSD	Configured as cache
	6 x 3.8-TB NVMe SSD	Capacity disks for each node
Hypervisor	VMware vSphere, 7.0.2	Virtual platform for Cisco HyperFlex HX Data Platform software
Cisco HyperFlex HX Data Platform software	Cisco HyperFlex HX Data Platform Release 5.0(1b)	
Replication factor	3	Failure redundancy from two simultaneous, uncorrelated failures

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 the following 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.
- **VMware vSphere Distributed Resource Scheduler (DRS)** – DRS monitors resource utilization and intelligently allocates system resources as needed.
- **VMware High Availability (HA)** – HA monitors hardware and OS failures and automatically restarts the affected virtual machine, providing cost-effective failover.
- **VMware vSphere Fault Tolerance (FT)** – 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 modernize their transactional and analytical workloads on premises easily and cost-effectively. 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 5. Reference architecture software components

Software	Version	Function
Cisco HyperFlex HX Data Platform	Release 5.0(1b)	Data platform
Oracle Enterprise Linux	Version 7.9	OS for Oracle RAC
Oracle UEK Kernel	4.14.35-2047.511.5.3.el7uek.x86_64	Kernel version in Oracle Linux
Oracle Grid and Oracle ASM	Version 19c Release 3	Automatic storage management

Software	Version	Function
Oracle Database	Version 19c Release 3	Oracle Database system
Oracle Silly Little Oracle Benchmark (SLOB)	Version 2.5.4	Workload suite
Swingbench, Order Entry workload	Version 2.6 (1150)	Workload suite
Oracle Recovery Manager (RMAN)	Version 19c Release 3	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 Cisco HyperFlex HX240C-M6SN 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). Using **HyperFlex iSCSI Storage** and network configuration, create initiator groups, targets, and iSCSI LUNs. The storage architecture for this environment is shown in Figure 6. This reference architecture uses 3.8-TB NVMe SSDs.

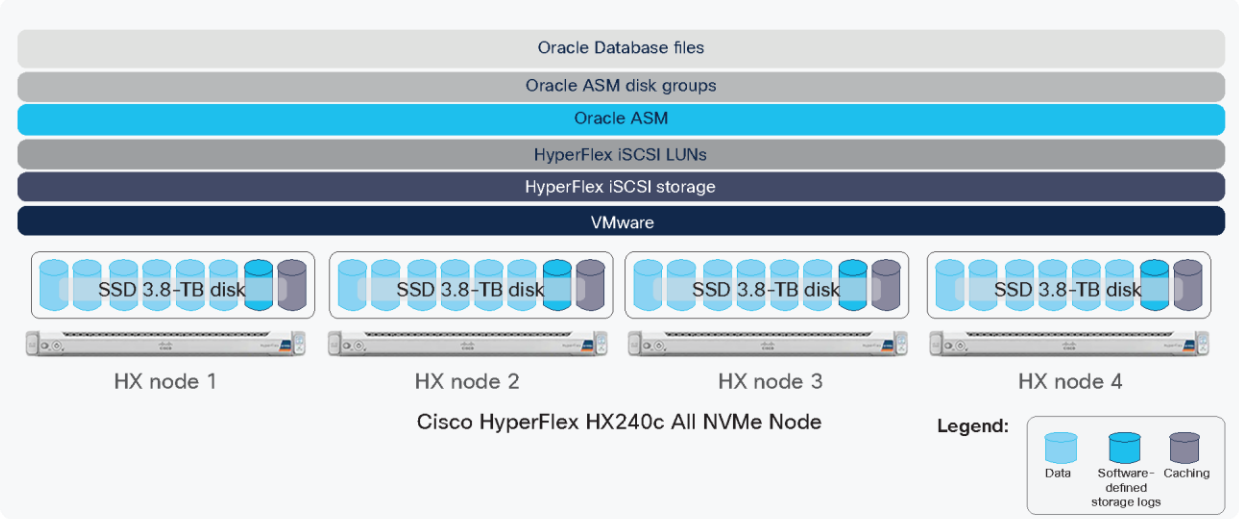


Figure 6.
Storage architecture

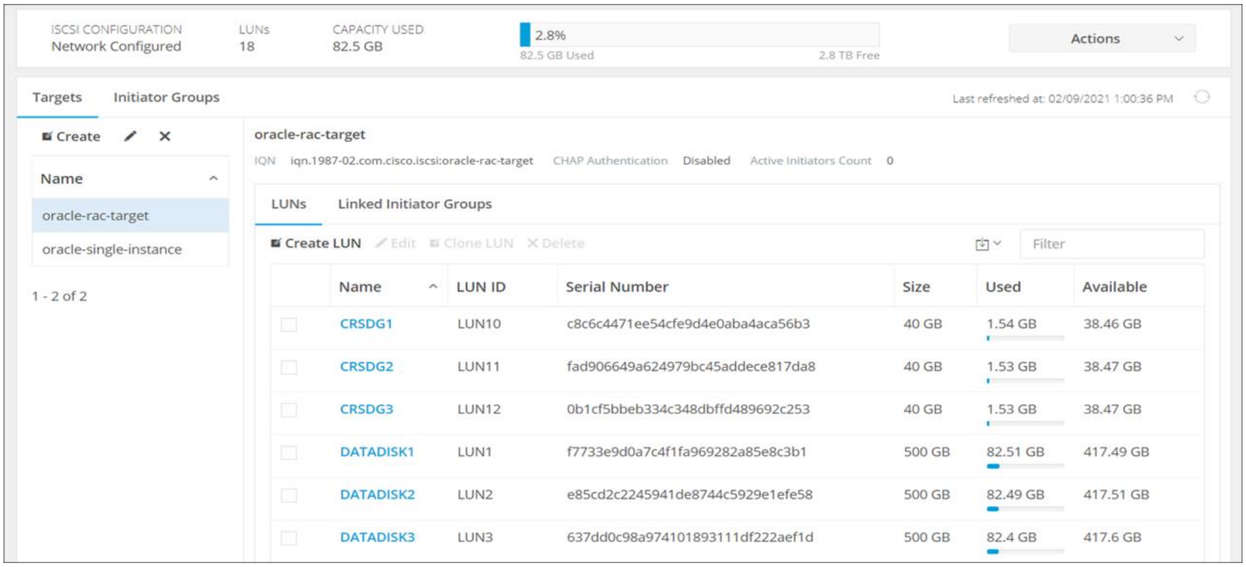
Storage configuration

This solution uses iSCSI LUNs to create shared storage that is configured as an Oracle Automatic Storage Management or ASM disk group. One of the major advantages with the iSCSI configuration is that it **supports "shared disk access"** feature by just adding the virtual machines initiator names (iqn) to the same initiator group (IG). With this **iSCSI support**, Oracle RAC virtual machines must be able to **access the iSCSI LUNs concurrently** to share the virtual machine disk configuration **without the need to enable multi-writer option** and **update the kernel** about presence and numbering of on-disk partitions.

Note: In general, both HyperFlex (HX) and Oracle ASM provide Replication Factor (RF). In our test environment, the resiliency provided by HX is applicable to all disk groups, and in addition, the DATA disk group is configured with normal redundancy provided by Oracle ASM. Table 6 shows the data disk group capacity after applying normal redundancy. The capacities vary depending on the RF being set. (For instance, if RF is set to 2, actual capacities are one-half of raw capacity. If RF is set to 3, actual capacities are one-third of raw capacity.)

Table 6. Assignment of HyperFlex iSCSI LUNs to ASM disk groups (All LUNs are shared with all four Oracle RAC virtual machines.)

Root/boot disk	DATA disk group	CRS & FRA DISK groups	REDO disk group	RMAN backup
500 GB, OS LUN	500 GB, Data1	80 GB, FRA1	70 GB, Log1	1500 GB LUN
	500 GB, Data2	80 GB, FRA2	70 GB, Log2	
	500 GB, Data3	40 GB, CRS1	70 GB, Log3	
	500 GB, Data4	40 GB, CRS2		
		40 GB, CRS3		



oracle-rac-target

IQN iqn.1987-02.com.cisco.iscsi:oracle-rac-target CHAP Authentication Disabled Active Initiators Count 0

Name	LUN ID	Serial Number	Size	Used	Available
CRSDG1	LUN10	c8c6c4471ee54cfe9d4e0aba4aca56b3	40 GB	1.54 GB	38.46 GB
CRSDG2	LUN11	fad906649a624979bc45addece817da8	40 GB	1.53 GB	38.47 GB
CRSDG3	LUN12	0b1cf5bbeb334c348dbffd489692c253	40 GB	1.53 GB	38.47 GB
DATADISK1	LUN1	f7733e9d0a7c4f1fa969282a85e8c3b1	500 GB	82.51 GB	417.49 GB
DATADISK2	LUN2	e85cd2c2245941de8744c5929e1efe58	500 GB	82.49 GB	417.51 GB
DATADISK3	LUN3	637dd0c98a974101893111df222aef1d	500 GB	82.4 GB	417.6 GB

Figure 7.
HyperFlex Connect iSCSI configuration

Table 7 summarizes the Oracle Automatic Storage Management (ASM) disk groups for this solution that are shared by all Oracle RAC virtual machines.

Table 7. Oracle ASM disk groups

Oracle ASM disk group	Purpose	Stripe size	Capacity
DATA-DG	Oracle database disk group	4 MB	1000 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 8 summarizes the configuration details.

Table 8. Oracle Database configuration

Settings	Configuration
SGA_TARGET	30 GB
PGA_AGGREGATE_TARGET	30 GB
Data files placement	ASM and DATA DG
Log files placement	ASM and REDO DG
Redo log size	32 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 8 shows the network topology for this environment.

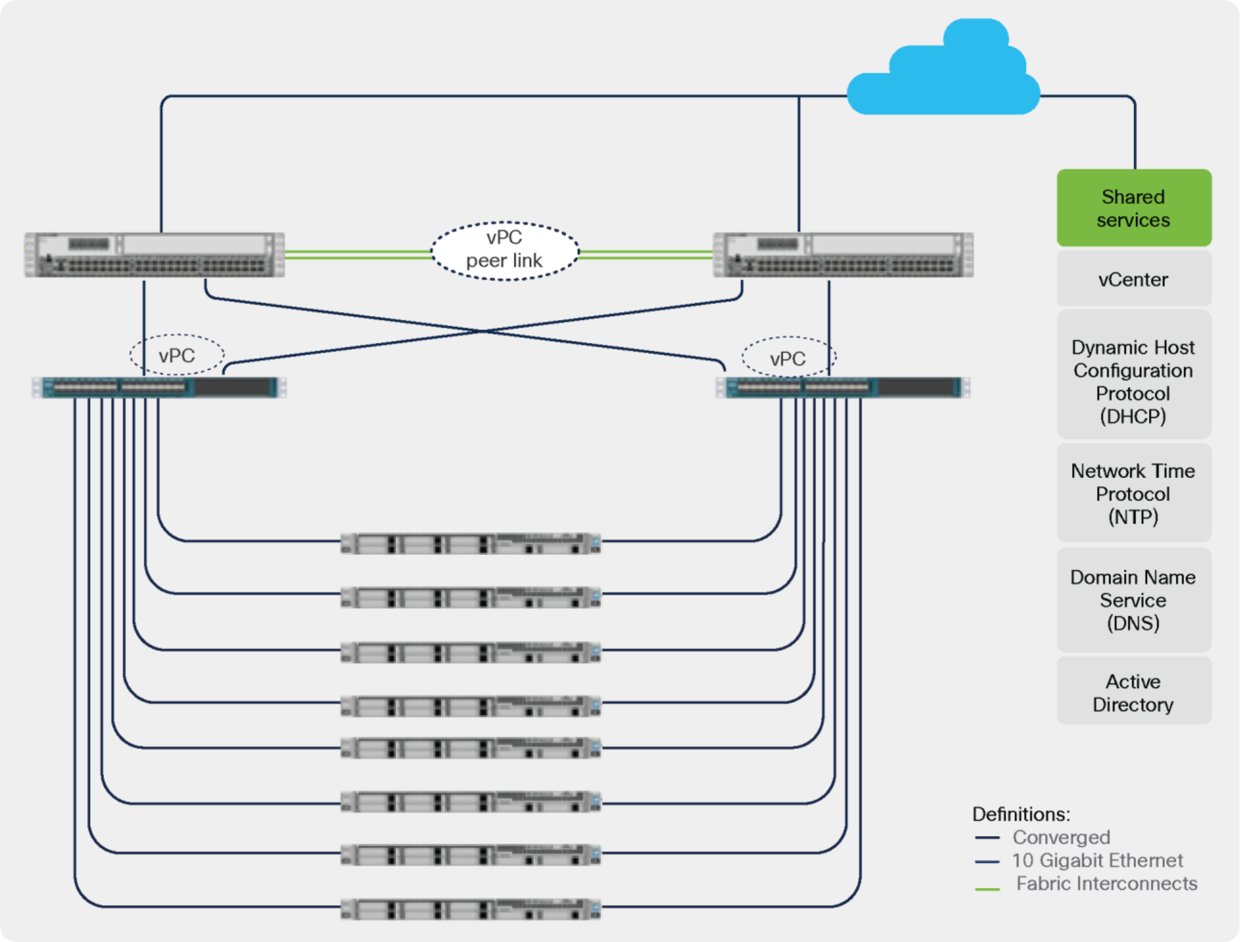


Figure 8.
Network topology

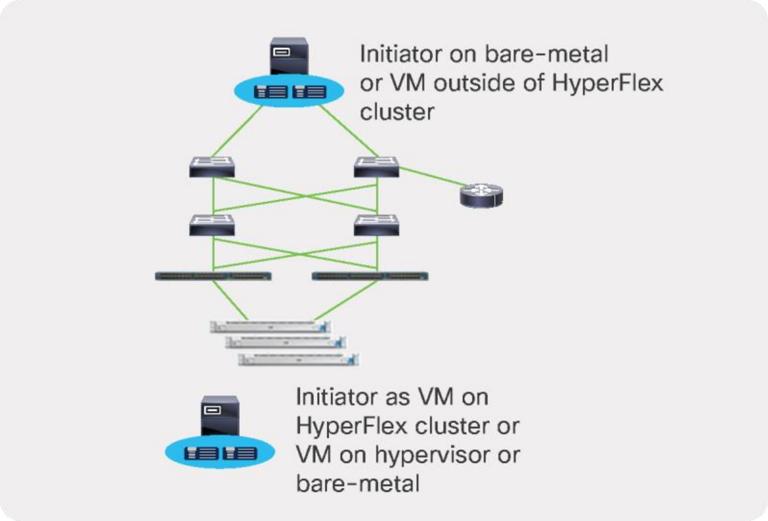

Initiator on bare-metal or VM outside of HyperFlex cluster
Initiator as VM on HyperFlex cluster or VM on hypervisor or bare-metal

Figure 9.
HyperFlex iSCSI network topologies

iSCSI access via centralized iSCSI portal

- Cisco HyperFlex HX Data Platform (HXDP) simplifies iSCSI connection availability and load balancing by providing a cluster IP for login.
  - Use the HX iSCSI CIP as the target's IP when configuring the initiator.
- Initiator can open multiple connections to the target via CIP.
- HXDP may redirect a connection to another controller VM through an iSCSI "TargetMoved" response, for load balancing or high availability.

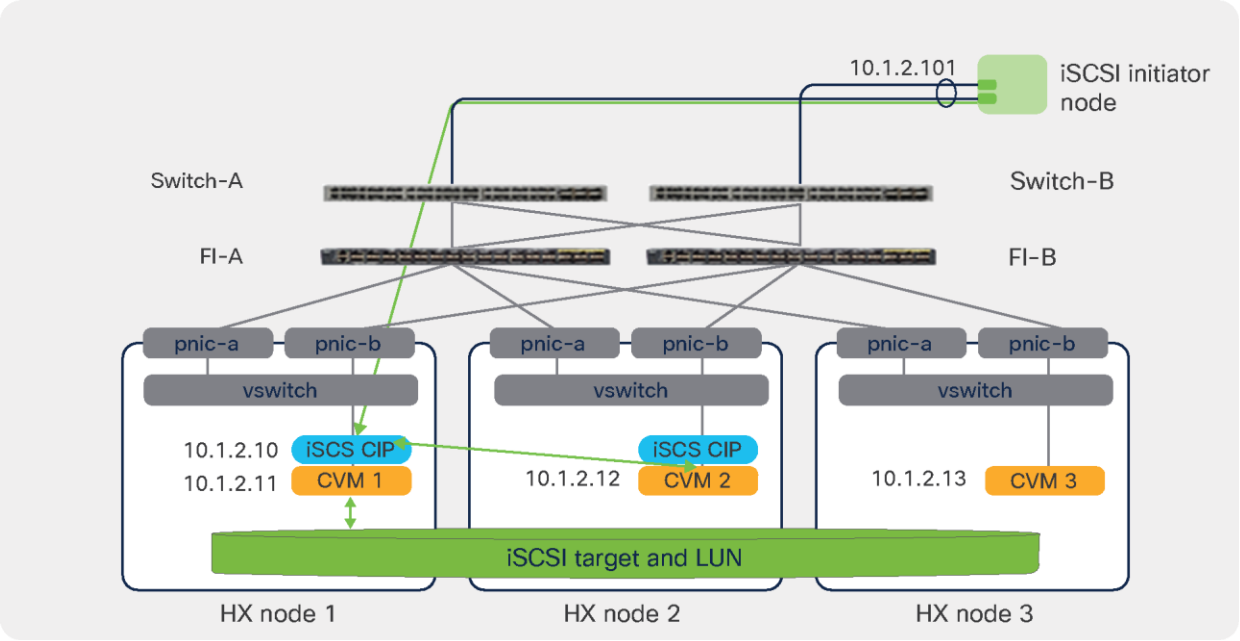


Figure 10.
Centralized iSCSI Portal reference diagram

iSCSI access via direct logins to controller VMs

- Some deployments may prefer to achieve HA (and load balancing) through MPIO software from the initiator.
- Initiator can login directly to any controller VM at its iSCSI IP.
- Initiator can transact I/O over multiple active/active paths.
- MPIO driver must handle downlink(s) because direct logins bypass the iSCSI CIP and HXDP's connection-HA and load-balancing mechanisms.

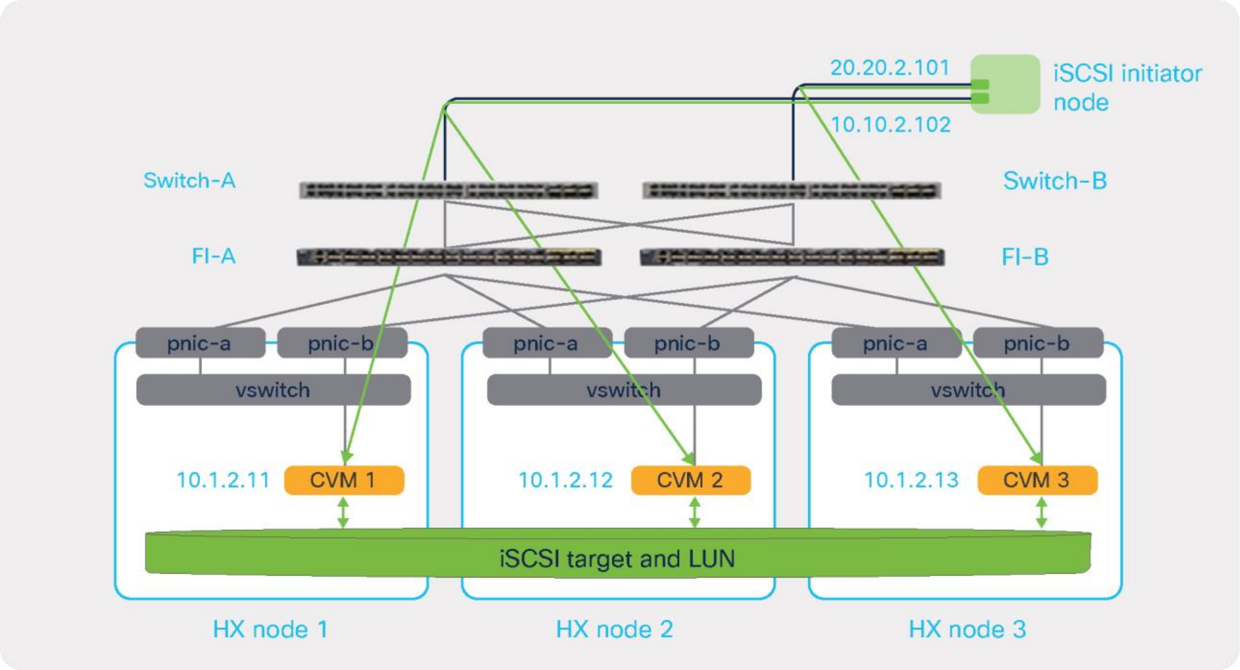


Figure 11.
Direct iSCSI logins to controller VMs reference diagram

Oracle-best-practice recommendations with iSCSI storage

- End-to-end jumbo frame solution:
  - Set MTU as 9000 (non-default MTU) while configuring iSCSI network at the Cisco HyperFlex cluster level.
  - Set MTU as 9000 while configuring iSCSI network interface in the initiator VM.
- Recommended to configure L2 (Layer 2) network on the same FIs (fabric interconnects).
- Recommended to enable HyperFlex boost mode. For more details, visit the Cisco HyperFlex Boost Mode white paper.
- Recommended to increase storage controller virtual machines' memory by +2GB for better iSCSI performance.
- Use UDEV to replace ASMLib on Red Hat 7 and OEL 7 for Oracle ASM configuration. For more details on how to configure UDEV rules, visit the Oracle ASM documentation.
- Recommended to change the device's queue depth size (default 32) in the iSCSI discovery configuration file based on the application requirement. For this reference white paper (Oracle application deployment), the device's queue depth size is set to 128.

Engineering validation

The performance, functions, and reliability of this solution were validated while running Oracle Database in a Cisco HyperFlex environment. The Oracle SLOB and Swingbench test kit was used to create and test an online transaction processing (OLTP)–equivalent database workload.

Performance testing

This section describes the results that were observed during the testing of this solution.

The test includes:

- Test profile of 70-percent read and 30-percent update operations
- Test profile of 100-percent read operations
- Test profile of 50-percent read and 50-percent update operations
- Testing of user and node scalability options using a 4-node RAC cluster with the SLOB test suite
- Testing of user and node scalability using a 4-node RAC cluster with the Swingbench test suite

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, use the Cisco HyperFlex Sizer, available at https://hyperflexsizer.cloudapps.cisco.com/.

Test methodology

The test methodology validates the computing, storage, and database performance advantages of Cisco HyperFlex systems for Oracle Database. These scenarios also provide data to help you understand the overall capabilities when scaling Oracle databases.

This test methodology uses the SLOB and Swingbench test suites to simulate an OLTP-like workload. It consists of various read/write workload configuration I/O distributions to mimic an online transactional application.

Test results

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.

SLOB performance on 4-node Cisco HyperFlex all-NVMe servers

The Silly Little Oracle Benchmark (SLOB) is a toolkit for generating and testing I/O through an Oracle database. SLOB is very effective in testing the I/O subsystem with genuine Oracle SGA-buffered physical I/O. SLOB supports testing physical random single-block reads (Oracle database file sequential read) and random single-block writes (DBWR [DataBase Writer]) flushing capability).

SLOB issues single-block reads for the read workload that are generally 8K (because the database block size was 8K). The following tests were performed and various metrics like IOPS and latency were captured along with Oracle AWR reports for each test. The database is approximately 800 GB in size.

Node-scalability performance using a 4-node Oracle RAC cluster

This test uses a 4-node Oracle RAC cluster to show the capability of the environment to scale as additional virtual machines were used to run the SLOB workload. Scaling is very close to linear. Figure 11 shows the scale testing results. **Think time is enabled and set to five** in the SLOB configuration file for this RAC node scalability test.

To validate the node scalability, we ran the SLOB test for the following scenario:

- Node-scalability testing with one, two, three, and four VMs
- 70-percent read, 30-percent update workloads
- 32 users per VM

Figure 12 shows performance of a 70-percent read and 30-percent update, respectively, at a **moderate workload**, and the corresponding latency as observed by the application; that is, measured from the virtual machine. This includes the overhead introduced by the virtual machine device layer and the I/O stack over and above the Cisco HyperFlex level storage latency.

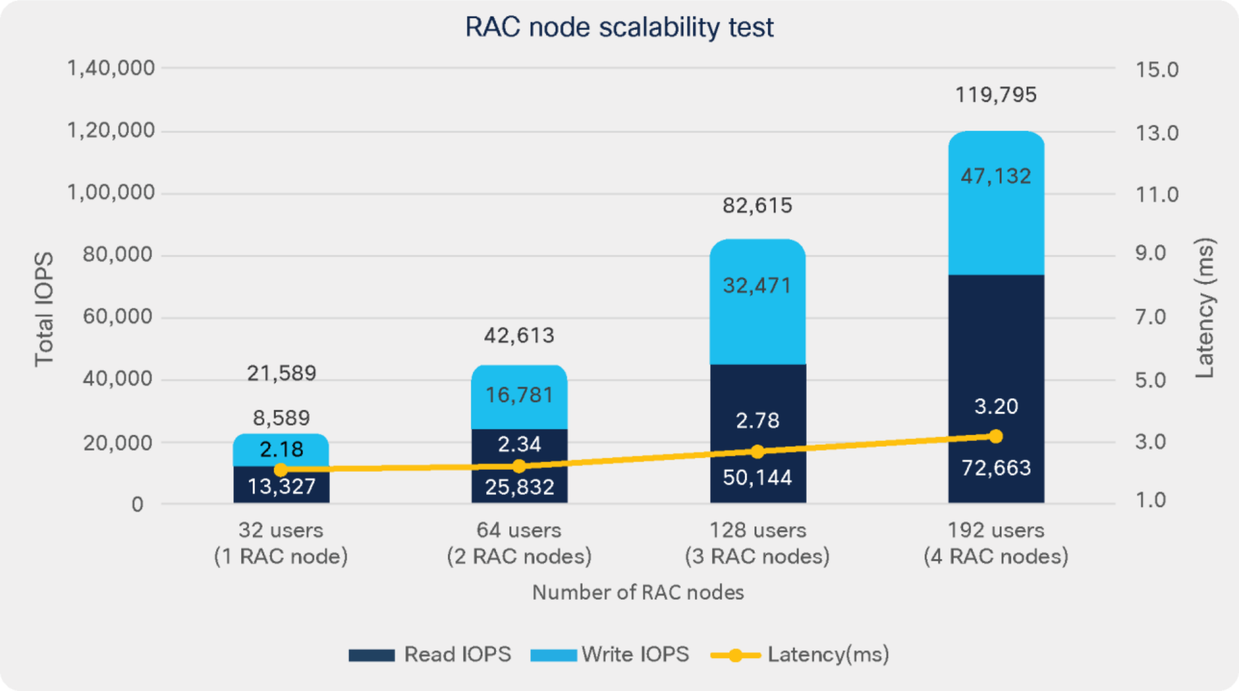


Figure 12.
Virtual machine scale test results (performance as seen by the application)

This feature gives you the flexibility to partition the workload according to the needs of the environment. The performance observed in Figure 12 includes the latency observed at the virtual machine and includes the overhead of the VM device layer, ESXi, and HyperFlex.

User scalability tests using a 4-node Oracle RAC cluster

The user scalability test uses 4-node Oracle RAC cluster instances to test the solution's capability to run many Oracle instances concurrently. The test ran for one hour and revealed that this solution stack has no bottleneck, and that the underlying storage system can sustain high performance throughout the test. Figure 13 shows the performance results for the test with different read and write combinations. The Cisco HyperFlex maximum IOPS performance for this specific all-NVMe configuration was achieved while maintaining ultra-low latency. Note that the additional IOPS come at the cost of a moderate increase in latency.

SLOB was configured to run against a 4-node Oracle RAC cluster, and the concurrent users were equally spread across all the nodes in the RAC cluster. We scale users from 32 to 192 for a 4-node Oracle RAC cluster and identify the maximum IOPS and latency:

- User scalability testing with 32, 64, 128, and 192 users using 4 Oracle RAC nodes
- Varying workloads:
    - 100-percent read
    - 70-percent read, 30-percent update
    - 50-percent read, 50-percent update

The following graphs illustrate user scalability in terms of the total IOPS (both read and write) when run at 32, 64, 128, and 192 concurrent users with 100-percent read, 70-percent read, and 50-percent read. SLOB was configured to run against a 4-node Oracle RAC cluster and the concurrent users were equally spread across all the virtual machines.

Figures 13 and 14 and Tables 9, 10, and 11 show performance of a 4-node Oracle RAC cluster for 100-percent read workload, 70-percent read, and 30-percent update, and 50-percent read and 50-percent update at a **moderate workload**, and the corresponding latency as observed by the application; that is, measured from the virtual machine. This includes the overhead introduced by the virtual machine device layer and the I/O stack over and above the Cisco HyperFlex–level storage latency.

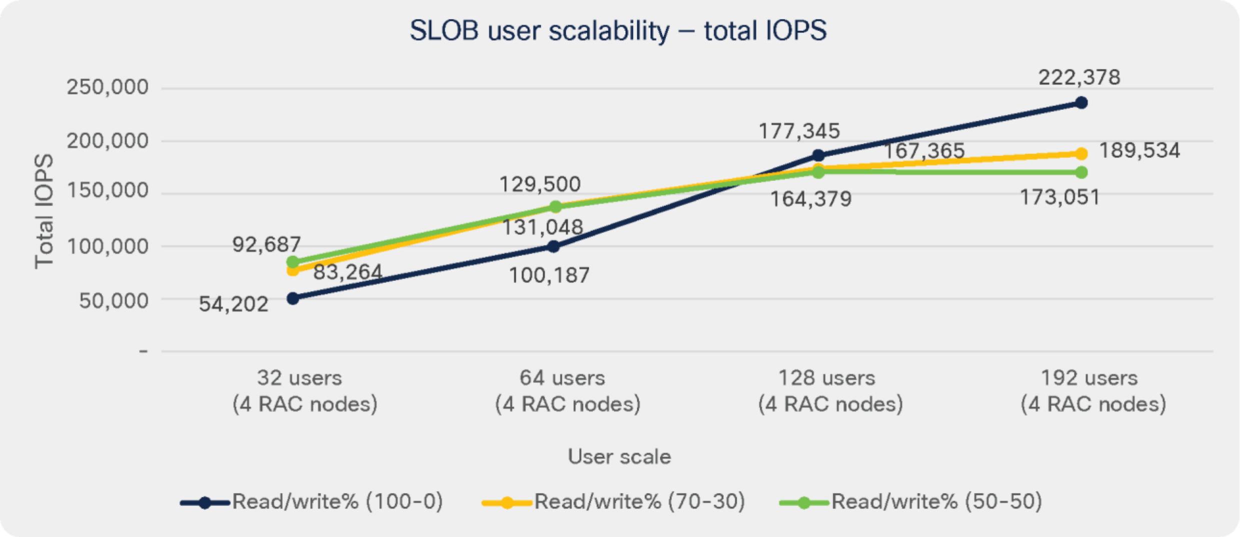


Figure 13.
User scalability test – IOPS graph (performance as seen by the application)

As expected, the graph illustrates the linear scalability with increased users and similar IOPS (around 189K) until it reaches 192 users across all workloads. Beyond 167K IOPS, the additional users yield higher IOPS but not at the same IOPS/user rate when the write percentages are higher.

The 50-percent update (50-percent read) resulted in 173K IOPS at 192 users, whereas 70-percent read (30-percent update) resulted in 189K IOPS at 192 users. The 100-percent read resulted in 222K IOPS at 192 users, which is excellent, considering the tests were performed at moderate load.

Figure 14 illustrates the latency exhibited across different workloads. The 100-percent read workload experienced less than 1.0ms, and it varies based on the workload. The 50-percent read (50-percent update) exhibited higher latencies with increased user counts.

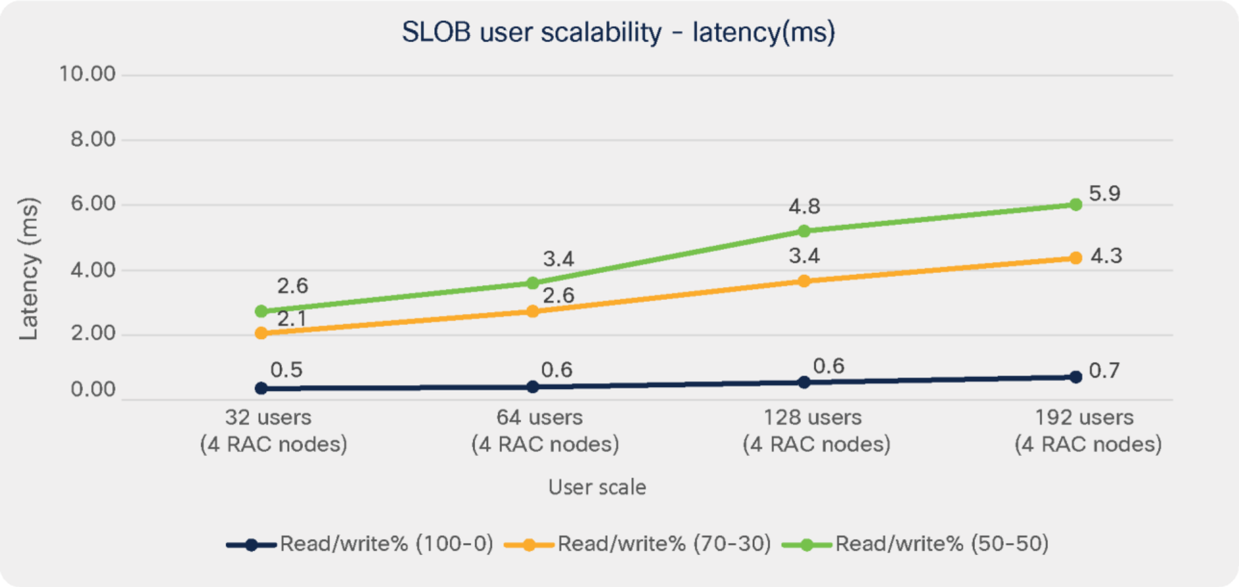


Figure 14.
User scalability test - latency graph (performance as seen by the application)

Table 9. User scalability application-level performance for 100-percent read

User count	Average IOPS	Average latency (ms)
32	54,202	0.5
64	100,187	0.6
128	177,345	0.6
192	222,378	0.7

Table 10. User scalability application-level performance for 70-percent read, 30-percent update

User count	Average IOPS	Average latency (ms)
32	83,264	2.1
64	131,048	2.6
128	167,365	3.4
192	189,543	4.3

Table 11. User scalability application-level performance for 50-percent read, 50-percent update

User count	Average IOPS	Average latency (ms)
32	92,687	2.6
64	129,500	3.4
128	164,379	4.8
192	173,051	5.9

Swingbench performance on 4-node Cisco HyperFlex All-NVMe servers

Swingbench is a simple-to-use, free, Java-based tool to generate database workloads and perform stress testing using different benchmarks in Oracle database environments. Swingbench can be used to demonstrate and test technologies such as Real Application Clusters, online table rebuilds, standby databases, online backup and recovery, etc.

Swingbench provides four separate benchmarks, namely, Order Entry, Sales History, Calling Circle, and Stress Test. For the tests described in this solution, the Swingbench Order Entry benchmark was used for OLTP workload testing and the Sales History benchmark was used for the DSS workload testing.

The Order Entry benchmark is based on a Swingbench Order Entry schema and is TPC-C-like by types of transactions. The workload uses a very balanced read/write ratio around 60/40 and can be designed to run continuously and test the performance of a typical Order Entry workload against a small set of tables, producing contention for database resources.

For this solution, we created an OLTP (Order Entry) database to demonstrate database consolidation, performance, and sustainability. The OLTP database is approximately 512 GB in size.

We tested a combination of scalability and stress-related scenarios that are typically encountered in real-world deployments and were run on a current 4-node Oracle RAC cluster.

Scalability performance

The first step after the database creation is calibration, regarding the number of concurrent users, nodes, OS, and database optimization.

For an OLTP database workload featuring an Order Entry schema, we used an OLTP database. For the OLTP database (OLTPDB) (512 GB), we used a 30-GB size of System Global Area (SGA). We also ensured that HugePages were in use. The OLTP database scalability test was run for at least 4 hours and ensured that results were consistent for the duration of the full run.

User scalability performance using a 4-node Oracle RAC cluster

The user scalability test uses 4-node Oracle RAC cluster instances to test the solution's capability to run many Oracle RAC instances concurrently. The test ran for four hours and revealed that this solution stack has no bottleneck, and that the underlying storage system can sustain high performance throughout the test. The Cisco HyperFlex maximum IOPS performance for this specific all-NVMe configuration was achieved while maintaining low latency. Note that the additional IOPS come at the cost of a moderate increase in latency.

Swingbench was configured to run against a 4-node Oracle RAC cluster, and the concurrent users were equally spread across all the virtual machines in the RAC cluster.

- 32, 64, 128, 256 and 480 are the number of concurrent users.

The graph in Figure 15 illustrates the Transactions Per Minute (TPM) for the OLTP database user scale on the 4-node Oracle RAC cluster. TPM for 32, 64, 128, 256, and 480 users are around 262K, 446K, 627K, 823K, and 937K.

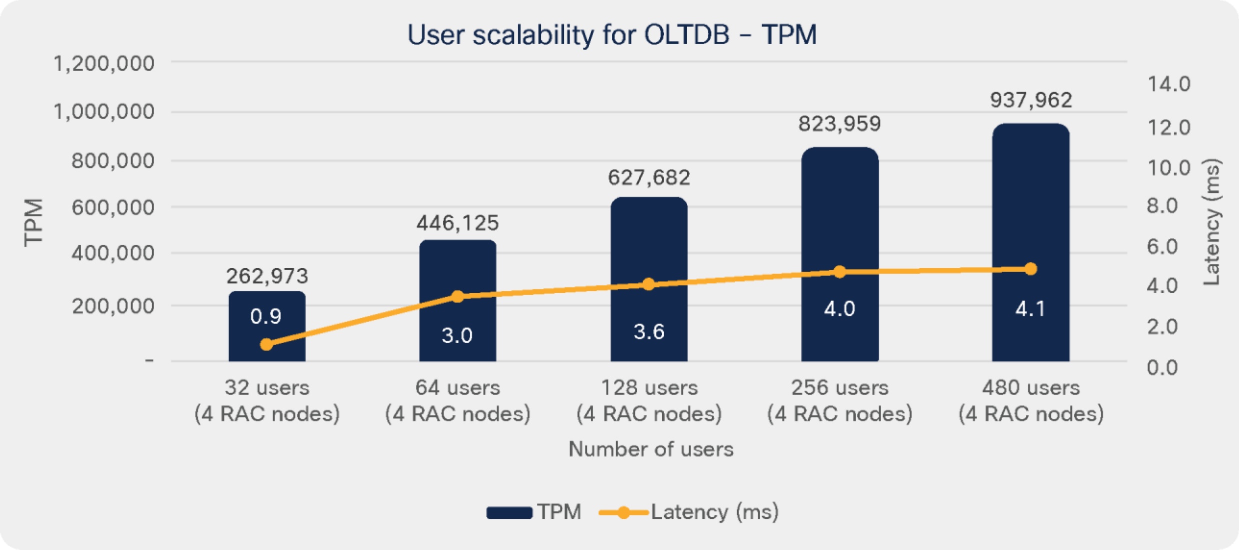


Figure 15.
User scalability TPM for OLTP database (OLTDB) (performance as seen by the application)

The graph in Figure 16 illustrates the total IOPS for the OLTP database user scale on a 4-node Oracle RAC cluster. Total IOPS for 32, 64, 128, 256, and 480 users are around 35K, 55K, 73K, 85K, and 97K.

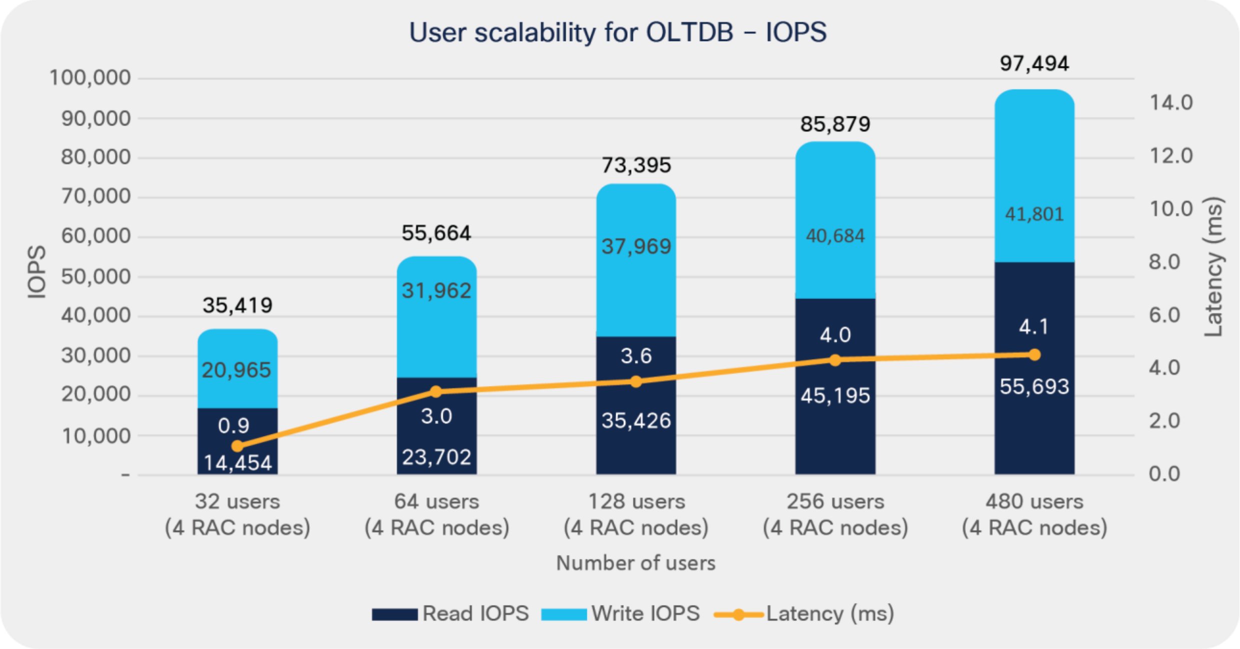


Figure 16.
User scalability IOPS for OLTP database (OLTPDB) (performance as seen by the application)

Table 12. User scalability application-level performance – IOPS and latency

User count	Average IOPS	Average latency (ms)
32 (4 RAC nodes)	35,419	0.9
64 (4 RAC nodes)	55,664	3.0
128 (4 RAC nodes)	73,395	3.6
256 (4 RAC nodes)	85,879	4.0
480 (4 RAC nodes)	97,494	4.1

Node scalability performance using a 4-node Oracle RAC cluster

The graph in Figure 17 illustrates the TPM for the OLTP database node scale. TPM for 1, 2, 3, and 4 RAC nodes are around 401K, 555K, 710K, and 802K, with latency under 4.3 milliseconds all the time.

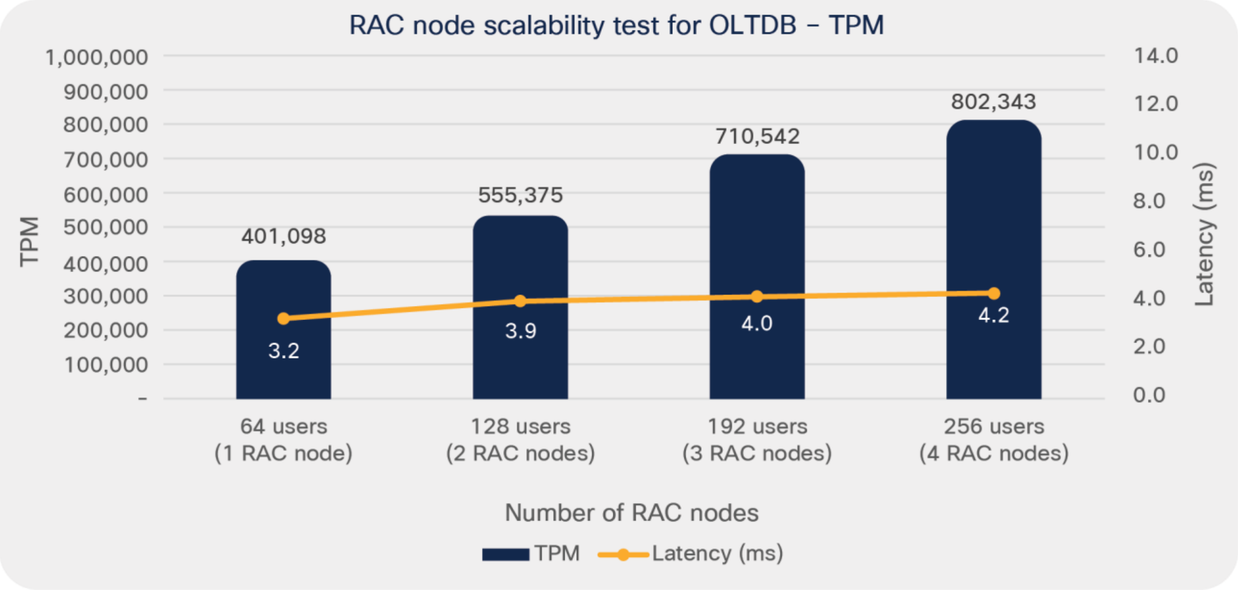


Figure 17.
Node scalability TPM for OLTP database (OLTBDB) (performance as seen by the application)

The graph in Figure 18 illustrates the total IOPS for the OLTPDB database node scale. Total IOPS for 1, 2, 3, and 4 RAC nodes are around 45K, 64K, 81K, and 91K, with latency under 4.3 milliseconds all the time.

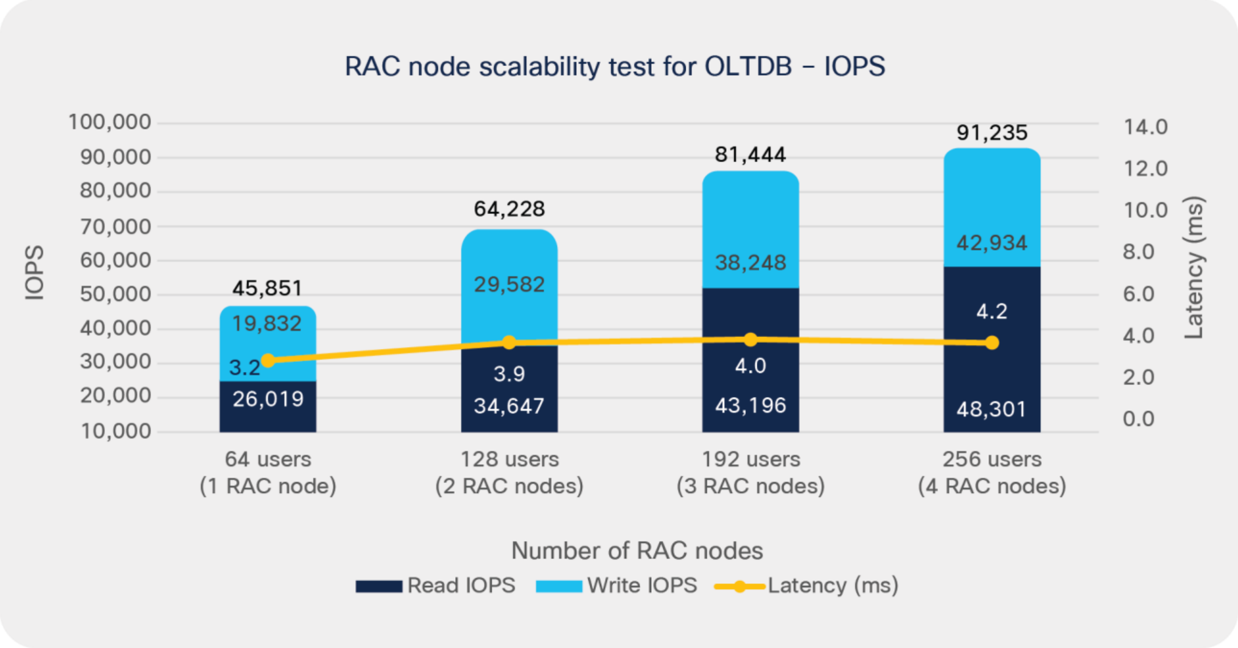


Figure 18.
Node scalability IOPS for OLTP database (OLTBDB) (performance as seen by the application)

Table 13. Node scalability application-level performance – IOPS and latency

User count	Average IOPS	Average latency (ms)
64 (1 RAC node)	45,851	3.2
128 (2 RAC nodes)	64,228	3.9
192 (3 RAC nodes)	81,444	4.0
256 (4 RAC nodes)	91.235	4.2

Oracle RAC database backup, restore, 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.

The list below is a **summary** of actions performed using Oracle RMAN; the actions are discussed in detail in the following sections:

- Create a full database backup to the shared iSCSI drive. In this reference paper, local Cisco HyperFlex iSCSI storage has been used to maintain the backup copies, but, In general, external iSCSI storage is also supported.
- Restore and recover the database from the backup copies in case of any media failure.
- Duplicate/clone a database from a backup copy or from an active database. In this reference paper, duplication of a production database to a remote host from the full backup copy has been tested, and the results are shown in their respective sections.

The list of Oracle RMAN operations mentioned above are all validated and tested using iSCSI Storage in the latest Cisco HyperFlexHX Data Platform, Release 5.0(1b).

Database backup

In the lab environment, a 2-node Oracle RAC cluster with a database of approximately 200 GB in size has been used for testing, and RMAN was configured on 1 RAC node. A dedicated 1500-GB iSCSI LUN 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. Using RMAN, initiate the full backup creation and provide the mount point location to store the backup sets that are generated.

With **iSCSI Support in Cisco HyperFlex HX Data Platform Release 5.0(1b):** Create iSCSI LUN of size 1500 GB on the remote cluster and add the virtual machine (where the backup LUN has been created) initiator (iqn) name to the same initiator group (where the RAC virtual machines iqn names have already been added) on the source cluster to make the backup LUN accessible and to be able to discover and log in to the targets present on the remote side. Once login to the target is successful, RAC virtual machines residing on the source cluster would be able to detect the backup LUN and mount it to the file system. For backup testing, mount the 1500-GB LUN on RAC VM1 (where RMAN is installed and configured) under the "/backup" directory and configure the backup location to the mount point (the "/backup" directory) while creating a full backup using RMAN. Once the backup creation is complete, mount the shared iSCSI LUN on the destination VM too, under the same "/backup" directory as in the source VM. Afterward, connect RMAN to the source database as TARGET and the remote database as AUXILLARY, use the CATALOG command to update the RMAN repository with the location of the mount point ("/backup" directory) where the backup sets have been created. With this iSCSI shared-drive feature, manual transfer/copy of the backup sets from the source cluster to the destination cluster using operating system utilities such as FTP or SCP can be eliminated. In a scenario where there are multiple backup copies, or the backup copy sizes are too big, and to manually transfer them to the remote cluster would be time consuming; if you use the iSCSI Support feature instead, disk backups can be made accessible across both source and remote clusters using a shared disk.

Note: One of the best-practice recommendations with HyperFlex iSCSI Storage is to mount the shared drive on one virtual machine at a time, because HyperFlex iSCSI Storage does not support a clustered file system with "Linux ext4," which was used to partition and format the iSCSI LUNs. For example, during the Oracle RMAN backup testing (results shown below), the iSCSI shared drive has to be mounted first on the local cluster for the Oracle RMAN to create backup copies and once complete mount the shared drive on the remote cluster to view the backup copies created.

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 RAC node 1 increased because of the RMAN processes.

Figure 19 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 residing on the remote cluster. A virtual machine with a mount point /backup has been created on a remote host and the same iSCSI drive has also been shared across the local cluster, to create and store the backup sets in the shared drive. However, if the intention is to duplicate the database in the remote cluster, and not just to store the backup copies, it needs the exact replica of the test setup that makes up the local cluster. The procedure is explained in detail later in this document, in the section titled "Duplicate database using RMAN."

Note: The recommended approach while creating database backups is always to store the backup set copies at a remote location but not at the same location where the source database exists. This is because if there is any kind of unexpected/sudden failure on the local cluster, the backup copies will still be present at the remote location to restore and recover the database. And to perform this, the iSCSI shared drive feature would be beneficial, because it makes the disk backups accessible across both clusters and eliminates the manual transfer procedure. In the figure below, if the local and remote clusters are connected to different FIs, make sure to set the MTU size to 9000 on the top-of-rack switches to enable the Jumbo frame packet transmission across clusters that span across different distribution switches.

In Figure 19, the local cluster uses HXAF240C-M6SN servers, whereas the remote cluster (where the backup is stored) uses HXAF220C-M5SN servers.

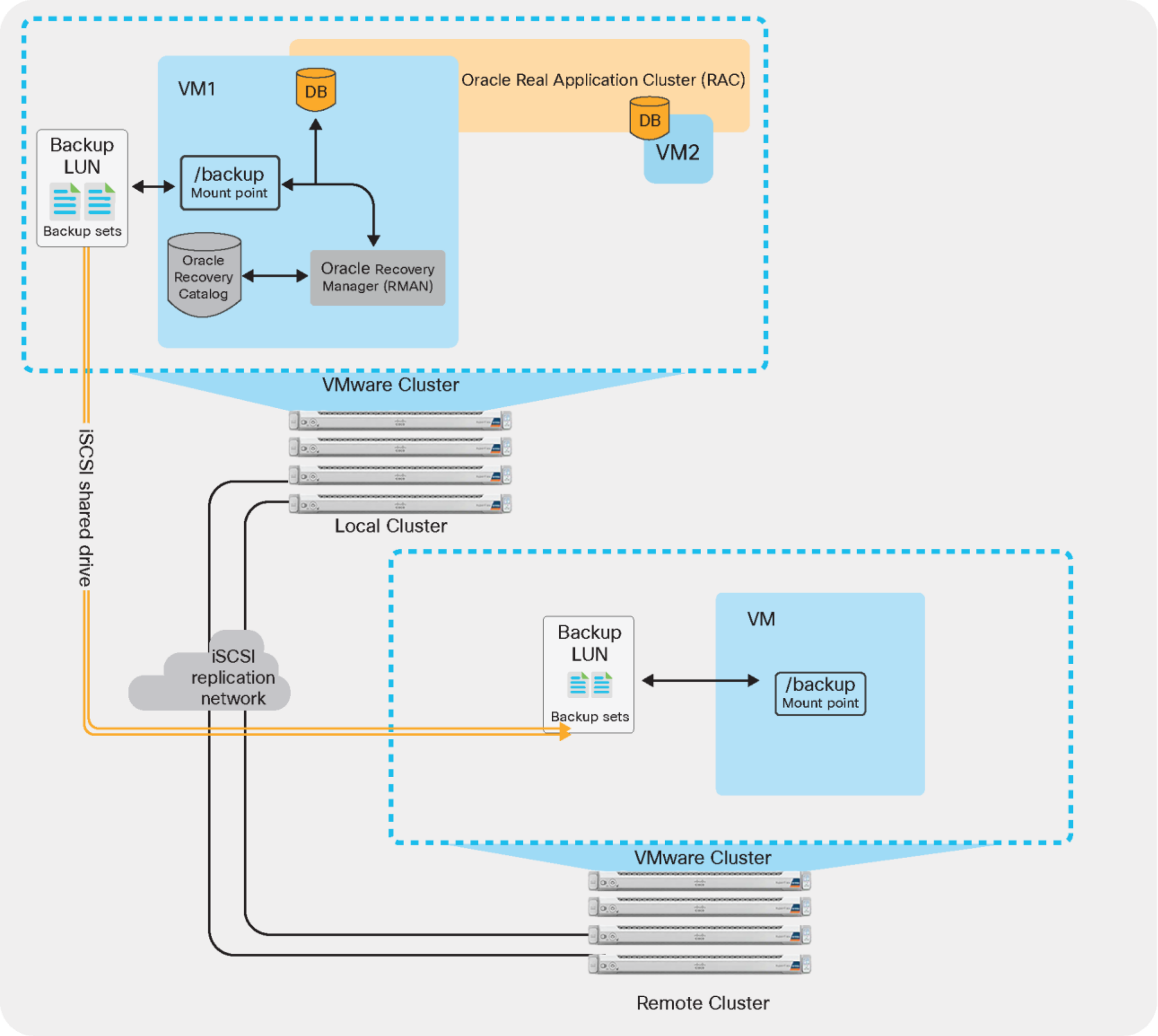


Figure 19.
RMAN test environment

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 14 shows the test results.

Table 14. RMAN backup results

	During idle database	During active Swingbench testing
Backup type	Full backup	Full backup
Backup size	250 GB	250 GB
Backup elapsed time	00:11:55	00:18:35
Throughput	591 MBps	1028 MBps

Table 15 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 15. Impact of RMAN backup operation

	Average TPM	Average TPS
RMAN hot backup during Swingbench test	674,975	11,249
Swingbench only	725,478	12,091

Note: These tests and several others were performed to validate the functions. The results presented here are not intended to provide detailed sizing guidance or guidance for performance headroom calculations, which are beyond the scope of this document.

Restore and recover an Oracle database

To restore and recover an Oracle database from backup, use the Oracle RMAN RESTORE command to restore, validate, or preview RMAN backups. Typically, the restore function is performed when a media failure has damaged a current data file, control file, or archived redo log, or before performing a point-in-time recovery.

To restore data files to their current location, the database must be started, mounted, or open with the table spaces or data files to be restored offline. However, if you use RMAN in an Oracle Data Guard environment, then connect RMAN to a recovery catalog. The RESTORE command restores full backups, incremental backups, or image copies, and the files can be restored to their default location or to a different location.

The purposes for using the RECOVER command are:

- To perform complete recovery of the whole database or one or more restored data files
- To apply incremental backups to a data file image copy (not a restored data file) to roll it forward in time
- To recover a corrupt data block or set of data blocks within a data file

On a high level, the following steps are performed in restoring and recovering the database from RMAN backup:

- Verify the current RMAN configuration on the server where the restore operation is performed and make sure you provide the right backup location
- Restore control file from backup
- Restore the database
- Recover the database

Duplicate a database using Oracle RMAN

Database duplication is the use of the Oracle RMAN DUPLICATE command to copy all, or a subset, of the data in a source database from a backup. The duplicate command functions entirely independently from the primary database. The tasks that can be performed and useful in a duplicate database environment, most of which involve testing, are:

- Test backup and recovery procedures
- Test an upgrade to a new release of oracle database
- Test the effect of applications on database performance
- Create a standby database
- Generate reports

For example, you can duplicate the production database to a remote host and then use a duplicate environment to perform the tasks just listed while the production database on the local host operates as usual.

As part of the duplicating operation, RMAN manages the following:

- Restores the target data files to the duplicate database and performs incomplete recovery by using all available backups and archived logs
- Shuts down and starts the auxiliary database
- Opens the duplicate database with the RESETLOGS option after incomplete recovery to create the online redo logs
- Generates a new, unique DBID for the duplicate database

In the test environment, an Oracle 2-node RAC cluster with database that is approximately 200 GB in size has been used. The virtual machine hosted on the remote server has a created backup disk (iSCSI LUN) that was shared across the local cluster to create the backup set in the shared drive. To prepare for database duplication, the first step would be to create an auxiliary instance, which should be a replica of the production environment. For the duplication to work, connect RMAN to both the target (primary) database and an auxiliary instance started in NOMOUNT mode. If RMAN is able to connect to both instances, the RMAN client can run on any server. However, all backups, copies of data files, and archived logs for duplicating the database must be accessible by the server on the remote host.

Figure 20 shows the RMAN environment for the database duplication testing. The figure shows a typical setup. In this case, a remote server with the same file structure has been created, and the backup target is the mount point /backup residing on the remote cluster. The auxiliary instance (Virtual machine1) of the Oracle RAC cluster on the remote cluster consists of the backup target and RMAN to initiate the duplicate process. In the below figure, the “/backup” directory created in the remote cluster is using the iSCSI LUN of size 1500 GB and is shared across the local cluster to create backup sets using RMAN on VM1. Therefore, the backup sets available on the iSCSI shared drive are accessible from both the clusters.

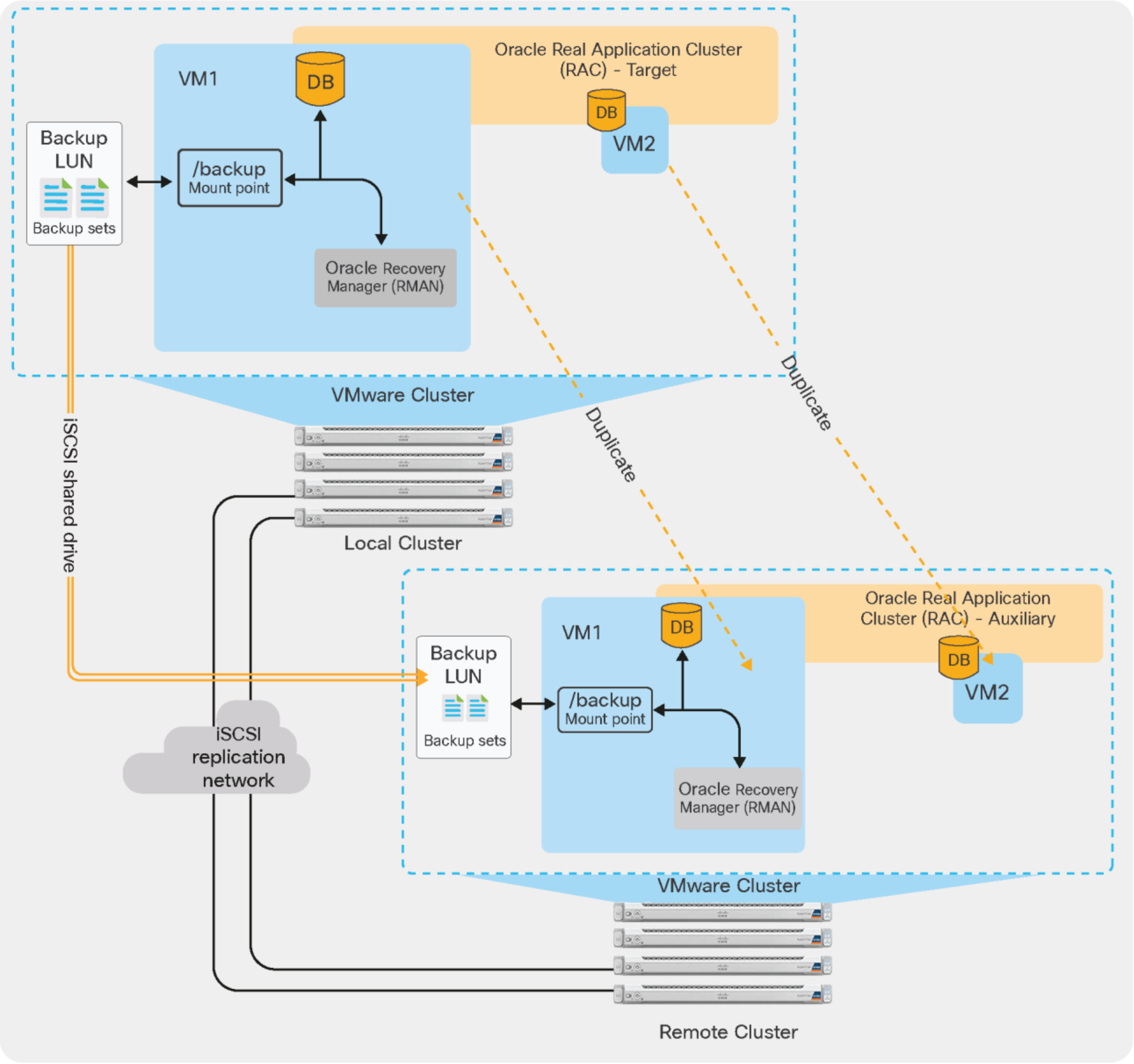
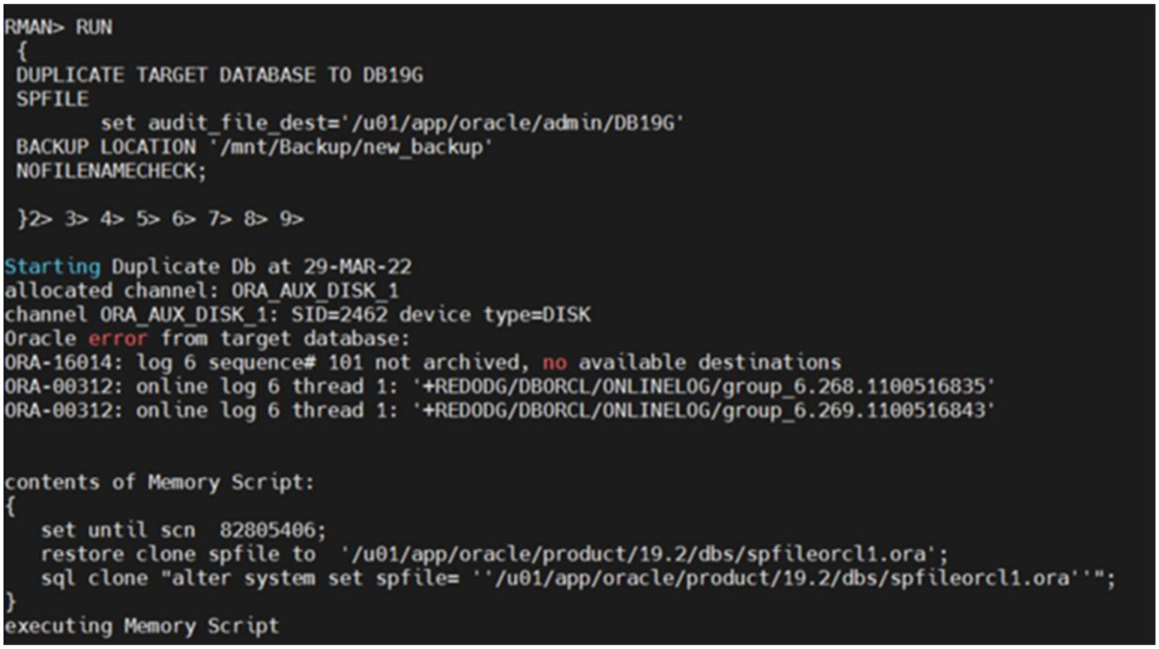


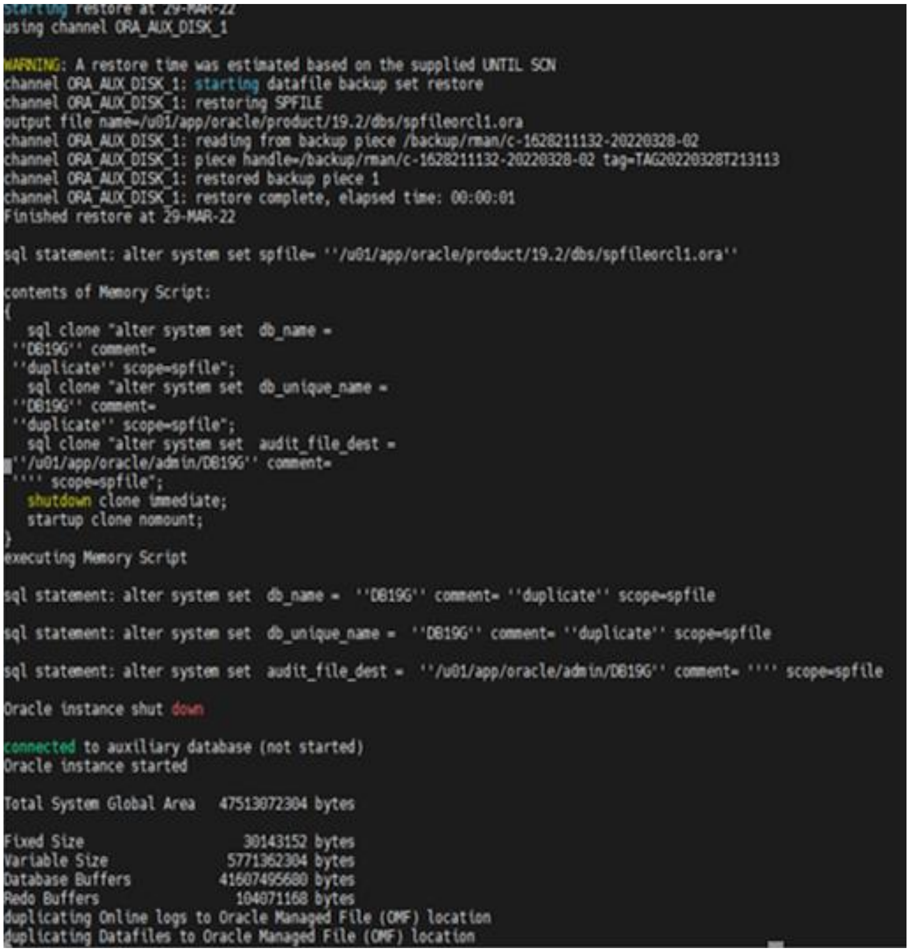
Figure 20.
Test environment for database duplication

While running RMAN duplicate command, make sure to configure the backup location to the mount point created on the remote cluster (/backup directory) where the backup copies are residing.



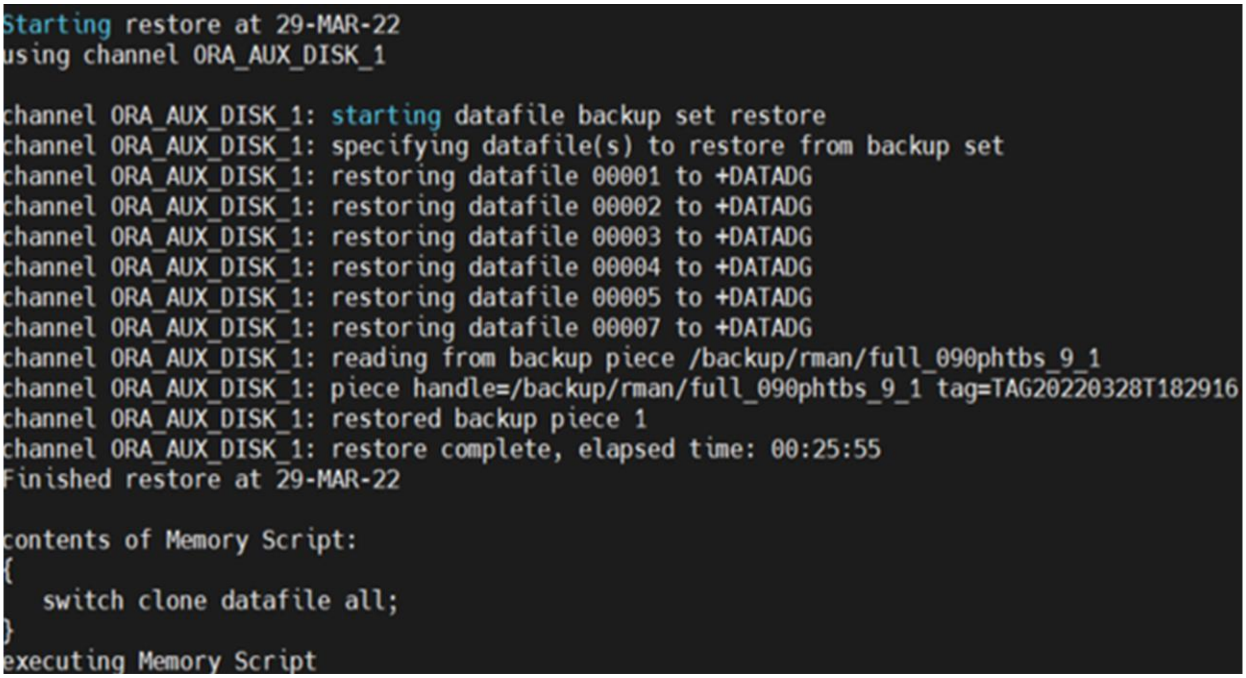
```
RMAN> RUN
 {
 DUPLICATE TARGET DATABASE TO DB19G
 SPFILE
        set audit_file_dest='/u01/app/oracle/admin/DB19G'
 BACKUP LOCATION '/mnt/Backup/new_backup'
 NOFILENAMECHECK;

 }2> 3> 4> 5> 6> 7> 8> 9>

Starting Duplicate Db at 29-MAR-22
allocated channel: ORA_AUX_DISK_1
channel ORA_AUX_DISK_1: SID=2462 device type=DISK
Oracle error from target database:
ORA-16014: log 6 sequence# 101 not archived, no available destinations
ORA-00312: online log 6 thread 1: '+REDODG/DBORCL/ONLINELOG/group_6.268.1100516835'
ORA-00312: online log 6 thread 1: '+REDODG/DBORCL/ONLINELOG/group_6.269.1100516843'


contents of Memory Script:
{
   set until scn  82805406;
   restore clone spfile to  '/u01/app/oracle/product/19.2/dbs/spfileorcl1.ora';
   sql clone "alter system set spfile= ''/u01/app/oracle/product/19.2/dbs/spfileorcl1.ora''";
}
executing Memory Script
```

Restore of the control file, duplicating online logs and Datafiles to the OMF location, has started on the remote cluster:



```
Starting restore at 29-MAR-22
using channel ORA_AUX_DISK_1

WARNING: A restore time was estimated based on the supplied UNTIL SCN
channel ORA_AUX_DISK_1: starting datafile backup set restore
channel ORA_AUX_DISK_1: restoring SPFILE
output file name=/u01/app/oracle/product/19.2/dbs/spfileorcl1.ora
channel ORA_AUX_DISK_1: reading from backup piece /backup/rman/c-1628211132-20220328-02
channel ORA_AUX_DISK_1: piece handle=/backup/rman/c-1628211132-20220328-02 tag=TAG20220328T213113
channel ORA_AUX_DISK_1: restored backup piece 1
channel ORA_AUX_DISK_1: restore complete, elapsed time: 00:00:01
Finished restore at 29-MAR-22

sql statement: alter system set spfile= ''/u01/app/oracle/product/19.2/dbs/spfileorcl1.ora''

contents of Memory Script:
{
   sql clone "alter system set  db_name =
 ''DB19G'' comment=
 ''duplicate'' scope=spfile";
   sql clone "alter system set  db_unique_name =
 ''DB19G'' comment=
 ''duplicate'' scope=spfile";
   sql clone "alter system set  audit_file_dest =
 ''/u01/app/oracle/admin/DB19G'' comment=
 '''' scope=spfile";
   shutdown clone immediate;
   startup clone nomount;
}
executing Memory Script

sql statement: alter system set  db_name =  ''DB19G'' comment= ''duplicate'' scope=spfile

sql statement: alter system set  db_unique_name =  ''DB19G'' comment= ''duplicate'' scope=spfile

sql statement: alter system set  audit_file_dest =  ''/u01/app/oracle/admin/DB19G'' comment= '''' scope=spfile

Oracle instance shut down

connected to auxiliary database (not started)
Oracle instance started

Total System Global Area   47513072304 bytes

Fixed Size                    30143152 bytes
Variable Size               5771362304 bytes
Database Buffers           41607495680 bytes
Redo Buffers                 104071168 bytes
duplicating Online logs to Oracle Managed File (OMF) location
duplicating Datafiles to Oracle Managed File (OMF) location
```

Restore data files from the backup copy in the mount location:



```
Starting restore at 29-MAR-22
using channel ORA_AUX_DISK_1

channel ORA_AUX_DISK_1: starting datafile backup set restore
channel ORA_AUX_DISK_1: specifying datafile(s) to restore from backup set
channel ORA_AUX_DISK_1: restoring datafile 00001 to +DATADG
channel ORA_AUX_DISK_1: restoring datafile 00002 to +DATADG
channel ORA_AUX_DISK_1: restoring datafile 00003 to +DATADG
channel ORA_AUX_DISK_1: restoring datafile 00004 to +DATADG
channel ORA_AUX_DISK_1: restoring datafile 00005 to +DATADG
channel ORA_AUX_DISK_1: restoring datafile 00007 to +DATADG
channel ORA_AUX_DISK_1: reading from backup piece /backup/rman/full_090phtbs_9_1
channel ORA_AUX_DISK_1: piece handle=/backup/rman/full_090phtbs_9_1 tag=TAG20220328T182916
channel ORA_AUX_DISK_1: restored backup piece 1
channel ORA_AUX_DISK_1: restore complete, elapsed time: 00:25:55
Finished restore at 29-MAR-22

contents of Memory Script:
{
   switch clone datafile all;
}
executing Memory Script
```

Table 16. RMAN duplicate database results

	Duplicate database
Backup type	Full backup
Backup size	250 GB
Duplication elapsed time	00:25:55

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 Database.

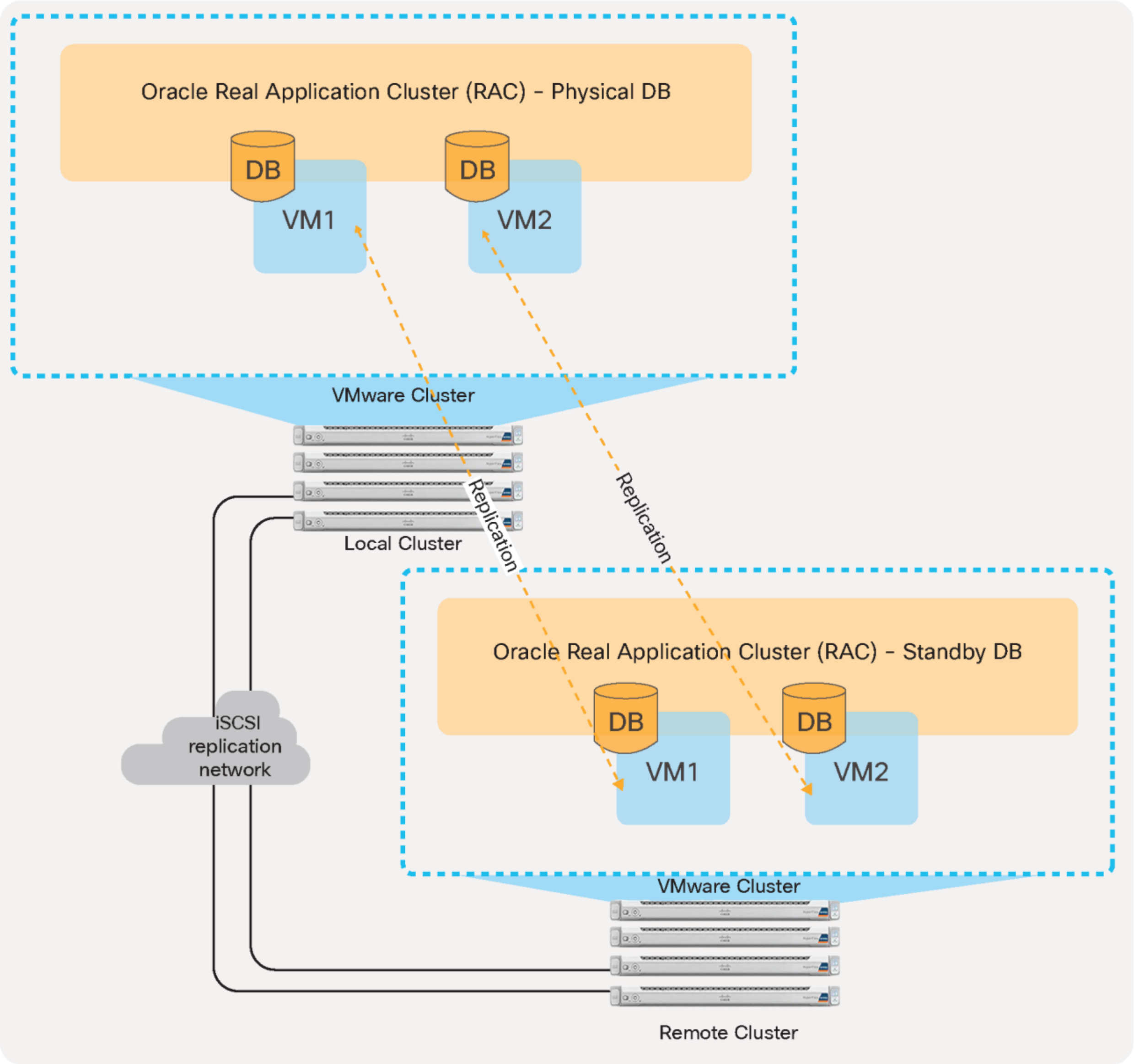

Oracle Real Application Cluster (RAC) - Physical DB
DB
VM1
DB
VM2
VMware Cluster
Local Cluster
Replication
Replication
iSCSI replication network
Oracle Real Application Cluster (RAC) - Standby DB
DB
VM1
DB
VM2
VMware Cluster
Remote Cluster

Figure 21.
Test environment of replicating database (DB)

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 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 Golden Gate for efficient and reliable transmission of redo logs from a source database to one or more remote destinations.

A Data Guard configuration consists of one primary database and up to nine standby databases. The databases in a Data Guard configuration are connected by Oracle Net and may be dispersed geographically. There are no restrictions on where the databases can be located so long as they can communicate with each other. For example, you can have a standby database on the same system as the primary database, along with two standby databases on another system.

The Oracle Data Guard Broker is the utility used to manage Data Guard operations. The Data Guard Broker logically groups the primary and standby databases into a broker configuration that allows the broker to manage and monitor them together as an integrated unit. Broker configurations are used from the Data Guard command-line interface. The Oracle Data Guard Broker is a distributed management framework that automates and centralizes the creation, maintenance, and monitoring of Data Guard configurations. The following list describes some of the operations the broker automates and simplifies:

- Creating Data Guard configurations that incorporate a primary database, a new or existing (physical, logical, or snapshot) standby database, redo-transport services, and log-apply services, where any of the databases could be Oracle Real Application Cluster (RAC) databases
- Adding new or existing (physical, snapshot, logical, RAC, or non-RAC) standby databases to an existing Data Guard configuration, for a total of one primary database, and from one to nine standby databases in the same configuration
- Managing an entire Data Guard configuration, including all databases, redo-transport services, and log-apply services through a client connection to any database in the configuration
- Managing the protection mode for the broker configuration
- Invoking switchover or failover with a single command to initiate and control complex role changes across all databases in the configuration
- Configuring failover to occur automatically upon loss of the primary database, increasing availability without manual intervention
- Monitoring the status of the entire configuration, capturing diagnostic information, reporting statistics such as the redo apply rate and the redo generation rate, and detecting problems quickly with centralized monitoring, testing, and performance tools.

The test environment consists of a two-node Oracle RAC cluster with a primary database that is approximately 200 GB in size, and a two-node Oracle RAC cluster with a standby database. Both the primary and standby servers have Oracle software installed on them (Oracle Linux 7.9). The primary RAC server has a running instance, whereas the standby RAC server has only software installed in it. However, both primary and standby database servers should have proper network configurations set up in the listener configuration file, and entries in the listener configuration can be configured using a Network Configuration Utility (NCU) or manually. In the test setup used, network entries are configured manually. Once network and listener configurations are complete, duplicate the active primary database to create a standby database. Create online redo log files in the standby server to match the configurations created on the primary server. In addition to the online redo logs, create standby redo logs on both the standby and primary databases (in case of switchovers). The standby redo logs should be at least as big as the largest online redo log, and there should be one extra group per thread compared to the online redo logs. Once the duplicate database is successful, bring up the standby database instance. Configure Data Guard Broker in both the primary and standby servers to access and manage databases.

Duplicate an active database from the primary database to the standby database:

```
using target database control file instead of recovery catalog allocated channel: p1
channel p1: SID=2190 instance=orcl1 device type=DISK

allocated channel: s1
channel s1: SID=271 device type=DISK

Starting Duplicate Db at 29-MAR-22



Starting backup at 29-MAR-22
Finished backup at 29-MAR-22



Starting restore at 29-MAR-22
channel s1: starting datafile backup set restore
channel s1: using network backup set from service dborcl
channel s1: restoring SPFILE
output file name=/u01/app/oracle/product/19.3/db_home/dbs/spfileorcl.ora channel s1:
restore complete, elapsed time: 00:00:01 Finished restore at 29-MAR-22

Starting restore at 29-MAR-22
channel s1: starting datafile backup set restore
channel s1: using network backup set from service dborcl
channel s1: specifying datafile(s) to restore from backup set
channel s1: restoring datafile 00001 to +DATADG
channel s1: specifying datafile(s) to restore from backup set
channel s1: restoring datafile 00008 to +DATADG
channel s1: restore complete, elapsed time:
```

```
00:00:06 Finished restore at 29-MAR-22


Finished Duplicate Db at 29-MAR-22
```

Set up the Oracle Data Guard Broker configuration on both the primary and standby databases:

```
DGMGRL> validate database dborcl
  Database Role:      Primary database
  Ready for Switchover:          Yes
  Managed by Clusterware:
    dborcl:           YES



DGMGRL> validate database stdbydb
  Database Role:      Physical standby database
  Primary Database:   dborcl
  Ready for Switchover:          Yes
  Ready for Failover: Yes (Primary Running)
  Flashback Database Status:
    dborcl :          On
    stdbydb:          Off
  Managed by Clusterware:
    dborcl :          YES
    stdbydb:          YES
DGMGRL>
```

Switch over from the primary database to the standby database:

```
DGMGRL> show configuration
Configuration - dg_config
  Protection Mode: MaxPerformance
  Members:
  stdbydb - Primary database
 dborcl  - Physical standby database
 Fast-Start Failover:  Disabled
Configuration Status:
 SUCCESS   (status updated 59 seconds ago)

DGMGRL> switchover to dborcl;
Performing switchover NOW, please wait...
Operation requires a connection to database "dborcl"
Connecting ...
Connected to "dborcl"
Connected as SYSDBA.
New primary database "dborcl" is opening...
Oracle Clusterware is restarting database "stdbydb" ...
Connected to "stdbydb"
Connected to "stdbydb"
Switchover succeeded, new primary is "dborcl"

SQL> select name, db_unique_name, database_role, open_mode from v$database;

NAME       DB_UNIQUE_NAME                   DATABASE_ROLE      OPEN_MODE
---------  ------------------------------   ----------------   --------------------
ORCL       stdbydb                          PHYSICAL STANDBY   READ WRITE
```

Note: These tests and several others were performed to validate the functions. The results presented here are not intended to provide detailed sizing guidance or guidance for performance headroom calculations, which are beyond the scope of this document.

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 customers' investments, 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.

With the **iSCSI feature** supported in the latest Cisco HyperFlex HX Data Platform Release 5.0(1b), Oracle application deployments will have an advantage of using the iSCSI shared disk feature to prepare the disks and make it accessible across all the virtual machines in the RAC cluster; further, they can be used for Oracle ASM storage to create disk groups. Also, Oracle backup creation and duplication of a database using RMAN on the remote cluster will be much easier and quick with the iSCSI shared disk feature, because it would eliminate the need to manually transfer the backup copies from the local cluster to the remote cluster. In this reference white paper, the "Engineering validation" section presents the node and user scale performance results tested using iSCSI storage, and the "Oracle Backup using RMAN" section covers backup, duplication, and reliability testing using the iSCSI shared drive feature.

This solution delivers many business benefits, including:

- 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
- The capability to start with a smaller investment and grow as business demand increases
- Enterprise application-ready solution
- Efficient data storage infrastructure
- Scalability
- Shared disk access

For more information

For additional information, consult the following resources:

- Cisco HyperFlex HX Data Platform, Release 5.0: https://www.cisco.com/c/en/us/td/docs/hyperconverged_systems/HyperFlex_HX_DataPlatformSoftware/HyperFlex-Release-Notes/HX-Release-5-0/Cisco-HXDataPlatform-RN-5-0.html
- 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 iSCSI Overview: https://www.cisco.com/c/en/us/products/collateral/hyperconverged-infrastructure/hyperflex-hx-series/hyperflex-iscsi-wp.html
- 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**.

Printed in USA C11-2399753-01 05/22