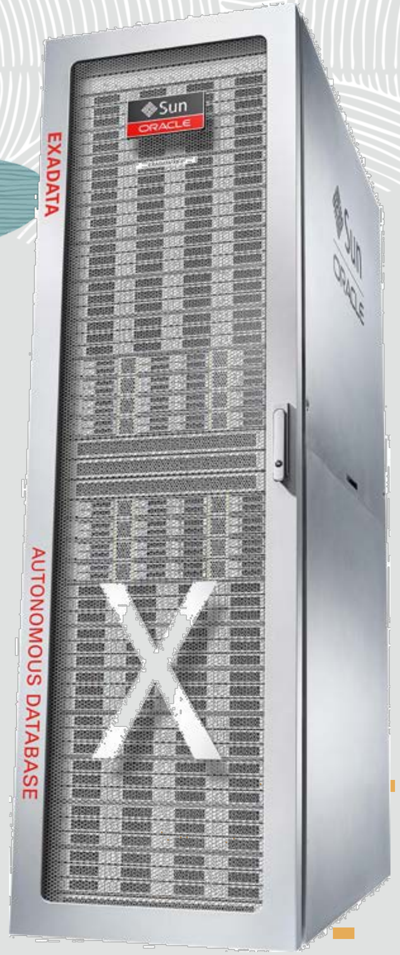
ORACLE

Oracle Exadata Database Machine

Oracle Exadata and OVM Best Practices

November 2019



Topics Covered

- Use Cases
- Exadata OVM Software Requirements
- Exadata Isolation Considerations
- Exadata OVM Sizing and Prerequisites
- Exadata OVM Deployment Overview
- Exadata OVM Administration and Operational Life Cycle
- Migration, HA, Backup/Restore, Upgrading/Patching
- Monitoring, Resource Management

Exadata Virtual Machines

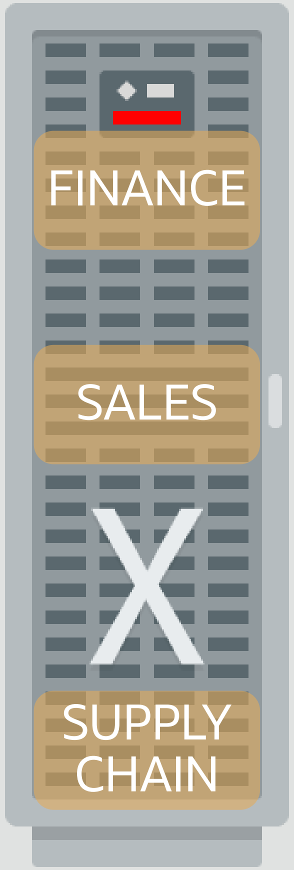
High-Performance Virtualized Database Platform

ORACLE VM

No Additional Cost

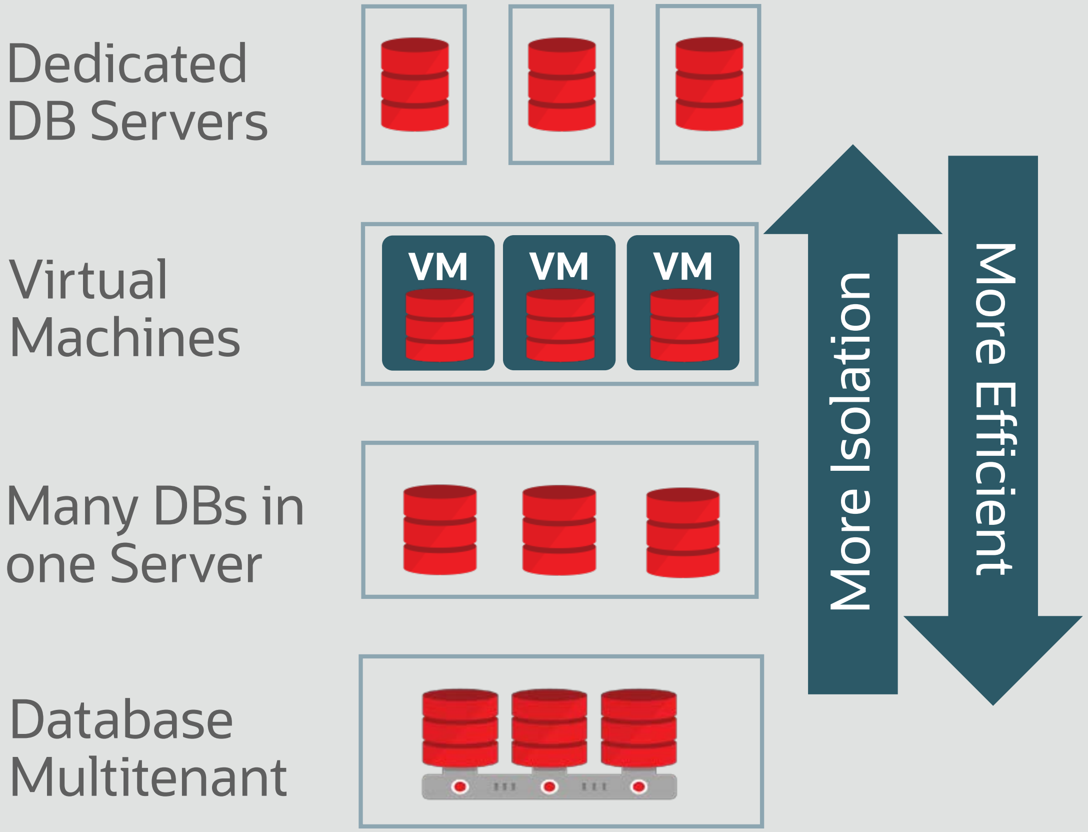
X8-2, X7-2, X6-2, X5-2, X4-2, X3-2, X2-2

DB 11.2 and higher



- XEN hypervisor
- VMs provide CPU, memory, OS, and sysadmin isolation for consolidated workloads
  - Hosting, cloud, cross department consolidation, test/dev,
    non-database or third party applications
- Exadata VMs deliver near raw hardware performance
  - I/Os go directly to high-speed InfiniBand bypassing hypervisor
- Combine with Exadata network and I/O prioritization to achieve unique full stack isolation
- Trusted Partitions allow licensing by virtual machine

Exadata Consolidation Options



Dedicated DB Servers

Virtual Machines

Many DBs in one Server

Database Multitenant

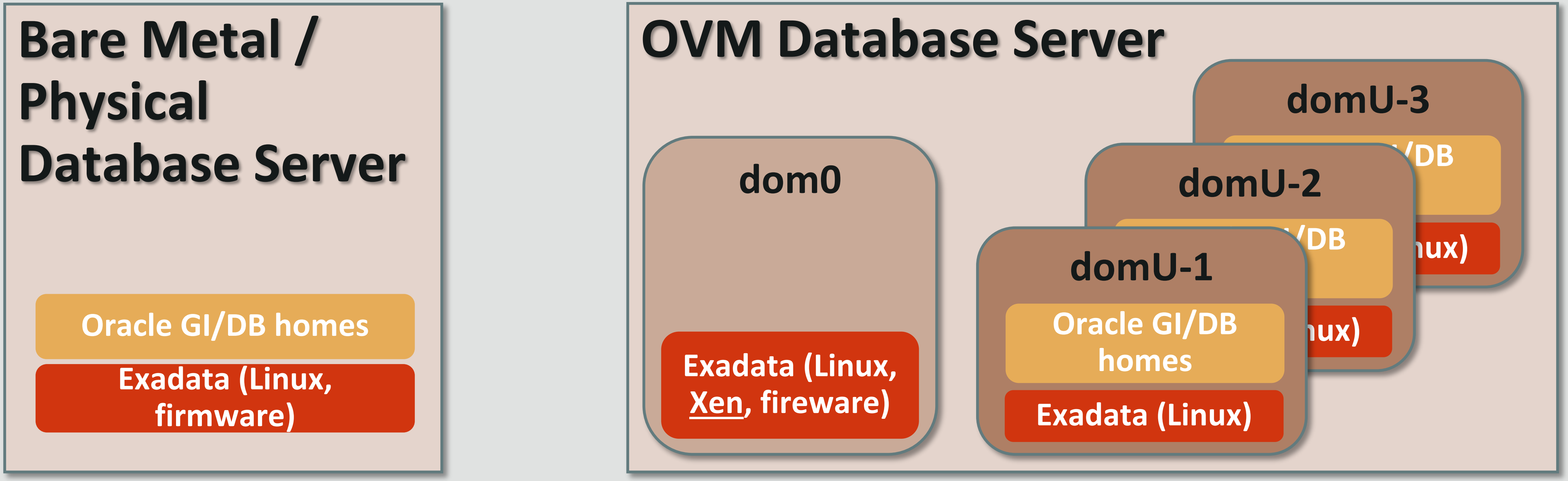
More Isolation

More Efficient

- VMs have good Isolation but poor efficiency and high management
  - VMs have separate OS, memory, CPUs, and patching
  - Isolation without need to trust DBA, System Admin
- Database consolidation in a single OS is highly efficient but less isolated
  - DB Resource manager isolation adds no overhead
  - Resources can be shared much more dynamically
  - But, must trust admins to configure systems correctly
- Best strategy is to combine VMs with database native consolidation
  - Multiple trusted DBs or Pluggable DBs in a VM
  - Few VMs per server to limit overhead of fragmenting CPUs/memory/patching etc.

Software Architecture Comparison

Database Server: Bare Metal / Physical versus OVM



Bare Metal / Physical Database Server

Oracle GI/DB homes

Exadata (Linux, firmware)

OVM Database Server

dom0

Exadata (Linux, Xen, fireware)

domU-3

/DB

ux)

domU-2

/DB

ux)

domU-1

Oracle GI/DB homes

Exadata (Linux)

No change to **Storage Grid**, **Networking**, or **Other**

Differences Between Physical and OVM

Details expanded throughout remaining slides

Topic	How OVM differs from Physical
Hardware support	2-socket only
Cluster config	System has one or more VM clusters, each with own GI/RAC/DB install
Exadata storage config	Separate griddisks/DATA/RECO for each VM cluster; By default no DBFS disk group
Dbnode disk config	VM filesystem sizes are small; GI/DB separate filesystems
Software updates	Dbnodes require separate dom0 (Linux+firmware) and domU (Linux) patchmgr updates
Exachk	Run once for dom0/cells/ibswitches, run once for each VM cluster
Enterprise Manager	EM + Exadata plugin + Virtualization Infrastructure plugin

Exadata VM Usage

- Primary focused on consolidation and isolation
- Can only run certified Oracle Linux versions
  - Windows, RedHat, and other guest operating systems are not supported
- Can virtualize other lightweight products
  - E.g. Lightweight apps, management tools, ETL tools, security tools, etc.
- Not recommended for heavyweight applications
  - E.g. E-business Suite or SAP application tier
  - Instead use Private Cloud Appliance

Exadata OVM Requirements

- Hardware
  - 2-socket database servers supported (X2-2 through X8-2)
- Software
  - Recommend latest Exadata 18.x or 19.x software
    - Supplied software (update with patchmgr - see MOS 888828.1)
      - domU and dom0 run same UEK kernel as physical
      - domU runs same Oracle Linux (OL) as physical
      - dom0 runs Oracle VM Server (OVS) 3.x
  - Grid Infrastructure / Database
    - Recommend 19c with latest quarterly update
    - Supported 19c, 18c, 12.2.0.1, 12.1.0.2, or 11.2.0.4

Exadata Security Isolation Recommendations

- Each VM RAC cluster has its own Exadata grid disks and ASM Disk Groups
  - Setting Up Oracle ASM-Scoped Security on Oracle Exadata Storage Servers
- 802.1Q VLAN Tagging for Client and Management Ethernet Networks
  - Dbnodes configured w/ OEDA during deployment (requires pre-deployment switch config)
  - Or configure manually post-deployment
    - Client network - MOS 2018550.1 Management network - MOS 2090345.1
- InfiniBand Partitioning with PKEYs for Exadata Private Network
  - OS and InfiniBand switches configured w/ OEDA during deployment
- Storage Server administration isolation through ExaCLI

Exadata OVM Sizing Recommendations

Use Reference Architecture Sizing Tool to determine CPUs, memory, disk space needed by each database

- Sizing evaluation should be done prior to deployment since OEDA will deploy your desired VM configuration in an automated and simple manner.
- Changes can be made post deployment, but requires many more steps
- Sizing approach does not really change except for accommodating DOM0, and additional system resources per VM
- Sizing tool currently does not size virtual systems
- Consider dom0 memory and CPU usage in sizing

Memory Sizing Recommendations

- Can not over-provision physical memory
  - Sum of all VMs + dom0 memory used cannot exceed physical memory
  - Sum of all VM memory <= 720 GB
    - X8, X7, X6 database servers support maximum 768 GB physical memory when deployed virtualized (non-virtualized systems support higher)
- dom0 memory sizing
  - 8 GB (do not change unless directed by Oracle)
- VM memory sizing
  - Initially set during OEDA configuration
  - Minimum 16 GB per VM (to support OS, GI/ASM, starter DB, few connections)
  - Maximum 720 GB for single VM
  - Memory size on Exadata can not be changed online (VM restart required)

CPU Sizing Recommendations

- CPU over-provisioning is possible
  - But workload performance conflicts can arise if all VMs become fully active
- Dom0 CPU sizing
  - Allocated 2 cores (4 vCPUs - do not change unless directed by Oracle)
- VM CPU sizing
  - Minimum per VM is 2 cores (4 vCPUs)
    - 1 vCPU == 1 hyper-thread; 1 core == 2 hyper-threads == 2 vCPUs
  - Maximum per VM per DB Server is number of cores minus 2 for dom0
    - E.g.: for X8-2, maximum per VM per DB Server is 46 cores (48 total minus 2 for dom0)
  - vCPU initially set during OEDA configuration
  - vCPU can be changed dynamically (online while VM remains up)

Local Disk Sizing Recommendations

- Total local disk space available for VMs
  - X8 - 3.2TB; X7,X6,X5 - 1.6TB, 3.7TB with disk drive expansion kit; X4 - 1.6TB
- 190GB used per VM at deployment, extendable post-deployment
- Actual allocated space for domU disk images <u>initially</u> much lower due to sparseness and shareable reflinks, but will grow with domU use as shared space diverges and becomes less sparse
    - Over-provisioning disk may cause unpredictable out-of-space errors inside VMs if dom0 space is exhausted
    - Restoring VM backup will reduce (may eliminate) space savings (i.e. relying on over-provisioning may prevent full VM restore)
    - Long lived / prod VMs should budget for full space allocation (assume no benefit from sparseness and shareable reflinks)
    - Short lived test/dev VMs can assume 100 GB allocation
- DomU local space can be extended after initial deployment by adding local disk images
  - Additionally, domU space can be extended with shared storage (e.g. ACFS, DBFS, external NFS) for user / app files
  - Avoid shared storage for Oracle/Linux binaries/config files. Access/network issues may cause system crash or hang.

Exadata Storage Recommendation

- DATA/RECO size for initial VM clusters should consider future VM additions
  - Using all space initially will require shrinking existing DATA/RECO before adding new
- Spread DATA/RECO for each VM cluster across all disks on all cells
  - By default no DBFS disk group
- Enable ASM-Scoped Security to limit grid disk access

VM Cluster	Cluster nodes	Grid disks (DATA/RECO for all clusters on all disks in all cells)
clu1	db01vm01 db02vm01	DATAC1_CD_{00..11}_cel01 RECOC1_CD_{00..11}_cel01 DATAC1_CD_{00..11}_cel02 RECOC1_CD_{00..11}_cel02 DATAC1_CD_{00..11}_cel03 RECOC1_CD_{00..11}_cel03
clu2	db01vm02 db02vm02	DATAC2_CD_{00..11}_cel01 RECOC2_CD_{00..11}_cel01 DATAC2_CD_{00..11}_cel02 RECOC2_CD_{00..11}_cel02 DATAC2_CD_{00..11}_cel03 RECOC2_CD_{00..11}_cel03

Deployment Specifications and Limits

	Hardware	X3-2	X4-2	X5-2	X6-2	X7-2	X8-2
VMs	Max VMs per database server	8					
Memory	Physical per database server (default/max)	256 GB 512 GB	256 GB 512 GB	256 GB 768 GB	256 GB 768 GB[2]	384 GB 768 GB[2]	
	Min per VM	16 GB					
	Max per VM	464 GB		720 GB			
	Default setting	Initially set during OEDA configuration					
CPU	Cores/vCPU[1] per database server	16	24	36	44	48	
	Min cores/vCPU per VM	2 core (4 vCPUs)					
	Max cores/vCPU per VM	Cores minus 2 (dom0 assigned 2 cores/4vCPUs)					
	Default setting	Initially set during OEDA configuration					
Disk	Total usable disk per dbserver for all VMs	700 GB	1.6 TB	1.6 TB (3.7 TB w/ DB Storage Expansion Kit)			3.2 TB
	Used disk per VM at deployment	190 GB Actual allocated space for domU disk images initially much lower due to sparseness and shareable reflinks, but will grow with domU use as shared space diverges and becomes less sparse, hence budget for these values when sizing.					

Footnotes: 1) 1 core = 1 OCPU = 2 hyper-threads = 2 vCPUs; 2) Systems deployed non-virtual support higher physical memory

Deployment Overview

OEDA is the only tool that should be used to create VMs on Exadata

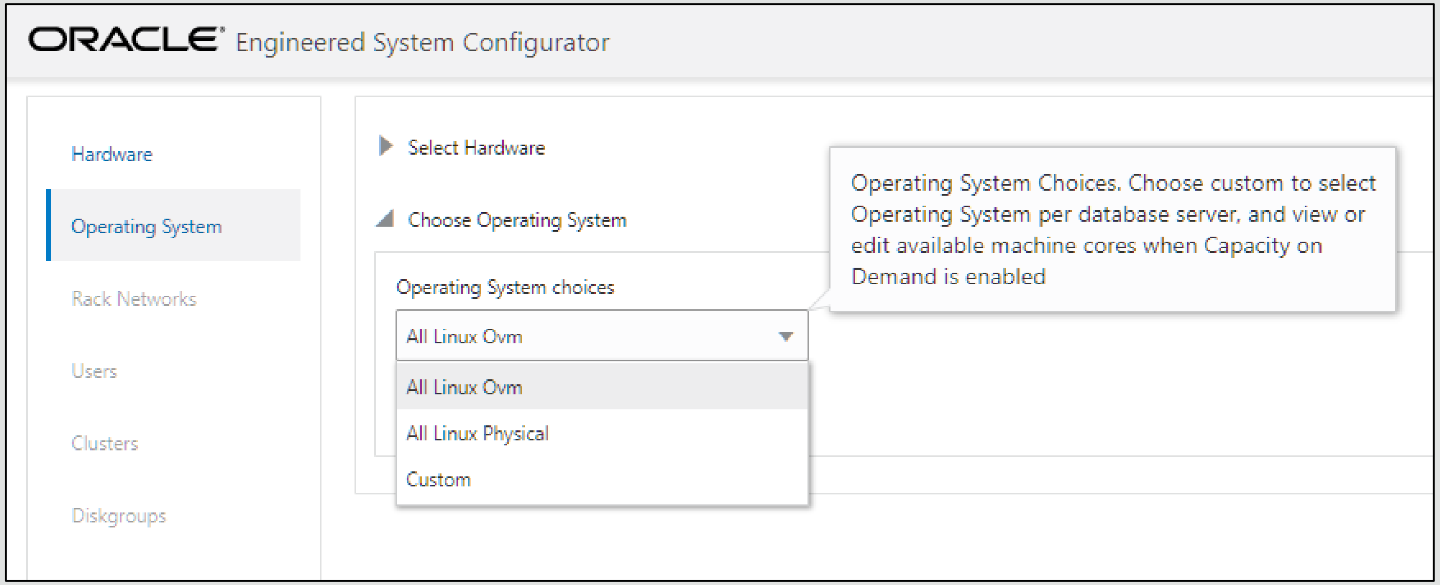
1. Create configuration with OEDA Configuration Tool
2. Prepare customer environment for OEDA deployment
   Configure DNS, configure switches for VLANs (if necessary)
3. Prepare Exadata system for OEDA deployment
   switch_to_ovm.sh; reclaimdisks.sh; applyElasticConfig.sh
4. Deploy system with OEDA Deployment Tool

Note: OS VLAN config can be done by OEDA or post deployment (MOS 2018550.1)

OEDA Configuration Tool

Configuring OVM

- Screen to decide OVM or Physical
  - All OVM
  - All Physical
  - Some OVM, some physical



ORACLE Engineered System Configurator

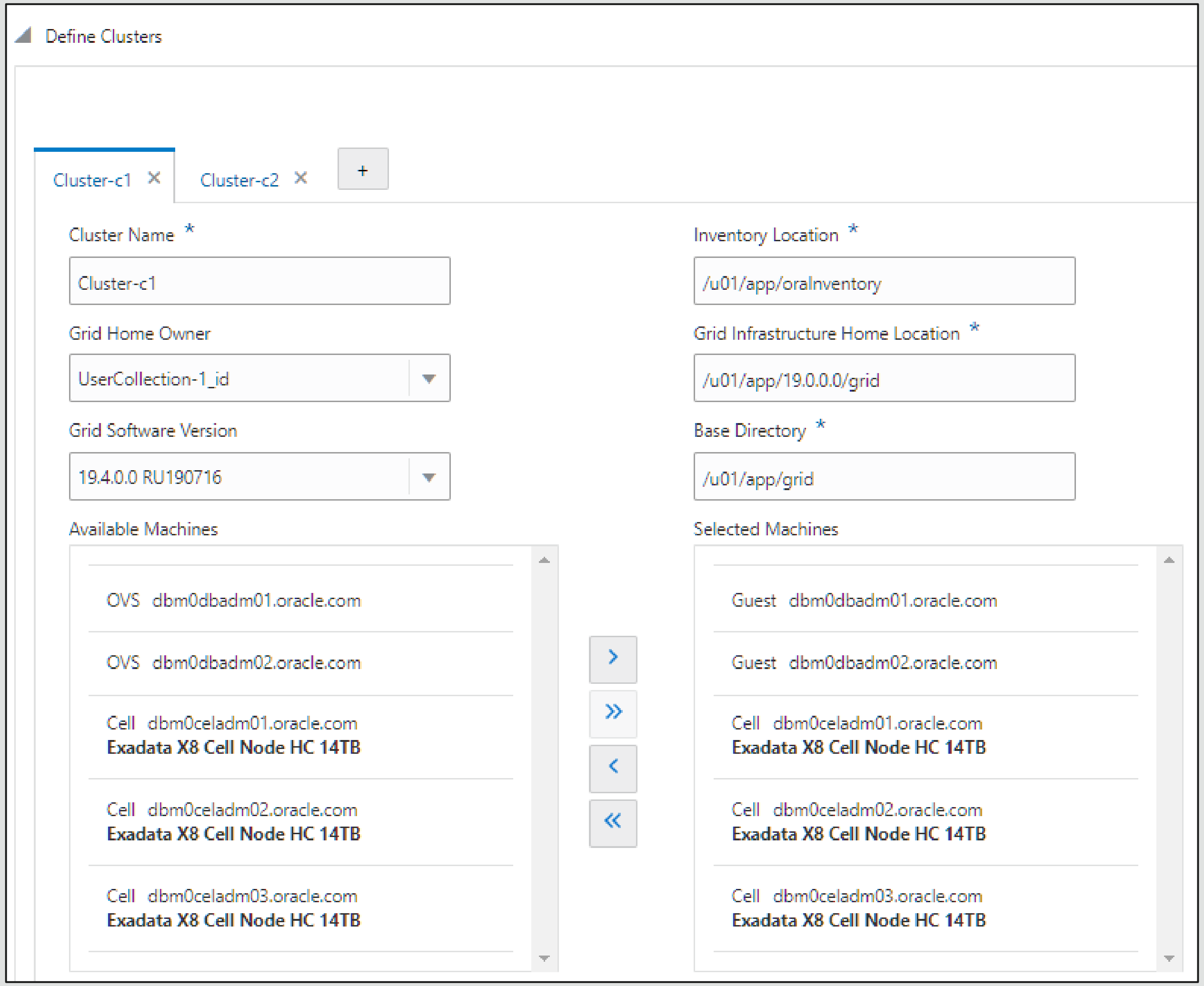
Operating System choices: All Linux Ovm

Operating System Choices. Choose custom to select Operating System per database server, and view or edit available machine cores when Capacity on Demand is enabled

OEDA Configuration Tool

Define Clusters

- Decide
  - Number of VM clusters to create
  - Dbnodes and Cells that will make up those VM clusters
    - Recommend using all cells
- What is a “VM cluster?”
  - 1 or more user domains on different database servers running Oracle GI/RAC, each accessing the same shared Exadata storage managed by ASM.



Define Clusters

Cluster Name * Cluster-c1

Grid Home Owner UserCollection-1_id

Grid Software Version 19.4.0.0 RU190716

Inventory Location * /u01/app/oraInventory

Grid Infrastructure Home Location * /u01/app/19.0.0.0/grid

Base Directory * /u01/app/grid

Available Machines

OVS dbm0dbadm01.oracle.com

OVS dbm0dbadm02.oracle.com

Cell dbm0celadm01.oracle.com
Exadata X8 Cell Node HC 14TB

Cell dbm0celadm02.oracle.com
Exadata X8 Cell Node HC 14TB

Cell dbm0celadm03.oracle.com
Exadata X8 Cell Node HC 14TB

Selected Machines

Guest dbm0dbadm01.oracle.com

Guest dbm0dbadm02.oracle.com

Cell dbm0celadm01.oracle.com
Exadata X8 Cell Node HC 14TB

Cell dbm0celadm02.oracle.com
Exadata X8 Cell Node HC 14TB

Cell dbm0celadm03.oracle.com
Exadata X8 Cell Node HC 14TB

OEDA Configuration Tool

Cluster Configuration

Each VM cluster has its own configuration

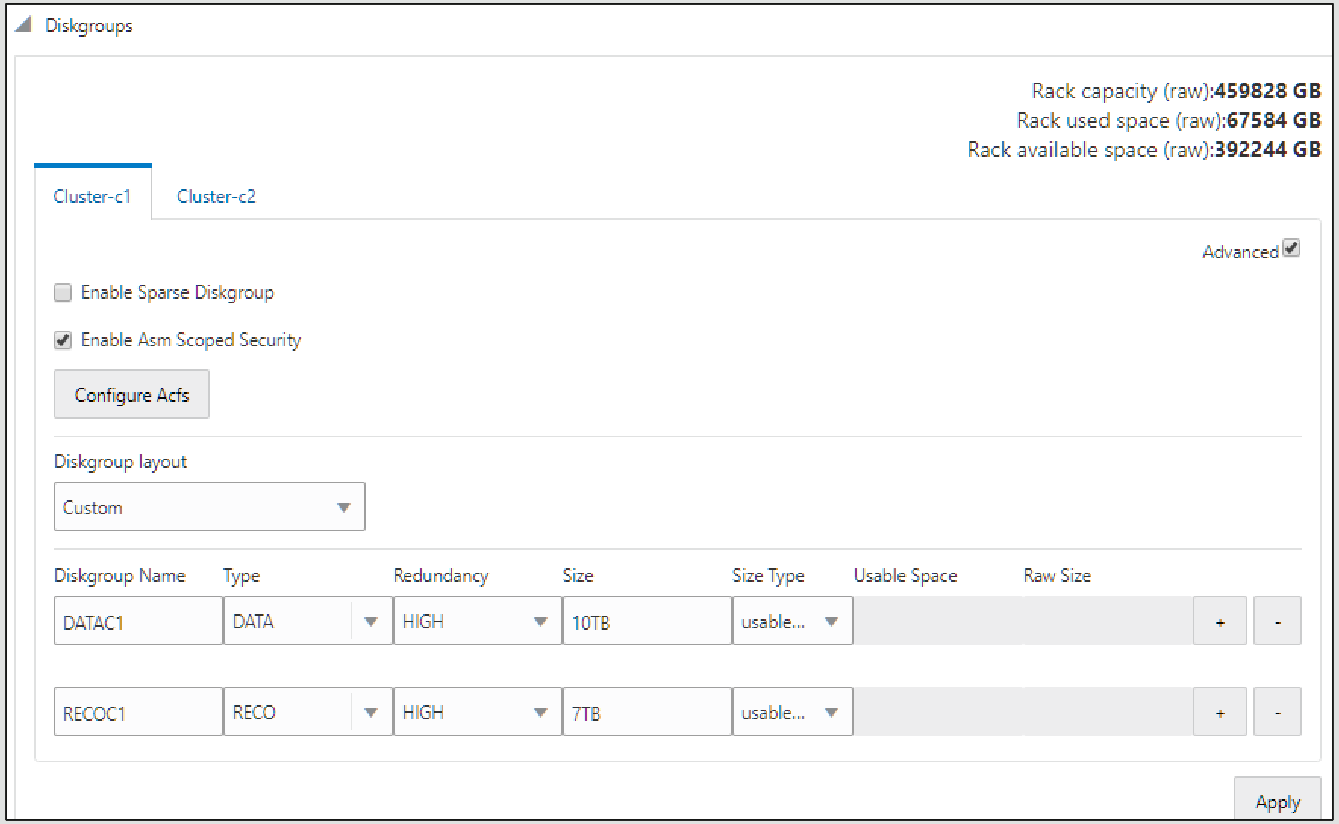
- VM size (memory, CPU)
- Exadata software version
- Networking config
- OS users and groups
- GI/DB version and location
- Starter database config
- ASM disk group config

OEDA Configuration Tool

Cluster Configuration

Grid infrastructure installed in each VM (grid disks “owned” by a VM cluster)

- Cluster 1 - DATAC1 / RECOC1 across all cells
- Cluster 2 - DATAC2 / RECOC2 across all cells
- Consider future clusters when sizing
- DBFS not configured
- **ASM-Scoped Security** permits a cluster to access only its own grid disks. Available with Advanced button.



Diskgroups

Rack capacity (raw):**459828 GB**
Rack used space (raw):**67584 GB**
Rack available space (raw):**392244 GB**

Cluster-c1 | Cluster-c2

Advanced ☑

☐ Enable Sparse Diskgroup

☑ Enable Asm Scoped Security

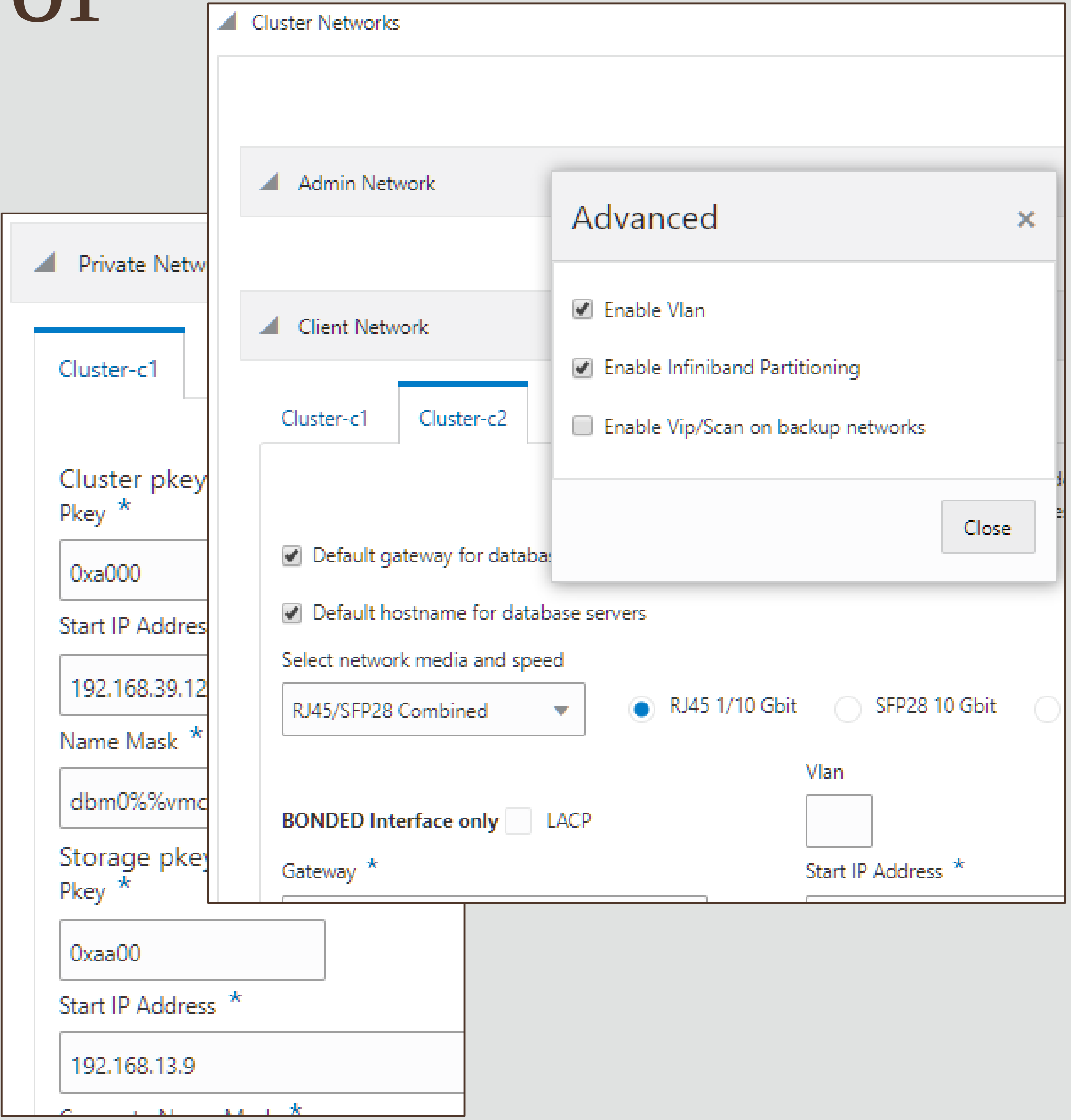
Diskgroup layout: Custom

Diskgroup Name	Type	Redundancy	Size	Size Type	Usable Space	Raw Size
DATAC1	DATA	HIGH	10TB	usable...		
RECOC1	RECO	HIGH	7TB	usable...		

OEDA Configuration Tool

Cluster Advanced Network Configuration

- Ethernet VLAN ID and IP details
  - To separate Ethernet traffic across VMs, use distinct VLAN ID and IP info for each cluster
  - Ethernet switches (customer and Cisco) must have VLAN tag configuration done before OEDA deployment
- InfiniBand PKEY and IP details
  - Typically just use OEDA defaults
  - **Compute Cluster network** for dbnode-to-dbnode RAC traffic. Separates IB traffic by using distinct Cluster PKEY and IP subnet for each cluster.
  - **Storage network** for dbnode-to-cell or cell-to-cell traffic - same PKEY/subnet for all clusters



Private Netw

Cluster-c1

Cluster pkey

Pkey *

0xa000

Start IP Addres

192.168.39.12

Name Mask *

dbm0%%vmc

Storage pkey

Pkey *

0xaa00

Start IP Address *

192.168.13.9

Cluster Networks

Admin Network

Client Network

Cluster-c1 Cluster-c2

☑ Default gateway for databa

☑ Default hostname for database servers

Select network media and speed

RJ45/SFP28 Combined

● RJ45 1/10 Gbit ○ SFP28 10 Gbit ○

Vlan

BONDED Interface only □ LACP

Gateway *

Start IP Address *

Advanced ×

☑ Enable Vlan

☑ Enable Infiniband Partitioning

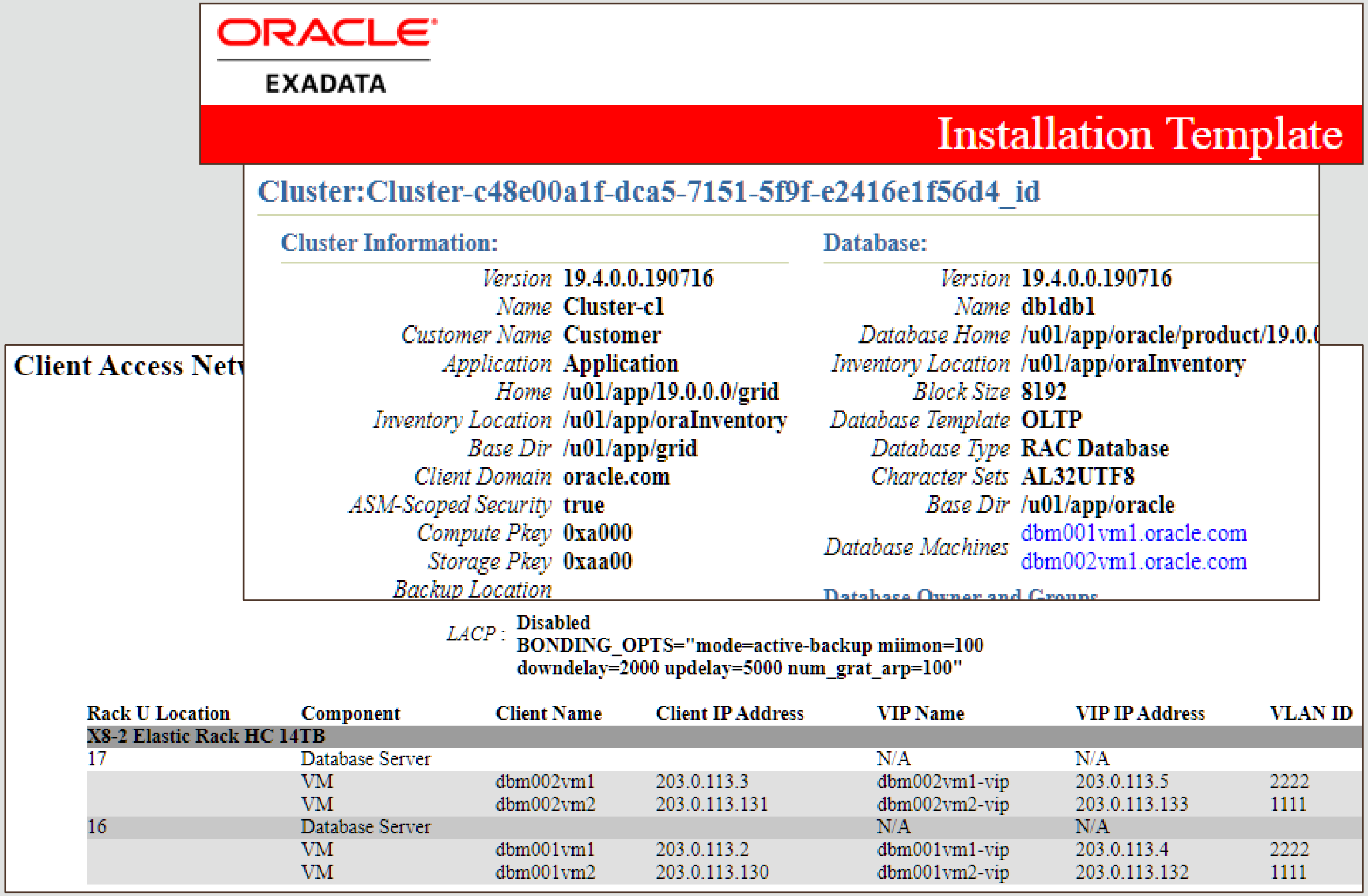
□ Enable Vip/Scan on backup networks

Close

OEDA Configuration Tool

Installation Template

Verify proper settings for all VM clusters in Installation Template so the environment can properly configured before deployment (DNS, switches, VLANs, etc.).



ORACLE
EXADATA

Installation Template

Cluster:Cluster-c48e00a1f-dca5-7151-5f9f-e2416e1f56d4_id

Cluster Information:

Version	19.4.0.0.190716
Name	Cluster-c1
Customer Name	Customer
Application	Application
Home	/u01/app/19.0.0.0/grid
Inventory Location	/u01/app/oraInventory
Base Dir	/u01/app/grid
Client Domain	oracle.com
ASM-Scoped Security	true
Compute Pkey	0xa000
Storage Pkey	0xaa00
Backup Location	

Database:

Version	19.4.0.0.190716
Name	db1db1
Database Home	/u01/app/oracle/product/19.0.
Inventory Location	/u01/app/oraInventory
Block Size	8192
Database Template	OLTP
Database Type	RAC Database
Character Sets	AL32UTF8
Base Dir	/u01/app/oracle
Database Machines	dbm001vm1.oracle.com dbm002vm1.oracle.com

Database Owner and Groups

Client Access Net

LACP : Disabled
BONDING_OPTS="mode=active-backup miimon=100 downdelay=2000 updelay=5000 num_grat_arp=100"

Rack U Location	Component	Client Name	Client IP Address	VIP Name	VIP IP Address	VLAN ID
X8-2 Elastic Rack HC 14TB						
17	Database Server			N/A	N/A	
	VM	dbm002vm1	203.0.113.3	dbm002vm1-vip	203.0.113.5	2222
	VM	dbm002vm2	203.0.113.131	dbm002vm2-vip	203.0.113.133	1111
16	Database Server			N/A	N/A	
	VM	dbm001vm1	203.0.113.2	dbm001vm1-vip	203.0.113.4	2222
	VM	dbm001vm2	203.0.113.130	dbm001vm2-vip	203.0.113.132	1111

OEDA Configuration Tool

Network Requirements

Component	Domain	Network	Example hostname
Database servers	dom0 (one per database server)	Mgmt eth0	dm01dbadm01
		Mgmt ILOM	dm01dbadm01-ilom
	domU (one or more per database server)	Mgmt eth0	dm01dbadm01**vm01**
		Client bondeth0	dm01client01**vm01**
		Client VIP	dm01client01**vm01**-vip
		Client SCAN	dm01**vm01**-scan
		Private ib	dm01dbadm01**vm01**-priv1
Storage servers (same as physical)		Mgmt eth0	dm01celadm01
		Mgmt ILOM	dm01celadm01-ilom
		Private ib	dm01celadm01-priv1
Switches (same as physical)		Mgmt eth0	dm01sw-*

Exadata OVM Basic Maintenance

Refer to Exadata Database Maintenance Guide: *Managing Oracle VM Domains on Oracle Exadata Database Machine*

- Show Running Domains, Monitoring, Startup, Shutdown
- Disabling User Domain Automatic Start
- Modify Memory, CPU, local disk space in a user domain
- Remove/Create RAC VM Cluster
- Expand Oracle RAC VM cluster
- Create User Domain without Grid Infrastructure (e.g. App VM)
- Moving a User Domain to a Different Database Server
- Deleting a User Domain from an Oracle RAC VM Cluster
- Running exachk

Exadata OVM Basic Maintenance

- Backing Up and Restoring Oracle Databases on Oracle VM User Domains
- Creating Oracle VM Oracle RAC Clusters
- Creating Oracle VM without GI and Database for Apps
- Add or Drop Oracle RAC nodes in Oracle VM
- Expanding /EXAVMIMAGES on User Domains after Database Server Disk Expansion
- Implementing Tagged VLAN Interfaces
- Implementing InfiniBand Partitioning across OVM RAC Clusters on Oracle Exadata
- Backing up the Management Domain (dom0) and User Domains (domU) in an Oracle Virtual Server Deployment
- Migrating a Bare Metal Oracle RAC Cluster to an OVM RAC Cluster

OEDACLI to Perform Maintenance Operations

- OEDA Command Line Interface
  - Orchestrate Exadata life cycle management tasks
- Supported post-deployment operations with VMs – Examples:
  - Add/Remove VM cluster
  - Add/Remove node
  - Add/Remove database
  - Add/Remove database home
  - Add/Remove storage cell
  - Resize ASM disk group
  - Upgrade Clusterware

Exadata OVM Migration

- Dynamic or online method to change physical to virtual
  - Data Guard or backups can be used to move databases – minimum downtime
  - Convert one node or subset of nodes to virtual at a time
- Migrating an existing physical Exadata rack to use virtual requires
  - Backing up existing databases, redeploying existing HW with OEDA and then Restoring Databases
  - Duplicating the databases to existing Exadata OVM configuration
  - If moving from source to a new target, standard Exadata migration practices still apply. Refer to Best Practices for Migrating to Exadata Database Machine

Exadata OVM Migration

Dynamic or online method to change physical to virtual using any of the procedures below

- Migrate to OVM RAC cluster using the existing bare metal Oracle RAC cluster with zero downtime
- Migrate to OVM RAC cluster by creating a new OVM RAC cluster with minimal downtime
- Migrate to OVM RAC cluster using Oracle Data Guard with minimal downtime
- Migrate to OVM RAC cluster using RMAN backup and restore with complete downtime

For requirements and detailed steps, refer to My Oracle Support note 2099488.1: *Migration of a Bare metal RAC cluster to an OVM RAC cluster on Exadata*

Backup/Restore of Virtualized Environment

- Dom0
  - Standard backup/restore practices to external location
- DomU – Two Methods
  - Backup within Dom0: Snapshot the VM image and backup snapshot externally
  - Backup within DomU: Standard OS backup/restore practices apply
  - If over-provisioning local disk space - Restoring VM backup will reduce (may eliminate) space savings (i.e. relying on over-provisioning may prevent full VM restore)
- Database backups/restore
  - Use standard Exadata MAA practices with RMAN, ZDLRA, and Cloud Storage

- Refer to Exadata Database Machine Maintenance Guide

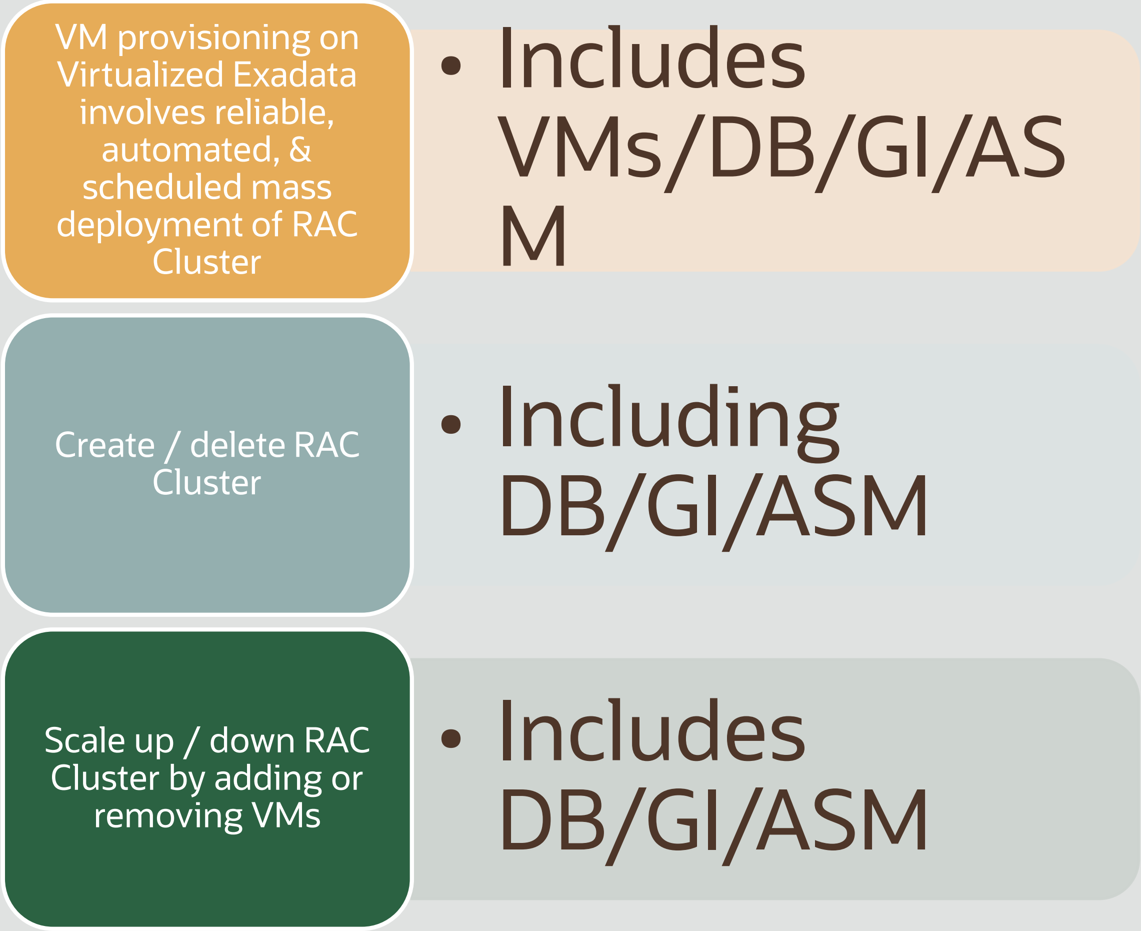
Updating Software

Component to update	Method
Storage servers	Same as physical - run patchmgr from any server with ssh access to all cells, or use Storage Server Cloud Scale Software Update feature (starting in 18.1).
InfiniBand switches	Same as physical - run patchmgr from dom0 with ssh access to all switches.
Database server – dom0	Run patchmgr from any server with ssh access to all dom0s. Dom0 update upgrades database server firmware. Dom0 reboot requires restart of all local domUs. DomU software not updated during dom0 update. Dom0/domU do not have to run same version, although specific update ordering may be required (see 888828.1).
Database server – domU	Run patchmgr from any server with ssh access to all domUs. Typically done on a per-VM cluster basis (e.g. vm01 on all nodes, then vm02, etc.), or update all VMs on a server before moving to next.
Grid Infrastructure / Database	Standard upgrade and patching methods apply, maintained on a per-VM cluster scope. GI/DB homes should be mounted disk images, like initial deployment. 12.2 upgrade MOS 2111010.1.

Health Checks and Monitoring

- Exachk runs in Dom0 and DomU (cells and IB switches checks run with Dom0)
  - Run in one dom0 for all dom0s, cells, switches
  - Run in one domU of each VM cluster for all domUs, GI/DB of that cluster
- EM Monitoring support (MOS 1967701.1)
- Exawatcher runs in Dom0 and DomU
- Database/GI monitoring practices still apply
- Considerations
  - Dom0-specific utilities (xmtop)
  - Dom0 is not sized to accommodate EM or custom agents
  - Oracle VM Manager not supported on Exadata

EM Support for Exadata Virtualization Provisioning



VM provisioning on Virtualized Exadata involves reliable, automated, & scheduled mass deployment of RAC Cluster

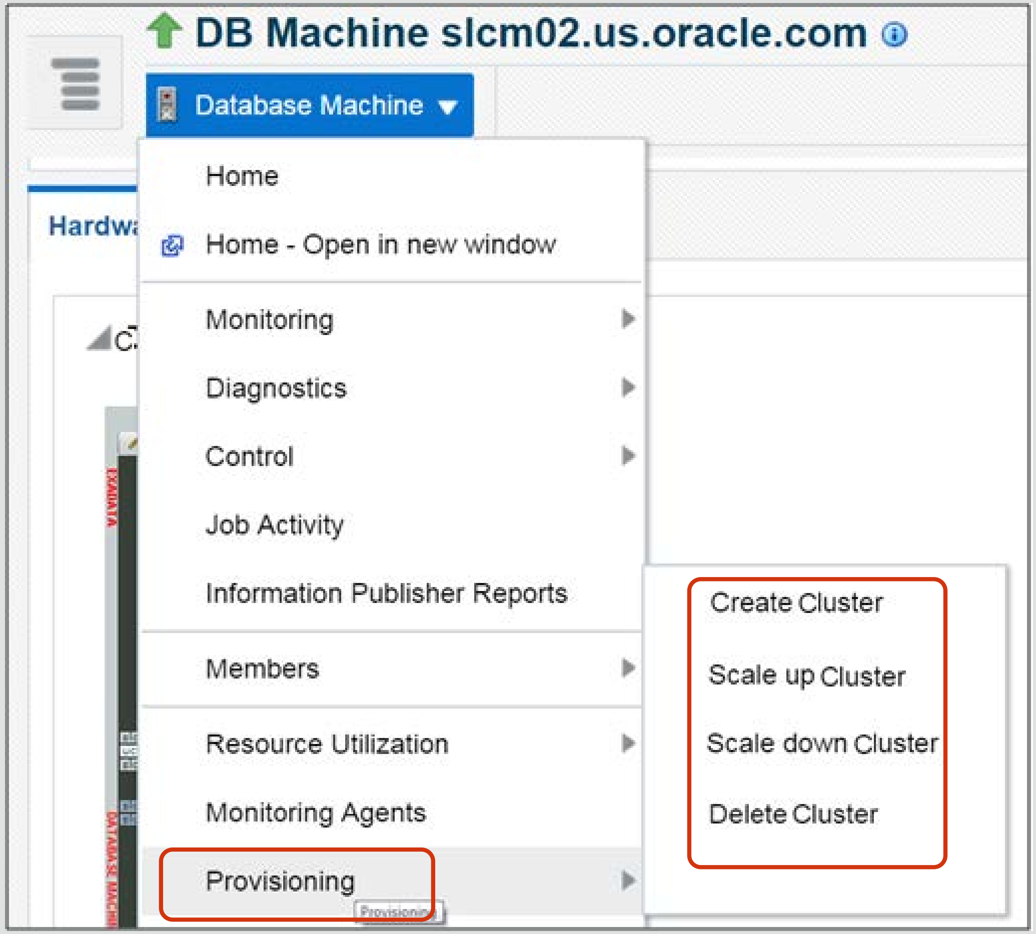
- Includes VMs/DB/GI/ASM

Create / delete RAC Cluster

- Including DB/GI/ASM

Scale up / down RAC Cluster by adding or removing VMs

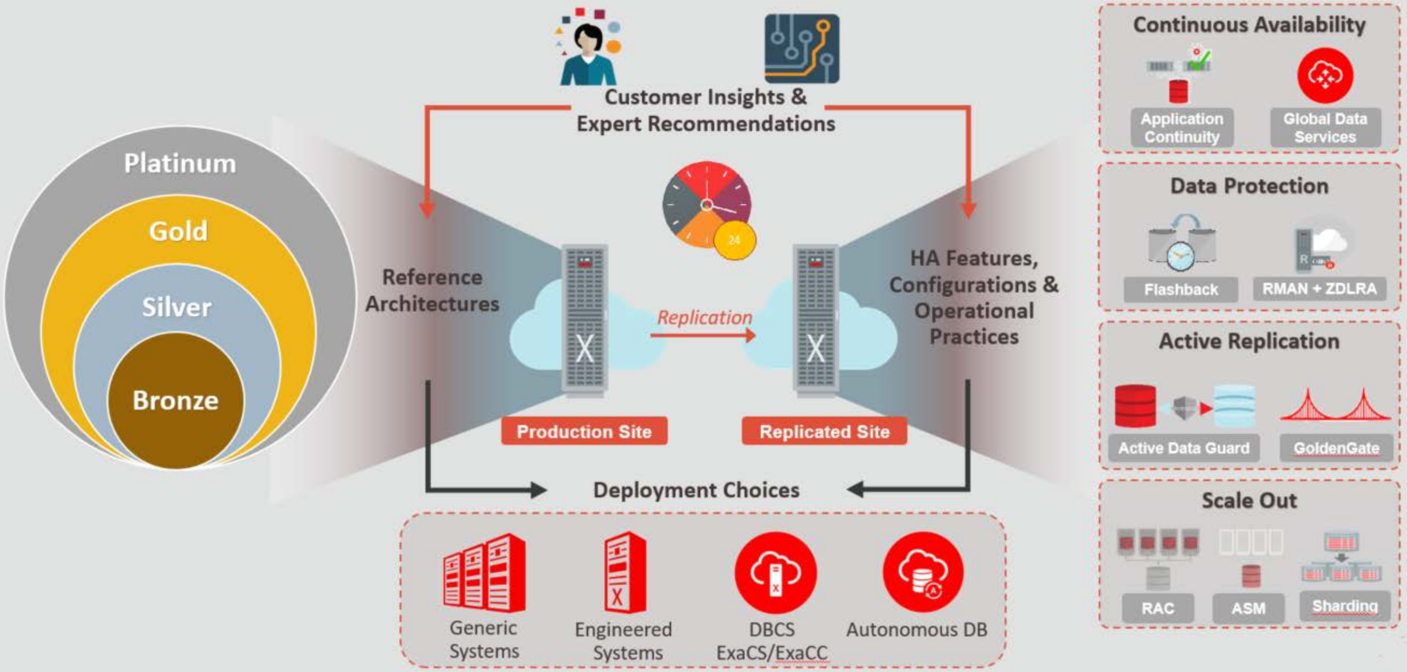
- Includes DB/GI/ASM



Increase Operational Efficiency by Deploying RAC Cluster Faster on Virtualized Exadata

Exadata MAA/HA

- Exadata MAA failure/repair practices still applicable. Refer to MAA Best Practices for Oracle Exadata Database Machine
- OVM Live Migration is not supported – use RAC to move workloads between nodes



Resource Management

- Exadata Resource Management practices still apply
  - Exadata IO and flash resource management are all applicable and useful
- Within VMs and within a cluster, database resource management practices still apply
  - cpu_count still needs to be set at the database instance level for multiple databases in a VM. Recommended min = 2
- No local disk resource management and prioritization
  - IO intensive workloads should not use local disks
  - For higher IO performance and bandwidth, use ACFS or DBFS on Exadata or NFS.

Integrated Cloud

Applications & Platform Services