Hewlett Packard Enterprise

Reference Configuration

HPE Reference Configuration for Oracle 19C Database with VMware VSphere on HPE Synergy and HPE Primera

Consolidation of Oracle databases with VMware vSphere 7.x

CONTENTS

Executive summary ... 3
Introduction ... 3
Solution overview ... 3
HPE Synergy 12000 Frame ... 4
HPE Synergy Gen10 Compute Modules ... 4
HPE Synergy Composer 2 ... 4
HPE Primera 600 Storage ... 5
HPE OneView for VMware vCenter (OV4VC) ... 5
HPE InfoSight for servers and storage ... 6
Oracle Database 19c ... 7
HPE Serviceguard for Linux ... 7
Solution components ... 7
Hardware ... 9
Software ... 9
Best practices and configuration guidance for the solution ... 10
HPE Synergy configuration ... 10
Storage configuration ... 11
Network configuration ... 15
CPU and Memory allocation for Oracle database VMs ... 17
Verifying the optimal configuration of a VM ... 19
Capacity and sizing ... 20
Capacity planning for the Oracle database virtual machines ... 20
Workload description ... 22
Analysis and recommendations ... 22
Summary ... 27
Implementing a proof-of-concept ... 28
Appendix A: Bill of materials ... 28
Appendix B: RHEL kernel settings ... 29
Appendix C: Oracle user account limits ... 30
Appendix D: udev rules for vVols ... 30
Resources and additional links ... 31

EXECUTIVE SUMMARY

Today, large IT organizations managing hundreds of database instances are under constant pressure to reduce operations and licensing costs without sacrificing quality or functionality. One common approach to reduce database sprawl and operational cost is to identify and consolidate all the underutilized databases and databases that are running on age-old Unix systems to a virtualized environment, such as VMware vSphere®, which is the industry's leading server virtualization software by VMware technology. A virtualized environment enables IT administrators, to allocate only the required compute resources and scale-out, as the workload demands more resources. It also enables internal organizations to share license costs and operation costs among them.

Hewlett Packard Enterprise (HPE) is the global leader in server virtualization solutions based on VMware vSphere, and HPE Synergy is the most prevalent and composable hardware infrastructure for workload virtualization. As a composable infrastructure platform, HPE Synergy composes physical and virtual compute, storage, and fabric pools into any configuration for any application through a single management interface, called HPE Synergy Composer 2, powered by HPE OneView.

This Reference Configuration describes the consolidation of Oracle RAC 19c databases and Oracle 19c single-instance databases on VMware vSphere with HPE Synergy 480 2-socket systems for low to medium intensive database workloads. HPE Primera A650 all-flash array is used as storage for Oracle database virtual machines (VM).

The move towards an HPE Synergy-based virtualized environment provides the following advantages for Oracle database deployment:

- HPE Synergy is an open infrastructure, which can accommodate most of the common operating systems, hypervisors, and workloads. Choosing HPE Synergy avoids an Oracle infrastructure lock-in situation.
- HPE Synergy supports a wide range of Intel CPUs. Among them, those having a higher clock speed delivers a higher throughput for the same amount of cores. This reduces the CPU core requirements, which results in reduced license cost.
- Reduced hardware footprint frees up physical data center space and cuts operating costs including power and cooling.
- Bring down IT administration overhead.
- Ability to quickly deploy resources and shuffle them as demand changes.
- Business continuity solutions for Oracle database using HPE Serviceguard for Linux.
- With the help of HPE OneView for VMware vCenter plug-in for VMware vCenter®, IT administrators can view and directly manage relationships between virtual machines (VM) running Oracle database and HPE Infrastructure. This increases productivity and ensures an improved quality of service.
- HPE Synergy composable ecosystem actively supports Ansible, Python, Golang, Terraform, Chef, Puppet, PowerShell, and Ruby SDKs. Using the unified HPE OneView API and these popular tools, IT administrators can deploy and update servers, storage, and networking simultaneously, using a few lines of code.

Target audience: This Reference Configuration is meant for data center managers, infrastructure architects, and DBAs who are considering consolidation to reduce license cost, database sprawl, streamlining lifecycle management, and optimizing resource utilization.

Document purpose: The purpose of this document is to demonstrate the Oracle database deployment on VMware vSphere environment using HPE Synergy and HPE Primera storage and present best practices for configuring the environment.

This Reference Configuration describes solution testing performed in May 2021.

INTRODUCTION

Consolidating Oracle database workloads onto a virtualized shared infrastructure helps to increase the utilization of infrastructure and saves license costs. HPE Synergy is a composable bladed infrastructure that powers any workload within a hybrid cloud environment and most prevalent in HPE server offerings for workload virtualization using VMware vSphere.

SOLUTION OVERVIEW

The key components of this solution are HPE Synergy Composer 2, HPE Synergy Gen10 Compute Modules, HPE Primera A650 all-flash storage, VMware vSphere 7 Update 2, and Oracle 19c Database software.

HPE Synergy 12000 Frame

The HPE Synergy 12000 Frame is the base for an HPE Synergy intelligent infrastructure with embedded management and scalable links that expand to meet business demand. The frame is the base infrastructure that pools resources of compute, storage, fabric, cooling, power, and scalability. IT can manage, assemble, and scale resources on demand by using the HPE Synergy Frame with an embedded management solution combining the HPE Synergy Composer 2 and Frame Link Modules. The HPE Synergy Frame is designed to meet today's needs and future needs with continuing enhancements to compute and fabric bandwidths, including photonics-ready capabilities.

NOTE

For development and Proof of Concept environments, a single HPE Synergy Frame with one or two HPE Synergy Composer 2 Management Appliances may be deployed. For production environments, Hewlett Packard Enterprise recommends the deployment of two or more HPE Synergy Frames, which allows for frame-level redundancy by placing two management appliances in separate frames.

HPE Synergy Gen10 Compute Modules

HPE Synergy Gen10 Compute Modules deliver more choice for performance, capacity, efficiency, and flexibility to power most workloads with support for the full range of Intel® Xeon® Scalable Family processors in a 1S, 2S, or 4S form factor. They provide a 25% performance increase over prior generations. The compute modules are designed to create a pool of flexible compute capacity within a composable infrastructure. In addition, Gen10 iLO 5 and silicon root of trust are designed to meet challenges such as attacks on firmware, ensuring effective security protections.

The HPE Synergy 480 Gen10 Compute Module comes in a two-socket, half-height form factor to support demanding workloads. Powered by the 2nd generation of Intel® Xeon® Scalable processors, HPE Synergy 480 Gen10 utilizes HPE DDR4 SmartMemory supporting up to 3TB. It has flexible storage controller options with up to two SFF drives (4 uFF drives) and up to two internal M.2 drives and three I/O mezzanine slots. It is designed to create a pool of flexible compute capacity within a composable infrastructure, the HPE Synergy 480 Gen10 Compute Module is the ideal platform for general-purpose enterprise workload performance now and in the future. Refer to the HPE Synergy 480 Gen10 Compute Module QuickSpecs for more details.

FIGURE 1. HPE Synergy 480 Gen10 Compute Module

HPE Synergy Composer 2

HPE Synergy Composer 2 provides an enterprise-level management user interface to compose and deploy system resources to match your application needs. This management appliance uses software-defined intelligence with embedded HPE OneView to aggregate Compute, Storage, and Fabric resources in a manner that scales to your application needs, instead of being restricted to the fixed ratios of traditional resource offerings. It simplifies IT operations and allows fast deployment and configuration. Refer to the HPE Synergy Composer QuickSpecs for more details.

HPE Primera 600 Storage

HPE Primera 600 Storage is a Tier-0 enterprise storage solution designed for simplicity, resiliency, and performance when running mission-critical applications and workloads. Built upon proven resiliency and powered by the intelligence of HPE InfoSight, HPE Primera 600 Storage delivers instant access to data with storage that sets up in minutes, upgrades transparently, and is delivered as a service. Ensure always fast and always on storage for all mission-critical applications. HPE Primera 600 Storage is a parallel, multi-node, all-active platform that achieves predictably high performance at extremely low latency.

HPE Primera 600 storage systems include the HPE Primera 630, HPE Primera 650, and HPE Primera 670. Each model is available in an all-flash-drive configuration or a converged-drive configuration. The converged drive configuration supports a mix of flash and no flash physical drives. All models come with factory-installed HPE Primera OS and HPE Primera UI.

The following summarizes basic configurations. It does not include information about available add-on drive enclosures and other options. For more information on the number of host ports and maximum storage capacities, see the HPE Primera 600 Storage QuickSpecs.

- **HPE Primera A630 and C630:** The HPE Primera A630 is an all-flash-drive configuration and the HPE Primera C630 is a converged-drive configuration. The 630 includes a 2U base enclosure that contains two (2) storage controllers. The 2U base enclosure can hold up to 24 small form factor physical drives.
- **HPE Primera A650 and C650:** The HPE Primera A650 is an all-flash configuration and the HPE Primera C650 is a converged flash configuration. The 650 is available with the 2U or 4U base enclosure that contains either two (2) or four (4) storage controllers. The 2U and 4U base enclosure can hold up to 24 or 48 small form factor physical drives.
- **HPE Primera A670 and C670:** The HPE Primera A670 is an all-flash-drive configuration and the HPE Primera C670 is a converged-drive configuration. The 670 is available with the 2U or 4U base enclosure that contains either two (2) or four (4) storage controllers. The 2U and 4U base enclosures can hold up to 24 or 48 small form factor physical drives.

FIGURE 2. HPE Primera 600 Storage Series

Management Interfaces for HPE Primera:

- **HPE Primera UI:** The HPE Primera UI is a graphical user interface for managing a single HPE Primera 600 Storage system. HPE Primera UI software is included in each HPE Primera 600 Storage system and does not require installation on a server.
- **HPE Primera CLI:** HPE Primera CLI is a text-based command-line interface for managing one HPE Primera 600 Storage system at a time. The functionality is included in the HPE Primera OS on HPE Primera 600 Storage systems. HPE Primera CLI client software can be installed on hosts (servers) running various computer operating systems.
- **HPE 3PAR SSMC (StoreServ Management Console):** HPE 3PAR SSMC is a graphical user interface for managing multiple HPE Primera 600 Storage systems or HPE 3PAR StoreServ Storage systems or both at a time. The software is available as a virtual appliance and can be downloaded from the HPE Software Depot. The software can be deployed in several supported virtual machine environments.

HPE OneView for VMware vCenter (OV4VC)

HPE OneView for VMware vCenter is a plug-in for VMware's vCenter management console, which enables vSphere administrators to quickly obtain context-aware information about HPE servers and HPE storage in their VMware vSphere environment directly from within vCenter. This enables the VMware vSphere administrators to easily manage physical servers and storage, data stores, and virtual machines, all from a single familiar console. By providing the ability to clearly view and directly manage relationships between virtual machines, and HPE Infrastructure, the VMware administrator's productivity increases, as does the ability to ensure the quality of service.

Figure 3 shows the network topology for the deployed virtual machines and HPE Synergy Compute Module within the VMware vCenter console.

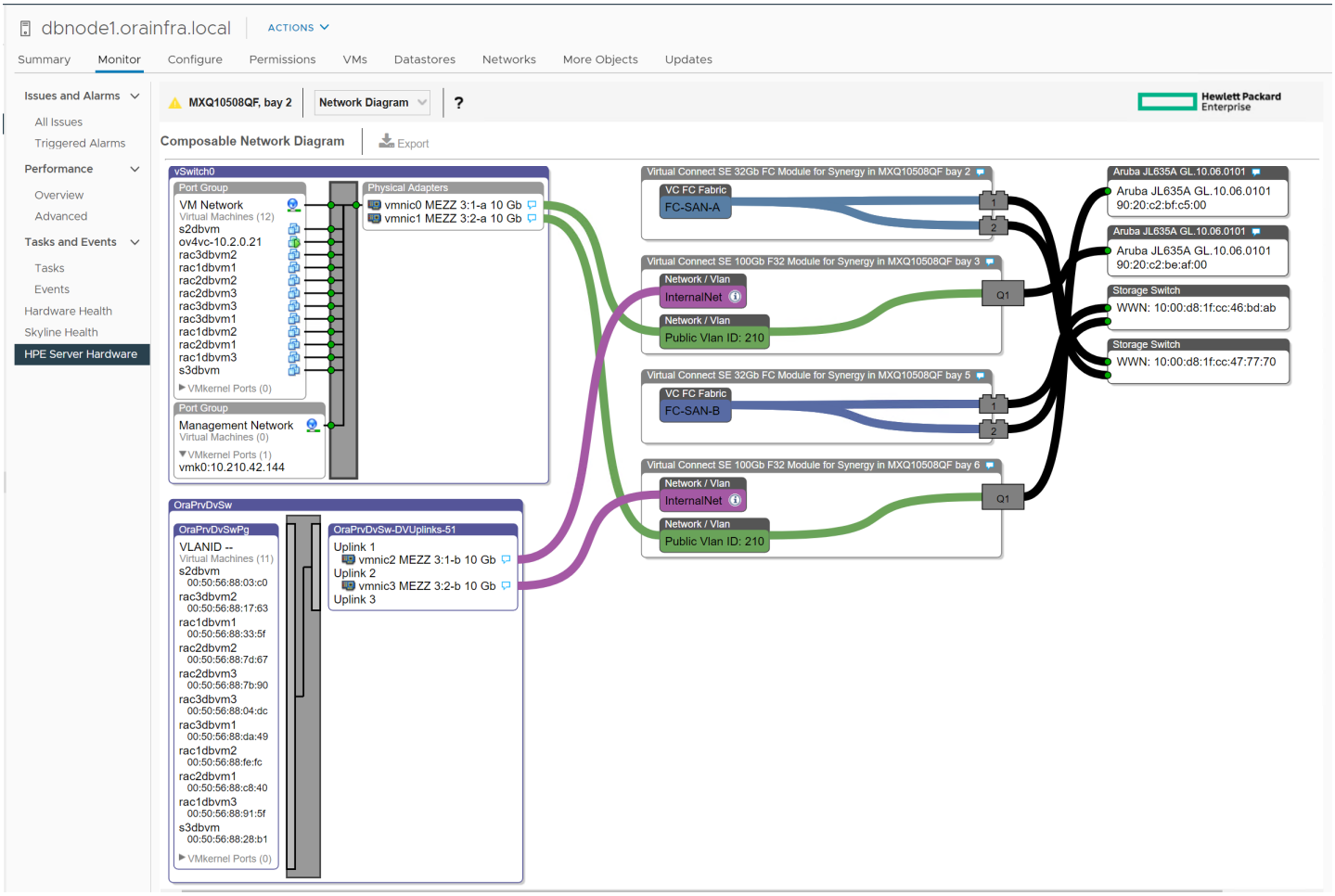

FIGURE 3. HPE OneView for VMware vCenter displaying network topology for the VMs deployed on HPE Synergy Compute Module

One of the key capabilities of HPE OneView for VMware vCenter is to increase capacity easily and consistently on demand for provisioning and deployment. It enables you to deploy ESXi hosts on bare-metal HPE ProLiant and HPE Synergy servers directly from VMware vCenter with a few simple steps to expand an existing cluster, or easily deploy a new cluster. This capability is referred to as the Grow Cluster feature. Refer to the HPE OneView for VMware vCenter Grow Cluster White Paper for more information.

HPE InfoSight for servers and storage

HPE InfoSight is a secured web interface hosted by Hewlett Packard Enterprise that allows you to monitor supported devices through a graphical interface. It uses artificial intelligence to provide global insights into the status and health of infrastructure and watches over the customer infrastructure 24x7, so they don't have to spend days, nights, and weekends dealing with infrastructure issues anymore.

HPE InfoSight for servers combines the machine learning and predictive analytics of HPE InfoSight with the health and performance monitoring of Active Health System (AHS) and HPE iLO to optimize performance and predict and prevent problems. For more information on getting started and using HPE InfoSight for servers, go to https://www.hpe.com/info/infosight-servers-docs.

Refer to HPE InfoSight for HPE Storage User Guide for more details on HPE Primera support for HPE Infosight.

Oracle Database 19c

Oracle Database 19c[1] builds upon the innovations of previous releases such as Multitenant, In-Memory, JSON support, Sharding, and many other features. This release introduces new functionality, providing customers with a multi-model enterprise-class database for all their typical use cases, including:

- Operational database use cases such as traditional transactions, real-time analytics, JSON document stores, and Internet of Things (IoT) applications.
- Analytical database use cases such as traditional and real-time data warehouses and data marts, big data lakes, and graph analytics.

HPE Serviceguard for Linux

HPE Serviceguard for Linux (SGLX) increases the availability and uptime of your mission-critical applications, allows business continuity during maintenance windows, and minimizes the impact of unplanned outages. Providing both HA and DR features, HPE SGLX packages applications and other services with their associated resources and monitors the entire package for any failure. Each package is monitored for faults related to hardware, software, OS, virtualization layer, virtual machine guests, network, and storage. When any failure is detected, SGLX shuts down the application quickly, and smartly relocates the application or service to another system with the necessary resources to bring it into production again. An HPE SGLX cluster allows for multiple node failures using redundant hardware to remove any single point of failure.

Refer to HPE Serviceguard for Linux with VMware Virtual Machines for more details on deploying highly available Oracle single-instance databases on the VMware platform.

SGLX-based HA and DR solution for Oracle Database

HPE SGLX provides automatic HA and DR protection for Oracle databases. This solution manages and completely automates the transition of Oracle databases from one cluster node to another in the event of a failure. When utilized without Oracle Data Guard (ODG), Oracle is installed on each of the cluster nodes and the databases are configured such that the database data is located on shared storage which can be accessed from all the cluster nodes. HPE Serviceguard deep integration with Oracle allows SGLX to start, stop, monitor, and recover the databases on the adoptive cluster nodes in case of failures.

SGLX-based HA and DR solution for Oracle Database using ODG

HPE SGLX along with ODG provides HA and DR protection for Oracle databases with much faster recovery times as compared to the solution with just SGLX (without ODG). HPE SGLX manages and completely automates the role transition process of ODG databases in the event of a failure. In this solution, primary and standby databases are configured with ODG. SGLX is utilized to start, stop, and monitor the databases and monitor the replication between primary and standby databases. In case of failures, the solution automatically recovers the Oracle Database by promoting the standby database instance to primary.

SOLUTION COMPONENTS

The HPE Synergy components used in this solution included a single HPE Synergy 12000 Frame, with two HPE Synergy Composer 2, four HPE Synergy 480 Gen10 compute modules, two HPE Virtual Connect SE 100 Gb F32 Modules for Synergy, and two HPE Virtual Connect SE 32Gb Fibre Channel Modules for Synergy. For this solution, one HPE Primera A650 all-flash storage with two 32G Fibre Channel switches was used for the guest OS deployment and Oracle database tablespaces, indexes, and logs using VMware vSphere vVols. For networking, two Aruba switches were deployed.

[1] Oracle Database 19c Introduction and Overview

Figure 4 shows the hardware components inside a rack that form the solution.

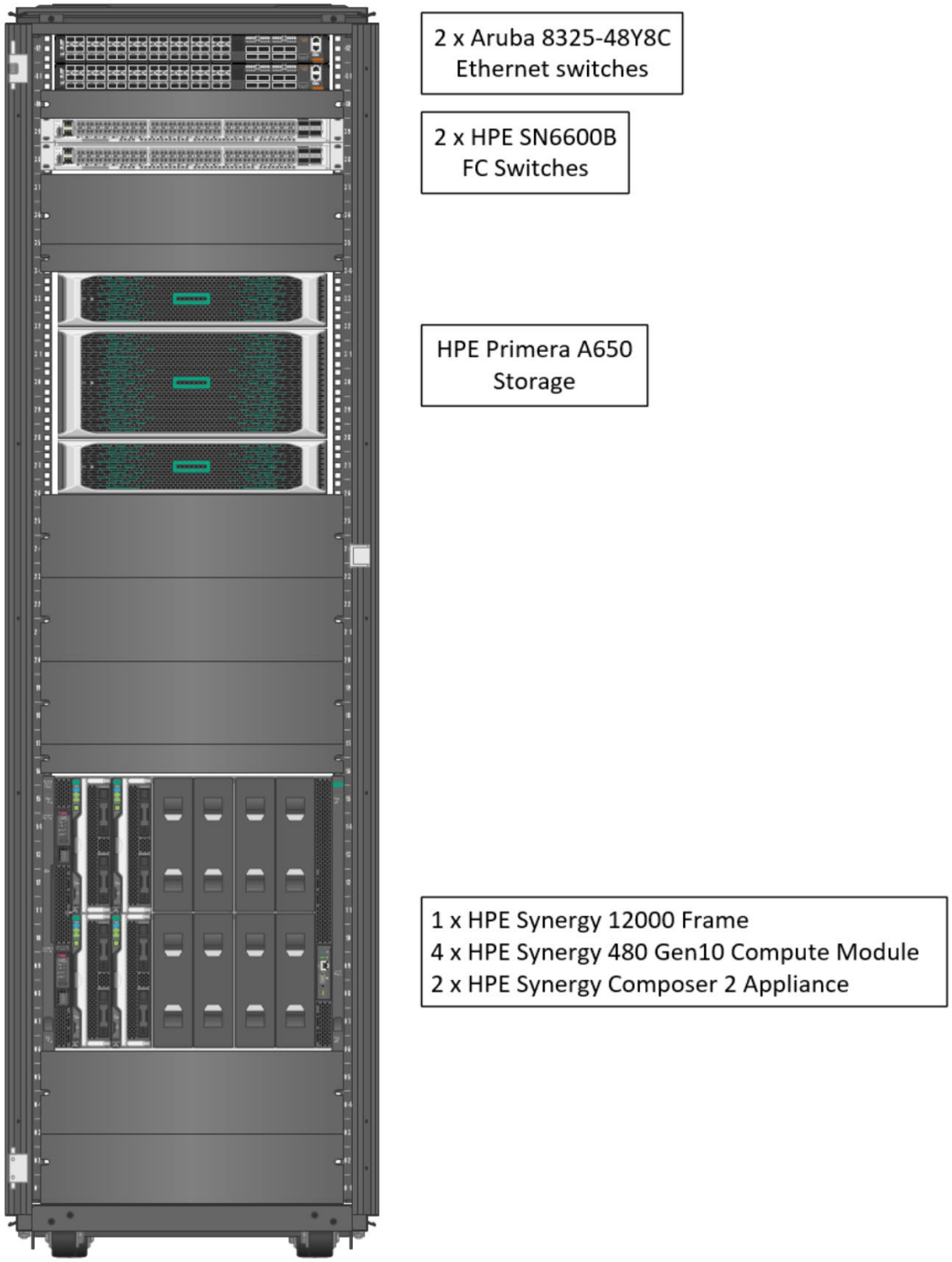

FIGURE 4. Front view of the rack installed with the solution components

Hardware

Table 1 summarizes hardware components that were utilized in the design and construction of this Reference Configuration.

TABLE 1. Hardware components utilized in this solution

Component	Qty	Description
HPE Synergy 12000 Frame	1	HPE Synergy 12000 Frame houses the composable infrastructure used for the solution.
HPE Synergy Composer 2	2	HPE Synergy Composer 2 for configuring and performing lifecycle management of the HPE Synergy components.
HPE Virtual Connect SE 100Gb F32 Module	2	HPE Virtual Connect SE 100Gb F32 Module for Synergy provides Ethernet network connectivity in and out of the frame.
HPE Virtual Connect SE 32Gb Fibre Channel Module	2	HPE Virtual Connect SE 32Gb Fibre Channel Module for Synergy provides FC connectivity between external FC SAN storage (HPE Primera A650) and compute modules
HPE Synergy 480 Gen10 Compute Module • 2 x Intel® Xeon® Gold 6246R Processor with 16 cores and max turbo frequency 4.10 GHz • 12 x 64 GB DRAM DIMMs • 1 x HPE Synergy 5830C 32Gb FC HBA • 1 x HPE Synergy 6820c 25/50Gb CNA	4	One (1) VMware ESXi Server hosting VMware vCenter appliance, HPE OneView for VMware vCenter Server Appliance, and DNS & NTP servers for Oracle RAC infrastructure. Three (3) VMware ESXi Servers forming a vSphere cluster.
HPE Primera A650 Storage • 4 Controller nodes • 48 x 1.92 TB SSD • 8 x 32 Gb 4 Port FC HBAs	1	Storage for VMs and Oracle databases
HPE SN6600B 32Gb	2	Fibre channel switches for providing FC connectivity between SAN storage and Compute Module via VC SE 32Gb FC Modules.
Aruba 8325-48Y8C 48p 25G • 48 Ports x 1G/10G/25 GbE • 8 Ports x 40G/100 GbE	2	Gigabit Ethernet switches, also known as Top of the Rack (ToR) switches for providing Ethernet connectivity between the datacenter and Compute Modules via HPE VC SE 100Gb F32 Modules. This connectivity is referred to as **uplink**. In this Reference Configuration, one 100 Gb port from each VC module was connected to ToR switches and Link Aggregation (LAG) was configured for load balancing and redundancy. **NOTE:** Within the frame, connectivity between HP VC SE 100Gb F32 Modules and Compute Module is known as **downlink**.

Software

Table 2 summarizes software components that were used in this solution configuration.

TABLE 2. Software used in this solution

Component	Version	Description
HPE Synergy Custom SPP	2021.02.01	HPE Synergy Custom SPP – Contains firmware and drivers for Synergy hardware
RHEL	8.3	Red Hat® Enterprise Linux® operating system deployed on virtual machines running Oracle databases
VMware ESXi™ Server	7.0 U2	VMware vSphere ESXi Server installed on compute modules using HPE Custom Image for ESXi 7.0 U2 Install CD
VMware vCenter	7.0 U2	VMware vCenter appliance for managing VMware vSphere environment
HPE OneView for VMware vCenter (OV4VC)	10.2	HPE OneView for VMware vCenter seamlessly integrates the manageability features of HPE Synergy with VMware solutions.

Application software

Table 3 shows the application software used in this solution configuration.

TABLE 3. Application Software used in this solution

Component	Version	Description
Oracle Database	19c	Oracle Database 19.3.0.0.0 Enterprise Edition with the RAC option and following patches • GRID Patch p32226239 (19.10) • DB SW Patch p32218454 (19.10)

BEST PRACTICES AND CONFIGURATION GUIDANCE FOR THE SOLUTION

As a first step, the HPE Synergy hardware environment must be set up with the required networks and storage connectivity for ESXi server implementation, which is also known as server profile deployment.

HPE Synergy configuration

Upon completion of HPE Synergy hardware deployment, proceed with the network, storage, and server profile configuration. Below are the high-level steps:

Steps to configure HPE Primera storage system in HPE Synergy Composer 2 - OneView

1. Add HPE Primera storage system under Storage Systems.
2. Add 32Gb Fibre Channel Switch under SAN Managers in OneView for automatic zone creation.

Steps to configure networks in HPE Synergy Composer 2 - OneView

1. Upload the latest HPE Synergy Custom SPP bundle under Firmware Bundles.
2. Create Ethernet networks with the required VLANs.
3. Create FC networks.
4. Create Logical Internet Group (LIG) for Ethernet and Fibre channel virtual connect (VC) modules, and add uplink set with above-configured networks.
5. Create Enclosure Group (EG) using the above created LIGs and allocate IPs for VC modules.
6. Configure Link Aggregation Group (LAG) in top of the rack (ToR) switches for the Ethernet VC modules.
7. Create Logical Enclosure (LE) using the above created EG and select the firmware bundle for the compute modules.

Steps to deploy server profile for compute modules in HPE Synergy Composer 2 - OneView

1. Click the Create profile under Server Profiles.
2. Provide a name for the profile and select the compute module for profile deployment.
3. Select firmware baseline, if required.
4. Configure Ethernet networks and Fibre channel storage connections.
5. Configure local storage – RAID 1 logical drive for OS installation on compute module.
6. Configure SAN storage, as required for VMware datastore(s) creation. In this Reference Configuration, VMware vVols were deployed for virtual machines (VM). At the time of the VM creation process, LUNs are automatically created in Primera by VMware vCenter. Refer to the Storage configuration section for more details.
7. Configure boot settings.
8. Configure BIOS settings for workload profile – "**Virtualization–Max Performance**".
9. Apply.

Storage configuration

In the VMware vSphere environment, traditionally shared disks for Oracle RAC were deployed using virtual disk (.vmdk) files, located on a VMware datastore. In this Reference Configuration, VMware vSphere vVols were used for deploying shared VMDKs for all the Oracle RAC databases. The vVols are much easier to manage and created by vCenter using the VASA (vSphere APIs for Storage Awareness) protocol instructing the HPE Primera storage array to present units of storage to a VM. VMware vSphere supports sharing of vVols across Oracle RAC database VMs and it is known as the **multi-writer** option in the VM settings.

Refer to VMware vVols With HPE Primera Storage document for setting up vVols. A VM Storage Policy called "Primera" was configured in vCenter and it was used in all the Oracle RAC database VMs for vVol creation.

Following are the VM Storage Policy parameters used for the testing:

- **Common Provisioning Group:** SSD_r6
- **Snapshot Common Provisioning Group:** SSD_r6
- **Thin Persistence:** Enabled
- **Data Reduction:** Disabled (Default value)

Figure 5 shows the VM Storage Policy parameters configuration in vCenter.

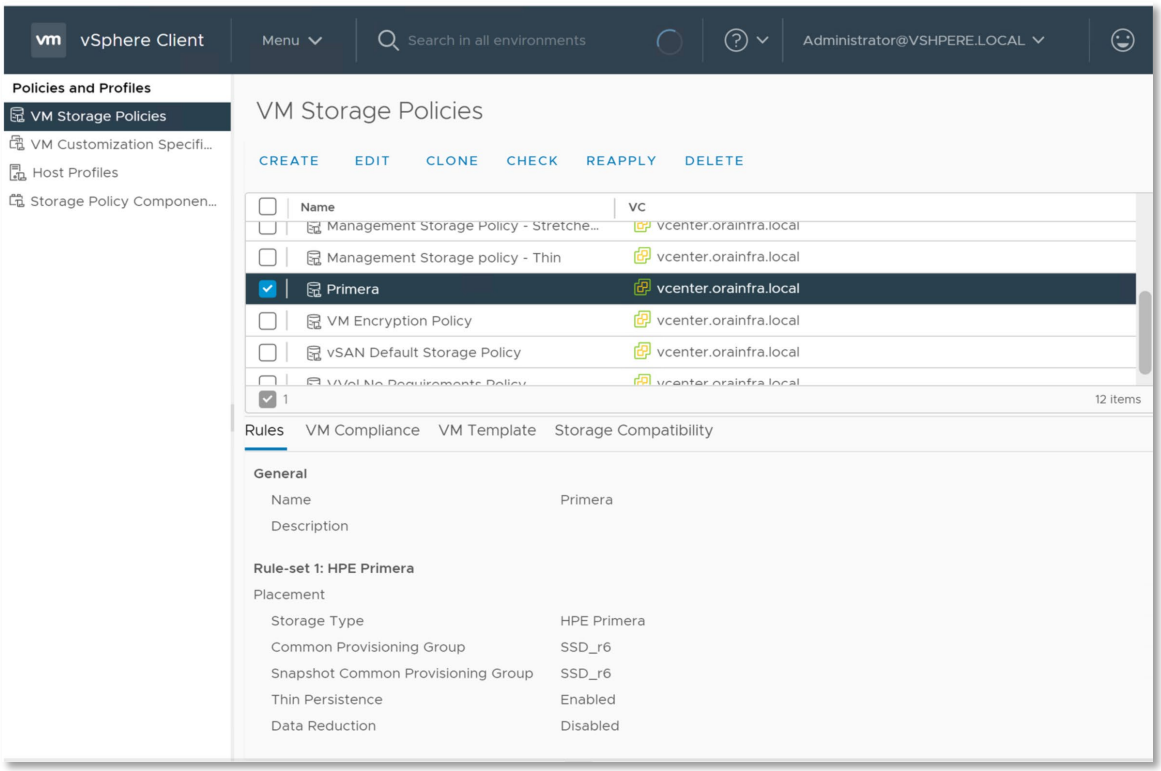

FIGURE 5. VMware vCenter - VM storage policy for HPE Primera vVol deployment

ESXi storage connections

Table 4 shows the compute module storage connection settings in the server profile.

TABLE 4. Storage connections for ESXi server(s)

OneView network name	Compute module FC HBA ports	Server profile connection name	Bandwidth	ESXI vmhba
SAN-A	FC HBA Port1	FC1	32 Gbps	vmhba1
SAN-B	FC HBA Port2	FC2	32 Gbps	vmhba2

Multipathing configuration for ESXi host(s)

In this Reference Configuration testing, the following multipathing policy was used for HPE Primera:

- **Path Selection Policy:** Round Robin (VMware)
- **Storage Array Type Policy:** VMW_SATP_ALUA

All multipathing policies are applied to a Protocol Endpoint LUN (PE LUN). This means, if path failover happens, it applies to all vVols bound on that PE LUN. Refer to HPE Primera: VMWare ESXi Implementation Guide for more details.

Virtual machine storage configuration

Oracle RAC database VMs were configured with the following storage layout and settings:

1. Disks for the operating system (OS), Oracle binaries (/u01 xfs filesystem), and shared disks for Oracle RAC ASM disk groups:
   - Deploy a new virtual machine, for example, **rac1dbvm1**
   - By default, a hard disk is added for the OS installation. In addition to that add one more hard disk by selecting "Hard Disk" in "**Edit Settings** > **ADD NEW DEVICE**". Now, set the following parameters for both hard disks:
     - Set the hard disk size as per the requirement
     - VM storage policy: **Primera**
     - Location: Store with the virtual machine
     - Disk Provisioning: **Thin Provision**
     - Sharing: **Unspecified**
     - Shares: Normal (Default value)
     - Limit – IOPs: Unlimited
     - Disk Mode: **Dependent**
     - Virtual Device Node: [Select the SCSI controller 0 for both disks]
   - Now, add as many as hard disks required for Oracle ASM disk groups by selecting "Hard Disk" in "**Edit Setting**s > **ADD NEW DEVICE**" and set their parameters as shown below:
     - Set the hard disk size as per the requirement
     - VM storage policy: **Primera**
     - Location: Store with the virtual machine
     - Disk Provisioning: **Thin Provision**

NOTE

There are two options available for disk provisioning here, namely Thin Provision and Thick Provision. When Thin Provision is selected, HPE Primera creates a Thin Provision volume. When Thick Provision is selected, HPE Primera creates a Thick Provision Lazy Zeroed volume. In the testing, Thin Provision was used.

- Sharing: **Multi-writer**
- Shares: Normal (Default value)
- Limit – IOPs: Unlimited
- Disk Mode: **Independent – Persistent**

NOTE
The Independent-Persistent disk mode is not mandatory here. But, the default Dependent disk mode would cause the "cannot snapshot shared disk" error when a VM snapshot is taken. Use of Independent-Persistent disk mode would allow taking a snapshot of the OS disk and non-shared disks (for example, disk with /u01 filesystem on it and single-instance Oracle database disks), while the shared disks would need to be backed up separately by the Oracle RAC database backup and recovery methods.

- Virtual Device Node: Select the SCSI controller 1 for Oracle ASM MGMT disk group; Select the SCSI controller 1 for Oracle ASM DATA disk group; SCSI controller 2 for REDO_A disk group; SCSI controller 3 for REDO_B disk group.

– Deploy other RAC VMs, for example, **rac1dbvm2** and **rac1dbvm3** and add the above created Multi-writer enabled disks by selecting "Existing Hard Disk" in "**Edit Settings** > **ADD NEW DEVICE**" and select the appropriate .vmdk file by navigating to the rac1dbvm1's folder location. Set the hard disk parameters as already mentioned for **rac1dbvm1**.

2. For all the Oracle RAC database VMs, the SCSI controller type was set to **VMware Paravirtual** and the SCSI bus-sharing was set to **None** for all the controllers shown below:

– SCSI controller 0: For the Operating System and /u01 filesystem (Oracle binaries)

– SCSI controller 1: For the disks belonging to the Oracle ASM DATA disk group

– SCSI controller 1: For the disks belonging to the Oracle ASM MGMT disk group

– SCSI controller 2: For the disks belonging to the Oracle ASM REDO_A disk group

– SCSI controller 3: For the disks belonging to the Oracle ASM REDO_B disk group

Table 5 lists the storage layout for an Oracle RAC database (rac1db) containing three Oracle RAC nodes. In the test, three separate Oracle RAC databases (rac1db, rac2db, and rac3db) were created with a similar storage layout, demonstrating the Oracle RAC consolidation scenario.

TABLE 5. Storage layout configured for an Oracle RAC database deployment

Oracle RAC virtual machine	vVol allocated to virtual machines	Virtual SCSi device id	Description
Oracle RAC database: rac1db / RAC VMs: rac1dbvm1, rac1dbvm2, rac1dbvm3			
rac1dbvm1	OS: 100 GB /u01: 100GB	SCSI 0:0 SCSI 0:1	For operating system For Oracle binaries
rac1dbvm2	OS: 100 GB /u01: 100GB	SCSI 0:0 SCSI 0:1	For operating system For Oracle binaries
rac1dbvm3	OS: 100 GB /u01: 100GB	SCSI 0:0 SCSI 0:1	For operating system For Oracle binaries
rac1dbvm1 rac1dbvm2 rac1dbvm3	8 x 128 GB (shared)	SCSI 1:0 to SCSI 1:7	Used for Oracle RAC ASM DATA disk group
rac1dbvm1 rac1dbvm2 rac1dbvm3	1 x 50 GB (shared)	SCSI 1:9	Used for Oracle RAC ASM MGMT disk group
rac1dbvm1 rac1dbvm2 rac1dbvm3	4 x 64 GB (shared)	SCSI 2:0 to SCSI 2:3	Used for Oracle RAC ASM REDO_A disk group
rac1dbvm1	4 x 64 GB	SCSI 3:0	Used for Oracle RAC ASM REDO_B disk group

Oracle RAC virtual machine	vVol allocated to virtual machines	Virtual SCSi device id	Description
rac1dbvm2 rac1dbvm3	(shared)	to SCSI 3:3	

Table 6 lists the storage layout for a single instance Oracle database (s1db). In the test, three separate single-instance Oracle databases (s1db, s2db, s3db) were created with a similar storage layout, demonstrating a consolidation scenario.

TABLE 6. LUNs exported for the single instance Oracle database deployment

Oracle db virtual machine	vVol allocated to the virtual machine	Virtual SCSi device id	Description
Oracle database: s1db / VM: s1dbvm			
s1dbvm	OS: 100 GB /u01: 100GB	SCSI 0:0 SCSI 0:1	For operating system For Oracle binaries
s1dbvm	8 x 128 GB	SCSI 1:0 to SCSI 1:7	Used for Oracle ASM DATA disk group
s1dbvm	4 x 64 GB	SCSI 2:0 to SCSI 2:3	Used for Oracle ASM REDO_A disk group
s1dbvm	4 x 64 GB	SCSI 3:0 to SCSI 3:3	Used for Oracle ASM REDO_B disk group

Figure 6 shows the VM placement in HPE Synergy Compute Modules and storage layout for the databases.

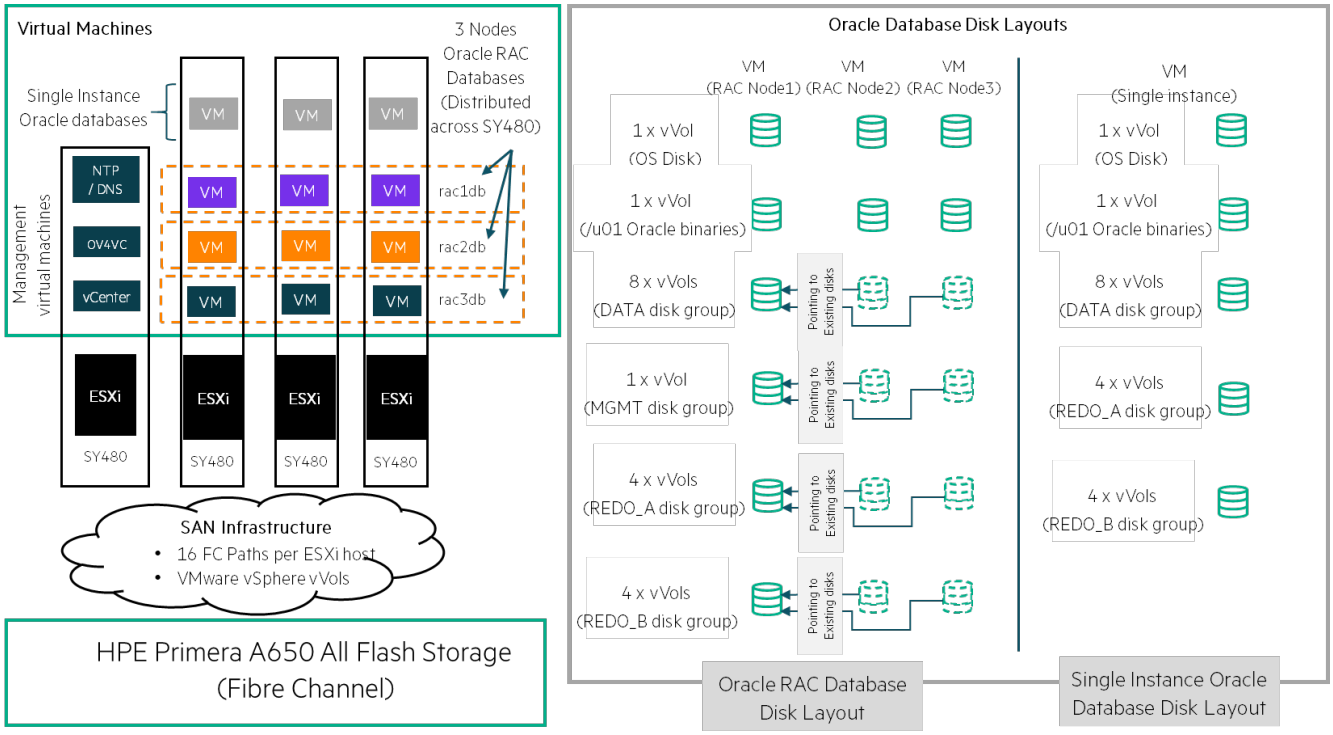

FIGURE 6. Oracle database VM placement in HPE Synergy Compute Modules and storage layout for the databases

Udev rules for vVols

By default, VMware vSphere doesn't provide the information needed for generating `/dev/disk/by-id` device file in the guest OS. But, this is mandatory for udev to create a persistent name(s) for the disk device(s). Hence, administrators must perform the following to generate this required device file:

1. Start the vSphere Client, and log in to a vCenter Server.
2. Power off the VM.
3. Go to **Edit Settings** and click the **VM Options** tab.
4. Select the **Advanced** and click "**EDIT CONFIGURATION**" in Configuration Parameters.
5. Click the "**ADD CONFIGURATION PARAMS**".
6. In the Name column, enter **disk.EnableUUID.**
7. In the Value column, enter **TRUE**.
8. Click **OK** and click **Save**.
9. Power on the VM.

Figure 7 shows the configuration screen where `disk.EnableUUID` is added for the VM.

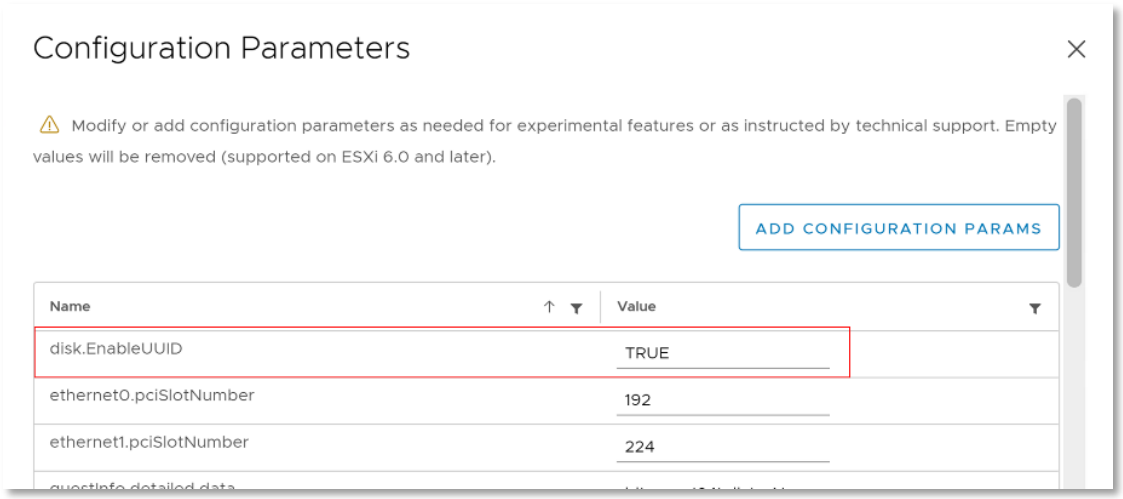

FIGURE 7. Setting up UUID for the disks

After rebooting the VM, the below command would return the WWID for the disk and it can be used for udev rules. Refer to APPENDIX D: UDEV RULES for the udev rules, which were deployed in this Reference Configuration.

```
[root@rac1dbvm1 ~]# /usr/lib/udev/scsi_id -verbose -g -u -d /dev/sdd
36000c29925a60256a91d1f60198f1437
```

Network configuration

All the nodes in the Oracle RAC cluster must be able to communicate via the following networks using TCP/IP.

- **Public Network:** Enables communication between database clients and the nodes in the RAC cluster.
- **Private Network:** This network is used for communications between instances running on the three nodes and not accessible from outside. This network is also known as a cluster interconnect.

For the single instance database nodes, only the public network is required and it is used for communication between database clients and the database node. Table 7 shows the uplink and internal network configuration for Oracle database deployment.

TABLE 7. Uplink and internal network configuration for Oracle database deployment

Oracle database network requirements	OneView network name	VLAN	Subnet id	Uplink set	Uplinks - link aggregation (multi-chassis lag)	Description
Public	vlan210	210	10.210.40.0	UplinkSet1	LAG1 consisting of the following uplink ports: • 1 x 100 Gb port from slot 3: VC 100G module • 1 x 100 Gb port from slot 6: VC 100G module	Enables communication between database clients and the database nodes. Based on the needs, more uplinks can be added to the LAG. Aruba switches also need to be configured for the LAG setup.
Private	InternalNet	None	192.168.1.0	None	None	Communication between RAC database nodes only. Not accessible from the public network and network packets are forwarded among the nodes within the HPE Synergy Virtual Connect modules themselves.

Table 8 shows the Compute Module network configuration. Networks for vMotion are not shown here.

TABLE 8. Compute modules network configuration for Oracle database deployment

OneView network name	Compute module network ports (physical functions)	Server profile connection name	Bandwidth	ESXi vmnic	VMware virtual switch	VMware network name	Vm vnic connected to	Description
mgmtnet	3:1-a	Mgmt1	2.5 Gbps	vmnic0	vSwitch0	Management Network	None	For VMware vSphere management network
mgmtnet	3:2-a	Mgmt2	2.5 Gbps	vmnic1	vSwitch0			
vlan210	3:2-b	Public1	10 Gbps	vmnic2	vSwitch1	VM Network	vnic1	For Oracle public network (RAC / Single instance databases). vnic was configured with the below settings: • Type: VMXNET 3 • DirectPath I/O: Enable
vlan210	3:2-b	Public2	10 Gbps	vmnic3	vSwitch1			
InternalNet	3:1-c	PrvNet1	10 Gbps	vmnic4	dvSwitch1	OraPrvDvSwPg (Jumbo frame enabled)	vnic2 (Jumbo frame enabled at OS level)	For Oracle RAC cluster interconnect. vnic was configured with the below settings: • Type: VMXNET 3 • DirectPath I/O: Enable
InternalNet	3:2-c	PrvNet2	10 Gbps	vmnic5	dvSwitch1			

Enabling jumbo frame for Oracle RAC private network (cluster interconnect)

In Oracle RAC, the cluster interconnect is used for various purposes such as transporting lock management, resource coordination messages, and moving the cache buffers from a node to another node. The setting of DB_BLOCK_SIZE multiplied by the MULTI_BLOCK_READ_COUNT determines the maximum size of a message for the Global Cache and the PARALLEL_EXECUTION_MESSAGE_SIZE determines the maximum size of a message used in Parallel Query.

These message sizes can range from 2K to 64K or more, and hence they can get fragmented more often with the default MTU size of 1500. Therefore, it is recommended to set MTU size to 9000 in all Oracle RAC VMs and distributed virtual switch, which connects the cluster interconnect. Since the cluster interconnect traffic goes within the virtual connect modules only, there is no need to configure the ToR switch for

this. Finally, check it using the ping command as shown below. While using the ping command, we need to take packet overhead (28 bytes per packet) into account. Hence, use 8972 as a size parameter in the ping command and check the VM to VM jumbo frame connectivity.

```
[root@rac1dbvm1 /]# ping -M do -s 8972 rac1dbvm2-prv1
PING rac1dbvm2-prv1.orainfra.local (192.168.1.2) 8972(9000) bytes of data.
8980 bytes from rac1dbvm2-prv1.orainfra.local (192.168.1.2): icmp_seq=1 ttl=64 time=0.108 ms
8980 bytes from rac1dbvm2-prv1.orainfra.local (192.168.1.2): icmp_seq=2 ttl=64 time=0.122 ms
```

CPU and Memory allocation for Oracle database VMs

When consolidating multiple virtual machines on a single ESXi host, make sure that cumulative physical CPU resources on a host are adequate to meet the needs of the VMs by testing the workload in the planned virtualized environment. Here are the recommendations for vCPU and memory allocation for Oracle database workloads.

1. It is generally recommended to not over-allocate vCPUs. If the exact workload is not known, start with fewer virtual CPUs or match the number of vCPUs with the source system's physical core count, if migrating, and increase the vCPU count later if necessary. Verify that proper monitoring processes and procedures are set in place to monitor the system. This can be done through the vSphere Client or by using `esxtop` or `resxtop`.
2. After the workload is established, vCPU over-commitment can be tried with caution. The vCPU over commitment is nothing but allocating more vCPUs to the VMs running on a host than the total number of physical processor cores in that host. Below is how one can calculate the available host CPU resources:

   Number of CPU sockets * Number of cores per processor = Physical Processors (pCPU)

   If hyperthreading is enabled, the number of vCPUs is calculated as **pCPU * 2 threads = Logical Processors**

   Please note that hyper-threading just enables the execution of two threads on a pCPU, where one thread is waiting on I/O or being idle and then the second thread is used. This increases the efficiency of the CPU resources are often in an idle state due to the nature of the workload. This means that the performance of two hyper-threaded cores is not equal to the 2x performance of a pCPU. Hyper-threaded core shares most of the core's resources, such as memory caches and functional units.

NOTE

vSphere will spread out a VM's virtual CPUs across as many physical cores as possible to provide the best performance to a virtual machine and it will try to give every vCPU its own physical core with nothing else scheduled on the secondary hyper-thread. It will first use all the physical cores before using the Hyper-threaded cores and this is the default behavior.

Using **PreferHT** tunable, this behavior can be modified either at the ESXi host level or at the VM level. When enabled, vSphere will use locally available HT cores rather than physical cores from other NUMA nodes for the VM that has vCPU allocation more than the number of physical cores available and less than the HT cores in the current NUMA node. In this Reference Configuration, this parameter was in the default (disabled) state, which means, guest OS was always scheduled with physical cores as long as they are available. This default setting is recommended for better performance.

---

3. Oracle database is a NUMA-aware application and to take advantage of NUMA in a multi-socket system, VMware provides a virtualized NUMA (vNUMA) environment. The maximum performance benefits can be obtained from vNUMA if ESXi clusters are composed entirely of hosts with matching NUMA architecture. If not, when a vNUMA virtual machine is moved to a host with a different NUMA topology, the virtual machine's vNUMA topology might no longer be optimal for the underlying physical NUMA topology, potentially resulting in reduced performance.
4. For the best performance, try to size your virtual machines to stay within a physical NUMA node. For example, if you have a host system with six physical cores per NUMA node, try to size your virtual machines with no more than six vCPUs. When a virtual machine needs to be larger than a single physical NUMA node, try to size it such that it can be split evenly across as few physical NUMA nodes as possible. Refer to VMware documentation for more details on the optimal number of sockets for a given number of vCPUs.

NOTE
By default, vNUMA is activated only for virtual machines with more than eight vCPUs. Beginning with vSphere 6.5, changing the cores per socket value no longer influences vNUMA or the configuration of the vNUMA topology. vNUMA will automatically determine the proper vNUMA topology to present to the guest OS based on the underlying ESXi host. vSphere is aware of Sub-NUMA Clustering (SNC) and Cluster on Die (CoD) feature of processors and accordingly vNUMA will be presented to the guest OS.

5. Below are the vCPU settings used for one of the test cases in this Reference Configuration and Figure 8 shows the same:
   - CPU: 20
   - Cores per Socket: 20
   - CPU Hot Plug: Disabled (checkbox not selected). **Note:** Enabling CPU HotAdd will disable virtual NUMA.
   - Reservation: 0

NOTE
Reservation is set to 0 by default. One can specify a reservation if there is a need to guarantee that the minimum required amount of CPUs are always available for the Oracle database virtual machine. For example, assume you have 50GHz available and specify a reservation of 25GHz for VM1 and 25GHz for VM2. Now each virtual machine is guaranteed to get 25GHz if it needs it. However, if VM1 is using only 15GHz, VM2 can use 35GHz. This is useful in a consolidation scenario where two separate Oracle databases have the highest resource usage at different times of the day.

   - Limit: 0
   - Shares: Normal
   - Hardware virtualization: Yes (checkbox selected)
   - Performance Counters: No (checkbox not selected)
   - Scheduling Affinity: None. **Note:** Do not set CPU affinity as it disables the scheduler's ability to balance load across processors.
   - I/O MMU: Enabled (checkbox selected)

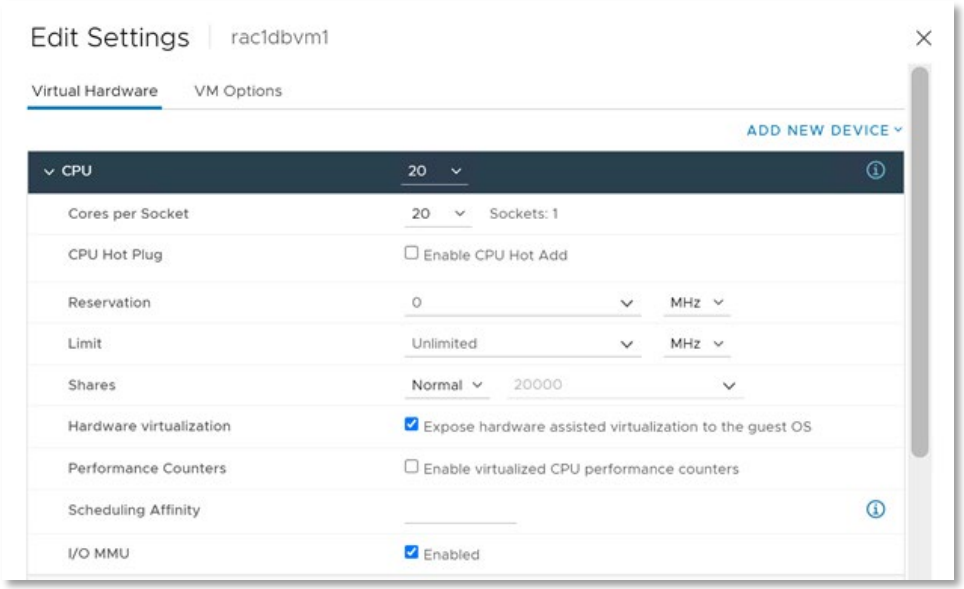

FIGURE 8. Setting up vCPUs

6. Set memory reservations equal to the sum of the total size of the Oracle SGA, the Oracle PGA, and the Oracle Background processes stack space and memory for OS operations. In this Reference Configuration, all the memory in the VMs were reserved as the VMs were dedicated for the Oracle databases only. Figure 9 shows the memory reservation deployed for rac1dbvm1.

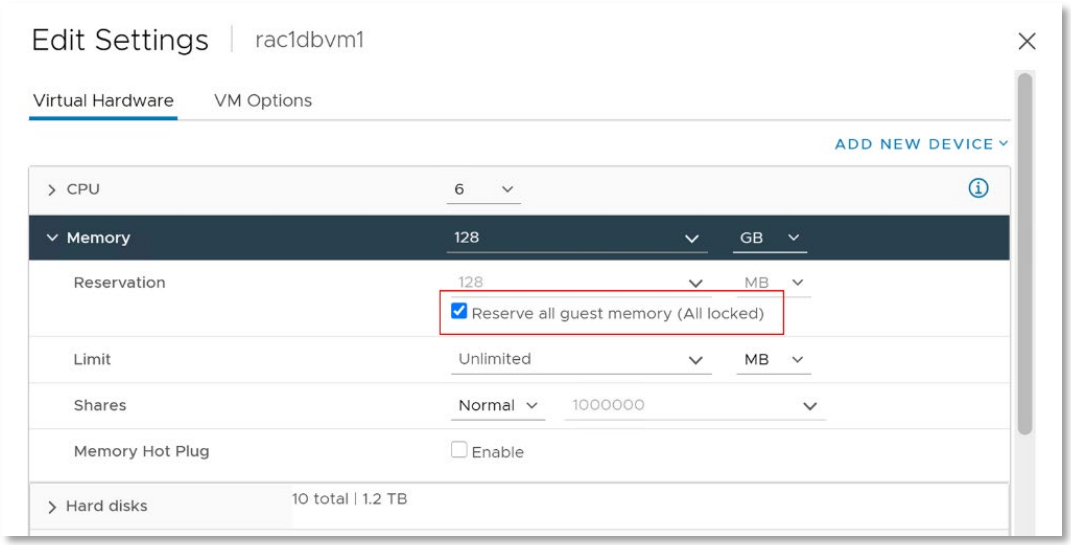

FIGURE 9. Setting up memory for the VM

NOTE
With memory reservation enabled, a VM can be live migrated (vSphere® vMotion®) only if the target ESXi host has free physical memory equal to or greater than the size of the reservation.

Verifying the optimal configuration of a VM

VMware provides a Powershell script known as Virtual Machine Compute Optimizer (VMCO), which captures information about the Hosts and VMs running in your vSphere environment, and reports back on whether the VMs are configured optimally based on the Host CPU and memory. Refer to VMware Virtual Machine Compute Optimizer for more details.

Linux kernel settings for Oracle database VMs

RHEL automatic NUMA balancing was disabled in the Linux kernel (kernel.numa_balancing=0). The number of huge pages (vm.nr_hugepages) was set large enough to hold the Oracle SGA. Specifying USE_LARGE_PAGES=ONLY in the Oracle database ensures the SGA is placed in the HugePages area.

VMware supports large pages at two levels, the hypervisor level, and the guest operating system level. ESXi hypervisor uses 2MB pages for backing guest memory by default and in this Reference Configuration, Oracle database VMs were configured with 2MB huge pages.

Transparent HugePages can cause memory allocation delays at runtime, leading to performance issues. Therefore, Transparent HugePages was disabled in the Linux kernel (transparent_hugepage=never).

Oracle RAC, VMware HA, and HPE Serviceguard for Linux

Oracle RAC and VMware HA solutions are completely complementary to each other. Oracle RAC on a VMware platform provides the application-level high availability and VMware HA provides ESXi host-level high availability. In the event of server failure, affected virtual machines are automatically restarted on other production servers with spare capacity.

HPE Serviceguard for Linux provides application-level high availability for the single instance Oracle database. Refer to HPE Serviceguard for Linux with VMware virtual machines for more details.

CAPACITY AND SIZING

Capacity planning for the Oracle database virtual machines

Consolidation will work only if there are adequate resources allocated to meet the peak load of individual databases. Databases with different resource usage behavior can be consolidated together with fewer resources. For example:

- Databases that have the highest resource usage at different times of the day can be consolidated together.
- Underutilized databases and their sum of peak resource usage do not go beyond the total capacity of a server.

When consolidating multiple virtual machines on one or more ESXi hosts, proper hardware sizing is critical for optimal performance. Following are the few recommendations for measuring capacity utilization prior to consolidation. Please note that measuring capacity utilization begins before each database is migrated to the consolidation environment. To begin with, the following questions would help to determine the right capacity for the consolidated environment:

1. Is the existing database server properly sized, to begin with for the database workload?
2. Was the hardware over-provisioned to support the workload for years?
3. Does the existing system require optimization for better performance? If yes, those optimizations must be considered for the consolidation environment as well.

Capacity usage measures may be gathered at OS level and database level:

NOTE
Hewlett Packard Enterprise has an internal tool called, **HPE Database Performance Profiler II (DPP)**, which provides consolidated performance information for existing Oracle environments. This data is invaluable when designing a new, consolidated Oracle infrastructure. Please contact your HPE Sales representative for more details.

- **OS level** – Check the existing system's capacity usage by database using OS-provided tools such as `sar`, `top`, `vmstat`, and `iostat`, etc. Here the assumption is that the system is dedicated to the Oracle database only. Analyze the logs and determine if the database needs more resources to run better or the currently allocated resources are sufficient.
  - Collect minimum, maximum, and average CPU usage. Database administrators with the help of performance logs collected in the past should be able to provide the range of time in which the system was underutilized or heavily utilized. The idea is to find out the minimum and maximum CPU usage trends. This data helps to find out the appropriate number of vCPUs for the VM.
  - Minimum, maximum, and average memory usage. No paging should be seen. The `vmstat` command can be used to monitor for zero or very low page-in and page-out rates if any.
  - For database storage, determine the total number of disks and their sizes allocated to ASM disk groups, Oracle binaries (ORACLE_HOME). Use `iostat` command to see if there are any high `iowait` and high queue sizes for these disks. If so, then it is better to deploy faster storage and/or more disks for VMs as compared to existing systems.
- **Database level** – In general, database performance depends more on the system capacity allocated to the database than what is available in the system.
  - The below SQL query helps to find out the minimum required CPUs for the database and it can be used to determine the number of vCPUs required for the Oracle database VM. In the Oracle database, `CPU_COUNT` parameter indicates the number of CPUs or CPU threads available for Oracle database use. One can either use this value right away or use the below query to determine the minimum required CPUs. This query provides a report for the past 1 hour period.

    In the query output shown below, the "AVG_RUNNING_SESSIONS" determines the CPU_COUNT utilized for the current workload. However, if "AVG_WAITING_SESSIONS" is greater than zero with any regularity and performance is not acceptable, the CPU_COUNT allocated is not sufficient and should be raised higher.

```
SQL> select to_char(begin_time, 'HH24:MI') time,sum(avg_running_sessions)
avg_running_sessions,sum(avg_waiting_sessions) avg_waiting_sessions from v$rsrcmgrmetric_history group by
begin_time order by begin_time;
```

```
TIME  AVG_RUNNING_SESSIONS AVG_WAITING_SESSIONS
----- -------------------- --------------------
06:31           12.6557912                    0
06:32           13.0541591                    0
06:33           11.2702657                    0
```

- Capture currently allocated SGA (SGA_TARGET) and PGA (PGA_AGGREGATE_TARGET or PGA_AGGREGATE_LIMIT) sizes of the database to calculate the memory required for the database and add more memory for OS requirements.
- Check out Automatic Workload Repository (AWR) and/or Enterprise Manager to see if there are any I/O wait events listed in the top wait events. If so, faster storage may help to reduce the I/O latencies.

Hypervisor – Resource overhead

The hypervisor reserves some CPU and memory capacity for its operations. Table 9 shows the available and allocated resources in the testing. In general, at the time of planning, it is recommended to leave up to 10% to 12% of total clock speed per host (Number of sockets per host x Number of cores per socket x Clock speed per core = Total clock speed per host) for the host operations and use the remaining for virtual machines. It is also important to plan for additional capacity for future growth of workloads.

In this test setup, around 9.27% (108.54 GHz/11.7 GHz = 9.27%) clock speed was automatically reserved for the host. vCenter provides this information at **ESXi host** -> **Configure tab** -> **System Resource Reservation** tab.

TABLE 9. VMware vSphere – Resource overhead

Resources	Per physical host	Available at hypervisor level	System resource reservation per host (vCenter reported)	Available at ESXi cluster level (3 node cluster)	Available for allocation to VM (available – overhead)	3 x node cluster overhead details
CPU (Calculated using CPU clock speed)	2 x Intel(R) Xeon(R) Gold 6246R CPU @ 3.40GHz Physical Core: 2 x 16 Cores	2 x 16 x 3.40 GHz = 108.8 GHz (vCenter reported: **108.54 GHz**)	**11.7 GHz** (11.7/3.40=3.4) ~ 3.4 Physical cores dedicated for an ESXi Host	3 Hosts x 108.8 GHz = 326.4 GHz (vCenter reported: **325.62 GHz)**	3 host x 11.7 GHz = 35.1 GHz 325.62 GHz – 35.1 GHz = **290.52 GHz**	Total 3 x 3.4 = 10.2 physical cores out of 96 physical cores dedicated for ESXi hosts operation at cluster level
Memory	768 GB	vCenter reported: **767.65 GB**	**13.86 GB**	3 Hosts x 768 GB = 2304 GB (vCenter reported: **2302.95 GB**)	3 x 13.86 GB = 41.58 GB 2302.95 – 41.48 GB = **2261.47 GB**	Total 41.58 GB is the memory overhead at cluster level

Similarly, some memory must be kept available for the hypervisor operations. Memory overhead calculation involves, a fixed system-wide overhead for the VMkernel and additional overhead for each virtual machine. Refer to the VMware documentation for more details. In the testing, every ESXi host had reserved around 13.86 GB memory for the host operation, such as VMkernel, and every Oracle database VM with 128 GB memory reservation had 1.32 GB additional memory overhead reserved as shown in Figure 10. The total memory reservation for the VM was 129.32 GB.

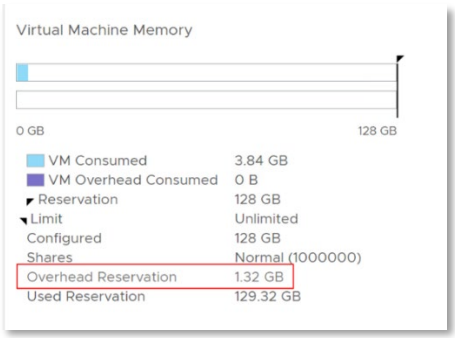

FIGURE 10. VM memory overhead

vSphere Cluster Service VM - overhead

In each ESXi cluster node, there was one virtual machine, called vCLS created by VMware vSphere. It is a vSphere Cluster Service VM, required to maintain the health of vSphere Cluster Services. The power state and resources of this VM are managed by vSphere Cluster Services. It has consumed 1 vCPU, 128 MB memory, and around 8 GB storage space.

At the time of sizing, these overheads should be considered and accommodated in the overall consolidation capacity requirement.

Workload description

Oracle performance tests were conducted using HammerDB, an open-source tool. For this Reference Configuration, HammerDB 3.3 was used to implement the online transaction processing (OLTP) workload. For the OLTP workload, HammerDB provides a real-world type scenario that consumes both CPU for the application logic and I/O. The HammerDB tool implements an OLTP-type workload with small I/O sizes of a random nature. The transaction results were normalized and used to compare test configurations. Other metrics collected during the testing came from the operating system and standard Oracle Automatic Workload Repository (AWR) statistics reports.

The OLTP test performed on a schema with 1000 warehouses was highly CPU intensive and moderately I/O intensive. The environment was tuned for maximum user transactions. After the database was tuned, the transaction rates were recorded at various Oracle connection counts. Because customer workloads vary in characteristics, the measurement was made with a focus on maximum transactions.

Analysis and recommendations

Two consolidation scenarios were tested with Oracle databases:

1. Running multiple single-instance Oracle databases (s1db, s2db, and s3db) in a VMware vSphere cluster node.
2. Running multiple three-node Oracle RAC database clusters (rac1db, rac2db, and rac3db) with RAC nodes distributed across three VMware vSphere cluster nodes.

Oracle RAC – Processing scalability

Figure 11 shows the relative throughput (measured in transactions per minute) of an Oracle RAC cluster with 8 vCPUs and 16 vCPUs configurations. When adding more vCPUs, throughput increased linearly. With 16 vCPUs, throughput increased up to 2 times the number of transactions as with 8 vCPUs. In this test, CPU reservation and limit features were used for restricting the CPU scheduling within allocated vCPU counts.

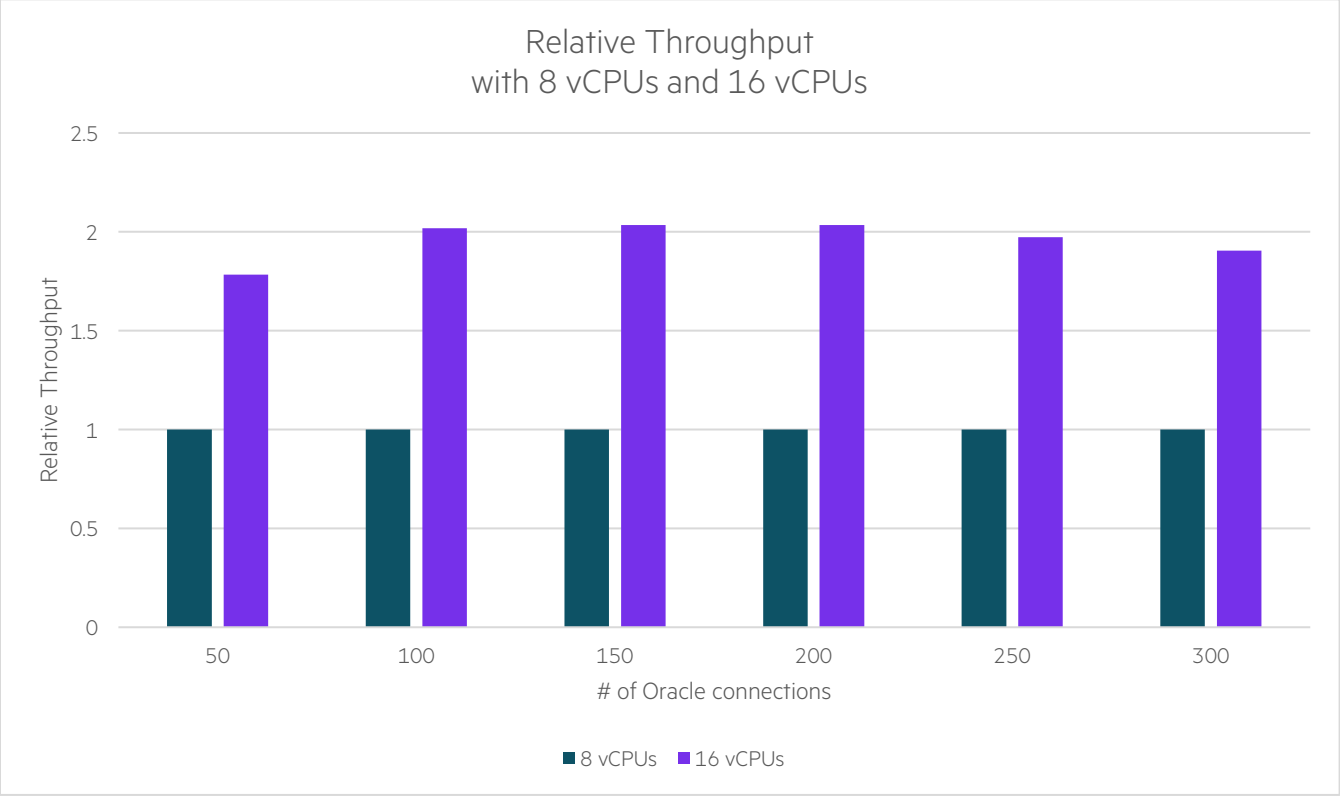

FIGURE 11. Oracle RAC database – 8 vCPUs vs 16 vCPUs in each Oracle RAC node

Oracle RAC – Ability to utilize the idle resources

One can specify a CPU reservation if there is a need to guarantee that the minimum required amount of CPUs are always available for the Oracle database VM. In this test, two throughput measurements were made with 20 vCPUs reserved in each VM (total 3 RAC Clusters x 3 VMs = 9 VMs):

- Run workload simultaneously on all three RAC clusters (rac1db, rac2db, and rac3db) and collected transaction throughput data.
- Run workload only on one RAC cluster (rac1db) while the other two RAC clusters were loaded with 20 oracle user connections.

As shown in Figure 12, transaction throughput for the RAC cluster (rac1db) increased up to 1.84 times when the ESXi cluster was partially loaded. This is useful in a consolidation scenario where two separate Oracle databases have their highest resource usage at different times of the day.

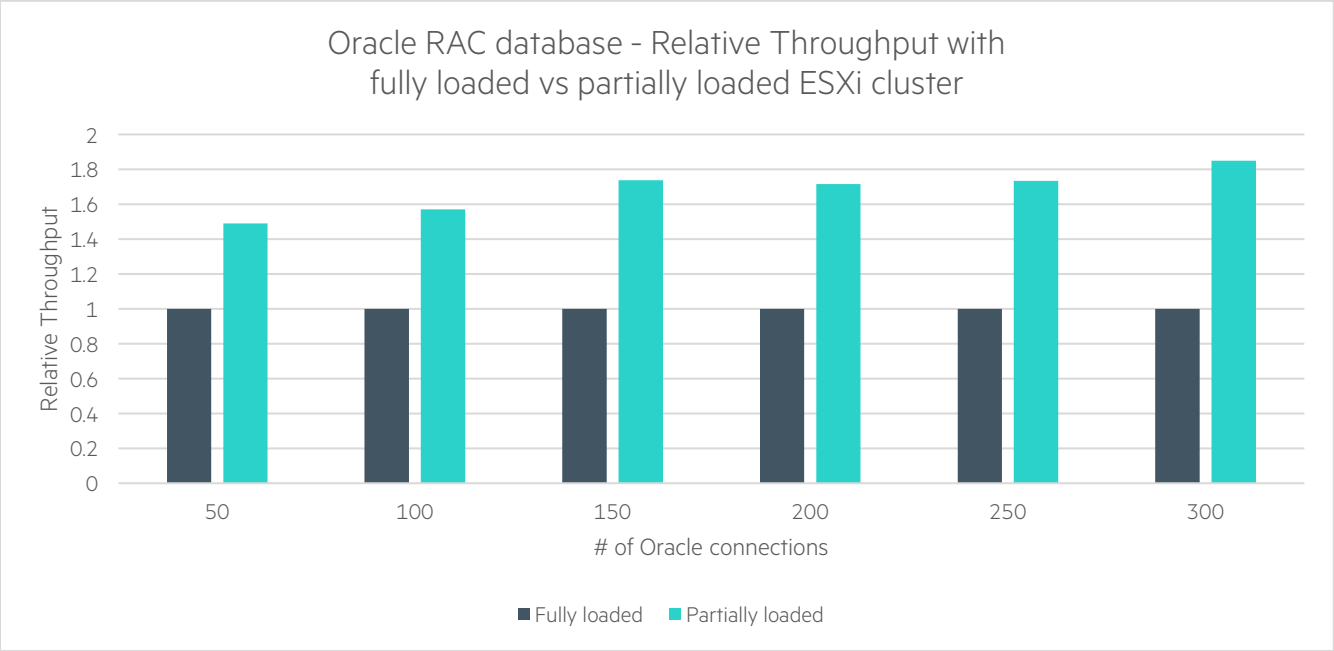

FIGURE 12. Oracle RAC database – Relative throughput with fully loaded vs partially loaded ESXi cluster

Oracle RAC - Throughput performance of three RAC clusters

Figure 13 shows the relative throughput performance of three RAC clusters, where workloads were run simultaneously on all three RAC clusters. All results are normalized to a single RAC cluster (rac1db) data. All the VMs in every RAC cluster had an equal amount of CPU and memory resources allocated. Test results show that resources were fairly allocated to all the VMs as throughput is almost the same on all the RAC clusters and only up to 6% throughput performance differences were seen.

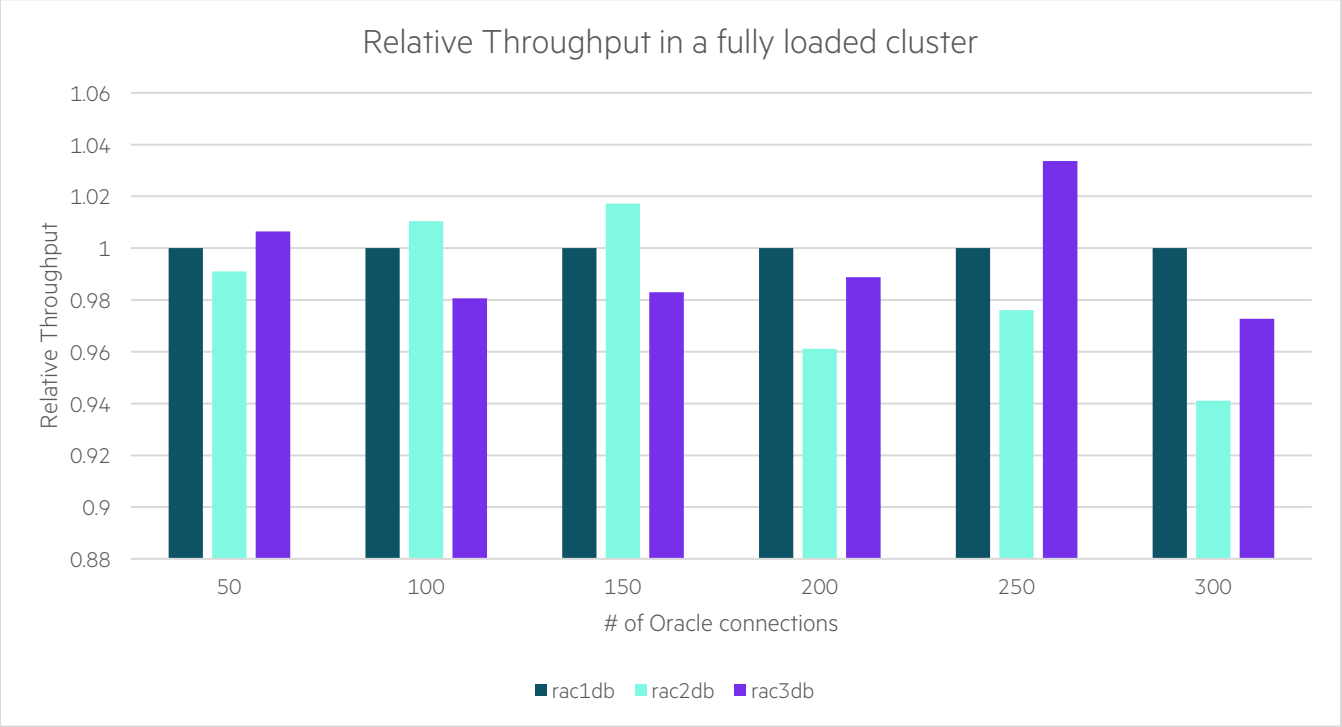

FIGURE 13. Comparison of relative throughput of three Oracle RAC cluster

Oracle single instance database – Bare metal versus virtual machine

Figure 14 shows the relative throughput performance of a single instance Oracle database running on a 2-socket bare metal server, having a total of 32 physical cores and a virtual machine with 64 vCPUs (hyper-threaded cores), which is equivalent to 32 physical cores. There were no other VMs configured on the same hypervisor. The number of warehouses, and PGA and SGA sizes were kept the same for both configurations. All results are normalized to the single instance database running on the virtual machine. Transaction throughput for the bare metal was up to 11% better than the virtual machine. This is as expected, as 11.7 GHz out of 108.54 GHz was reserved for the hypervisor operations. Therefore, the throughput of the virtual machine was lower than the bare metal.

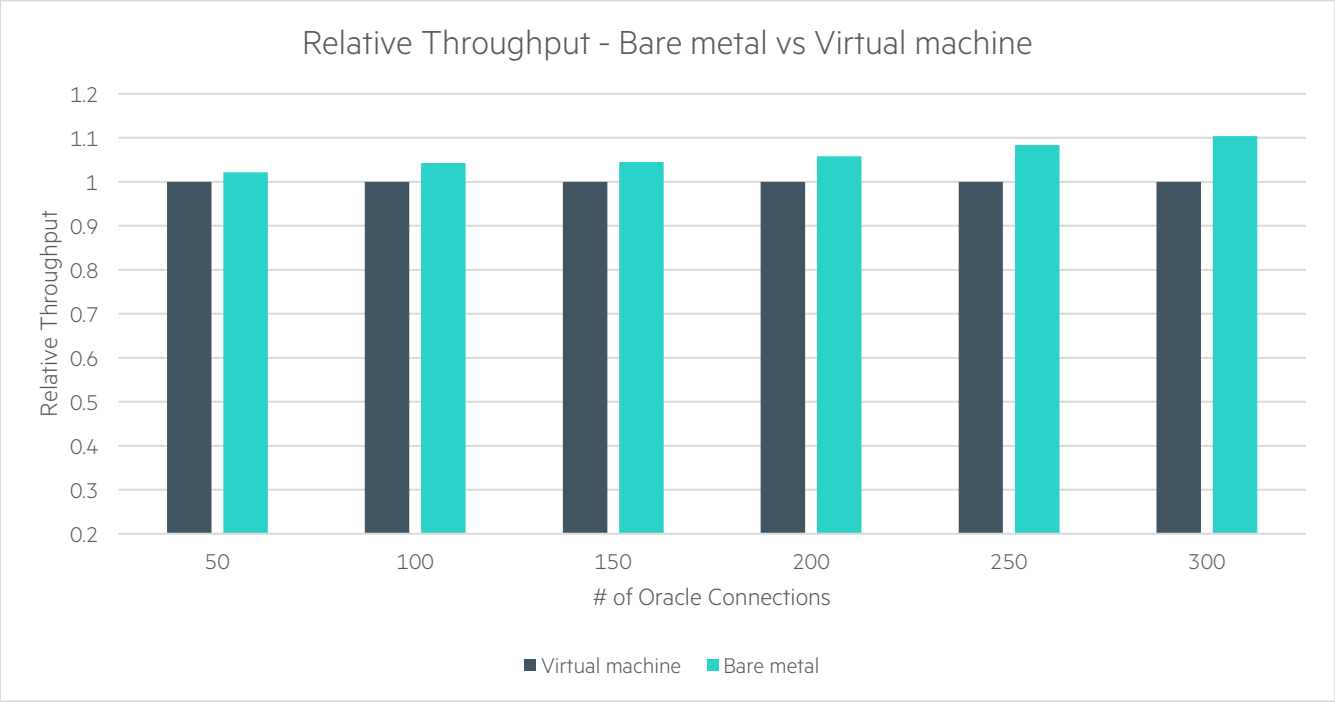

FIGURE 14. Comparison of relative throughput between Oracle single instance running on bare metal and virtual machine

Oracle single instance database – One virtual machine versus two virtual machine

Figure 15 shows the relative throughput performance of a single instance Oracle database running on a virtual machine with 64 vCPUs and two virtual machines with 32 vCPUs each. The number of warehouses, and PGA and SGA sizes were kept the same for both configurations. All results are normalized to the single instance database running on the virtual machine with 64 vCPUs. As shown in Figure 15, the sum of throughput from the two virtual machines configuration was almost the same as the single virtual machine configuration with small variations.

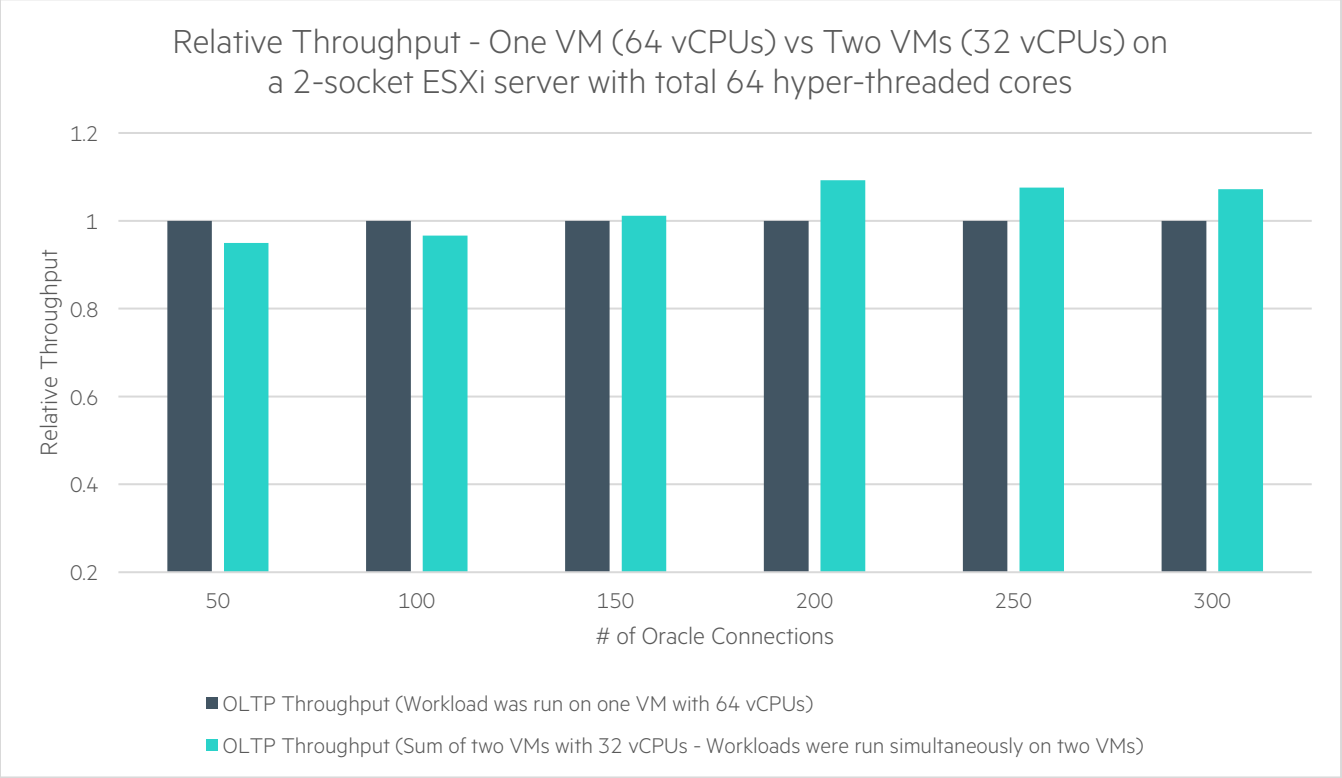

FIGURE 15. Comparison of relative throughput between one VM with 64 vCPUs and two VMs with 32 vCPU each

SUMMARY

This Reference Configuration described how to perform capacity planning, consolidate and deploy Oracle RAC 19c databases and single instance databases on HPE Synergy with VMware vSphere and HPE Primera A650 all-flash storage. With the help of fluid resource pools and software-defined intelligence from HPE Synergy Composer 2, IT administrators can rapidly compose any configuration required, reducing deployment time from days to hours. More specifically, this Reference Configuration shows the following benefits of utilizing HPE Synergy for Oracle database solutions.

- HPE Synergy Composer 2 with embedded HPE OneView seamlessly manages the entire environment, including the configuration of network resources required for Oracle databases on the compute modules, and the creation and management of HPE Primera virtual volumes for the Oracle database nodes.
- HPE OneView for VMware vCenter enables administrators to have an end-to-end resource view and provides the ability to create or grow VMware clusters from HPE OneView profiles.
- HPE Primera fully supports VMware vSphere vVols feature and supports Oracle databases with large datastores.
- Testing shows that VMware vSphere can best utilize the available resources and distribute resources to the most demanding Oracle databases in a consolidated environment.
- Testing shows that though there is a resource overhead due to virtualization, the benefits outweigh the overhead. For example, the following are a few advantages of virtualization, but not limited to:

  - Consolidate under-utilized database servers to easily manageable virtual machines environment and save database licensing cost.
  - Add or modify compute, memory, and storage resources in a virtual machine in minutes using VMware vCenter.
  - Clone virtual machines and create Test and Dev environment in minutes.
  - Easily move virtual machines between VMware vSphere cluster nodes for maintenance activities.
  - Perform online migrations of virtual machines.
- If the VMware vSphere cluster runs out of processing capacity for Oracle RAC clusters, the environment can easily be scaled out with more HPE Synergy Compute Modules added to the VMware vSphere clusters and add additional virtual machines to the existing RAC clusters. This is much easier to configure in an HPE Synergy environment due to the HPE OneView server profiles.

IMPLEMENTING A PROOF-OF-CONCEPT

As a matter of best practice for all deployments, Hewlett Packard Enterprise recommends implementing a proof-of-concept using a test environment that matches as closely as possible the planned production environment. In this way, appropriate performance and scalability characterizations can be obtained. For help with a proof-of-concept, contact a Hewlett Packard Enterprise Services representative (hpe.com/us/en/services/consulting.html) or your Hewlett Packard Enterprise partner.

APPENDIX A: BILL OF MATERIALS

NOTE

Part numbers are at the time of publication/testing and subject to change. The bill of materials does not include complete support options or other rack and power requirements. If you have questions regarding ordering, please consult with your Hewlett Packard Enterprise Reseller or Hewlett Packard Enterprise Sales Representative for more details. hpe.com/us/en/services/consulting.html.

TABLE A1. Bill of materials

Part number	Quantity	Description
		Rack and Network Infrastructure
P9K10A	1	HPE 42U 600mmx1200mm G2 Kitted Advanced Shock Rack with Side Panels and Baying
P9K10A 001	1	HPE Factory Express Base Racking Service
P9S21A	2	HPE G2 Metered/Switched 3Ph 14.4kVA/CS8365C 40A/208V Outlets (12) C13 (12) C19/Vertical NA/JP PDU
H6J85A	1	HPE Rack Hardware Kit
R9F64A	2	Aruba 8325-48Y8C 48p 25G SFP/+/28 8p 100G QSFP+/28 Front-to-Back 6 Fans and 2 PSU Bundle
R9F74A	2	Aruba 100G QSFP28 to QSFP28 3m Direct Attach Copper Cable
R9F59A	2	Aruba X474 4-post Rack Kit
R9F77A	2	Aruba 100G QSFP28 to QSFP28 1m Direct Attach Copper Cable
845406-B21	2	HPE 100Gb QSFP28 to QSFP28 3m Direct Attach Copper Cable
Q0U54B	2	HPE SN6600B 32Gb 48/24 Fibre Channel Switch
P9H32A	48	HPE B-series 32Gb SFP28 Short Wave 1-pack Transceiver
		Virtualization hosts
P06011-B21	1	HPE Synergy 12000 Configure-to-order Frame with 10x Fans
871942-B21	4	HPE Synergy 480 Gen10 Configure-to-order Premium Compute Module
P23592-L21	4	Intel Xeon-Gold 6246R (3.4GHz/16-core/205W) FIO Processor Kit for HPE Synergy 480 Gen10
P23592-B21	4	Intel Xeon-Gold 6246R (3.4GHz/16-core/205W) Processor Kit for HPE Synergy 480 Gen10
P28217-B21	48	HPE Synergy 64GB (1x64GB) Dual Rank x4 DDR4-2933 CAS-21-21-21 Registered Smart Memory Kit

Part number	Quantity	Description
P19905-B21	8	HPE 1.92TB SAS 12G Read Intensive SFF SC PM1643a SSD
P01367-B21	4	HPE 96W Smart Storage Lithium-ion Battery with 260mm Cable Kit
871573-B21	4	HPE Smart Array P416ie-m SR Gen10 SAS Cable Kit
777456-B21	4	HPE Synergy 5830C 32Gb Fibre Channel Host Bus Adapter
804428-B21	4	HPE Smart Array P416ie-m SR Gen10 (8 Int 8 Ext Lanes/2GB Cache) 12G SAS Mezzanine Controller
P02054-B21	4	HPE Synergy 6820C 25/50Gb Converged Network Adapter
867796-B21	2	HPE Virtual Connect SE 100Gb F32 Module for Synergy
876259-B21	2	HPE Virtual Connect SE 32Gb Fibre Channel Module for Synergy
876852-B21	2	HPE Synergy 4-port Frame Link Module
798096-B21	1	HPE 6x 2650W Performance Hot Plug Titanium Plus FIO Power Supply Kit
804938-B21	1	HPE Synergy Frame Rack Rail Kit
804943-B21	1	HPE Synergy Frame 4x Lift Handles
872957-B21	2	HPE Synergy Composer2 Management Appliance
		Storage
N9Z47A	1	HPE Primera 600 4-way Storage Base
N9Z61A	1	HPE Primera A650 4-node Controller
R0P95A	24	HPE Primera 600 1.92TB SAS SFF (2.5in) SSD
N9Z39A	8	HPE Primera 600 32Gb 4-port Fibre Channel Host Bus Adapter
716195-B21	8	HPE External 1.0m (3ft) Mini-SAS HD 4x to Mini-SAS HD 4x Cable
N9Z50A	2	HPE Primera 600 2U 24-disk SFF Drive Enclosure
R0P95A	24	HPE Primera 600 1.92TB SAS SFF (2.5in) SSD
QK734A	16	HPE Premier Flex LC/LC Multi-mode OM4 2 fiber 5m Cable
QK735A	48	HPE Premier Flex LC/LC Multi-mode OM4 2 fiber 15m Cable

APPENDIX B: RHEL KERNEL SETTINGS

The following RHEL kernel parameters were set in the `/etc/sysctl.conf` file for Oracle RAC Cluster nodes.

```
kernel.sem = 250 32000 100 128
kernel.shmall = 23488102
kernel.shmmax = 96207267430
fs.file-max = 6815744
kernel.shmmni = 4096
fs.aio-max-nr = 1048576
net.ipv4.ip_local_port_range = 9000 65500
net.core.rmem_default = 1048576
net.core.wmem_default = 1048576
net.core.rmem_max=26214400
net.core.wmem_max=26214400
net.ipv4.tcp_rmem = 1048576 1048576 4194304
net.ipv4.tcp_wmem = 1048576 1048576 1048576
vm.nr_hugepages = 45079
vm.hugetlb_shm_group = 54321
kernel.numa_balancing = 0
```

APPENDIX C: ORACLE USER ACCOUNT LIMITS

The following settings were included in the file `/etc/security/limits.d/oracle-limits.conf`.

```
oracle   soft   nofile    1024
oracle   hard   nofile    65536
oracle   soft   nproc    16384
oracle   hard   nproc    16384
oracle   soft   stack    10240
oracle   hard   stack    32768
oracle   hard   memlock    118575393
oracle   soft   memlock    118575393
```

APPENDIX D: UDEV RULES FOR VVOLS

Udev rule file called "`/etc/udev/rules.d/99-oracle-asmdevices.rules`" was created to set parameters for the VMware vVols presented from HPE Primera. ASM search string used for the installation was "`/dev/ora*`".

```
# cat /etc/udev/rules.d/99-oracle-asmdevices.rules
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c29815f9b95974f80b91763d8eac", SYMLINK+="oradata1", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c2952377a0c1bb0754ab1f3a2422", SYMLINK+="oradata2", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c290dd302f24a7ba43c00919afdf", SYMLINK+="oradata3", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c29b367e8db43b4c830fadb5e380", SYMLINK+="oradata4", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c290868781d0f5e795dd18ce52a6", SYMLINK+="oradata5", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c29965b9b3ef8b117ea6282e4e07", SYMLINK+="oradata6", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c29ff915a68c3a02399e776f2468", SYMLINK+="oradata7", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c29532b59e944312cff046bafd6c", SYMLINK+="oradata8", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c2949aa1eeb8d6d5c1f4cce7982a", SYMLINK+="oraredo1", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c29d6616e4dcc8cf301038d2a46c", SYMLINK+="oraredo2", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c299b1c424aa640a28ad6dd62280", SYMLINK+="oraredo3", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c2993588fecbb48dbc940eb5e675", SYMLINK+="oraredo4", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c2927c9f1fbfa9246e82a815e64b", SYMLINK+="oraredo5", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c294fbe36ec5717f1f0812ad3ed9", SYMLINK+="oraredo6", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c294a4b24dec4015c78b03555f25", SYMLINK+="oraredo7", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c29a3841fd224da61d930531bbe9", SYMLINK+="oraredo8", OWNER="oracle", GROUP="oinstall", MODE="0660"
KERNEL=="sd?", SUBSYSTEM=="block", PROGRAM=="/usr/lib/udev/scsi_id -g -u -d /dev/$name",
RESULT=="36000c29a3841cd324da61d93053d3ed9", SYMLINK+="oramgmt", OWNER="oracle", GROUP="oinstall", MODE="0660"
```

RESOURCES AND ADDITIONAL LINKS

HPE Reference Architectures, hpe.com/info/ra

HPE Servers, hpe.com/servers

HPE Storage, hpe.com/storage

HPE Networking, hpe.com/networking

HPE Advisory and Professional Services, hpe.com/us/en/services/consulting.html

HPE Primera: VMware ESXi Implementation Guide, https://support.hpe.com/hpesc/public/docDisplay?docId=sd00001341en_us

HPE OneView for VMware vCenter Grow Cluster White Paper, https://psnow.ext.hpe.com/doc/a00099279en_us

VMware Hybrid Cloud Best Practices Guide for Oracle Workloads, https://www.vmware.com/content/dam/digitalmarketing/vmware/en/pdf/solutions/vmware-oracle-databases-on-vmware-best-practices-guide.pdf

HPE Serviceguard for Linux with VMware Virtual Machines, https://psnow.ext.hpe.com/doc/4AA4-2016ENW

HPE Serviceguard for Linux Quick Specs, https://support.hpe.com/hpesc/public/docDisplay?docLocale=en_US&docId=c04154488

Managing HPE Serviceguard for Linux A.12.70.00, https://www.hpe.com/psnow/doc/c04154488.pdf?jumpid=in_lit-psnow-getpdf

HPE Serviceguard Toolkit for Oracle on Linux User Guide, https://support.hpe.com/hpsc/doc/public/display?docLocale=en_US&docId=emr_na-a00052274en_us&withFrame

HPE Serviceguard Toolkit for Oracle Data Guard User Guide, https://support.hpe.com/hpsc/doc/public/display?docLocale=en_US&docId=emr_na-a00052284en_us&withFrame

HPE Synergy 480 Gen10 Compute Module QuickSpecs, https://www.hpe.com/psnow/doc/a00008520enw?jumpid=in_lit-psnow-red

HPE Synergy Composer QuickSpecs, https://www.hpe.com/psnow/doc/c04815139.pdf?jumpid=in_lit-psnow-getpdf

HPE InfoSight for HPE Storage User Guide, https://support.hpe.com/hpesc/public/docDisplay?docId=a00094611en_us

VMware vVols With HPE Primera Storage, https://support.hpe.com/hpesc/public/docDisplay?docId=a00101451en_us

Virtual Machine vCPU and vNUMA Rightsizing – Guidelines, https://blogs.vmware.com/performance/2017/03/virtual-machine-vcpu-and-vnuma-rightsizing-rules-of-thumb.html

Virtual Machine Compute Optimizer, https://flings.vmware.com/virtual-machine-compute-optimizer

To help us improve our documents, please provide feedback at hpe.com/contact/feedback.

Learn more at

hpe.com/synergy

a50004657enw, June 2024

Hewlett Packard Enterprise