

Oracle 12c RAC Database on Dell EMC XC Family with NVMe Drives Reference Architecture

Abstract

This document outlines the architectural guidelines for deploying Oracle RAC database on Dell EMC XC Family using NVMe drives. This document also provides the best practices for achieving flexibility, availability, and optimal performance.

January 2019

Revisions

Date	Description
January 2019	Initial release
January 25, 2019	Update per Nutanix.

Acknowledgements

This paper was produced by the following members of the Dell EMC HCI solution engineering team.

Author: Sanjay Puttaswamy

The information in this publication is provided “as is.” Dell Inc. makes no representations or warranties of any kind with respect to the information in this publication, and specifically disclaims implied warranties of merchantability or fitness for a particular purpose.

Use, copying, and distribution of any software described in this publication requires an applicable software license.

© 2019 Dell Inc. or its subsidiaries. All Rights Reserved. Dell, EMC, Dell EMC and other trademarks are trademarks of Dell Inc. or its subsidiaries. Other trademarks may be trademarks of their respective owners.

Dell believes the information in this document is accurate as of its publication date. The information is subject to change without notice.

Table of contents

Revisions.....	2
Acknowledgements.....	2
Executive Summary	4
1 Introduction	5
1.1 Audience.....	5
1.2 Terminology.....	5
2 Product overview	6
2.1 Dell EMC XC Family overview.....	6
2.2 Dell EMC XC Family support for NVMe drives.....	6
2.3 Dell EMC XC Family Acropolis Block Services.....	7
3 Solution infrastructure	9
3.1 Oracle RAC database configuration using Nutanix volumes.....	9
3.2 Dell EMC XC Family hardware configuration.....	10
3.3 Dell EMC XC Family network configuration (ESXi).....	11
4 Sizing and best practices	13
4.1 Oracle database VM layout.....	13
4.1.1 Processor.....	13
4.1.2 Memory.....	13
4.1.3 HugePages.....	13
4.2 VM storage adapter and virtual drives.....	14
4.3 VM networking.....	14
4.3.1 Other networking best practices.....	14
4.4 Storage layout for Oracle database.....	15
4.5 Nutanix CVM.....	17
4.6 Oracle database performance and scalability.....	17
A Appendix	19
A.1 How to identify Dell EMC XC Family volume group in Linux Guest.....	19
B Configuration details	21
C Technical support and resources	22
C.1 Related resources.....	22

Executive Summary

NOTE: The information in this document applies to both Dell EMC XC Family Series Appliances, as well as the Dell EMC XC Family Core system offerings. Sections or information that apply to only one of the offerings (XC Family Series or XC Family Core).

This document provides the reference architecture for deployment of Oracle 12c Real Application Cluster database environment on Dell EMC XC Family with Linux® operating systems. This document also highlights the NVMe drive support on Dell EMC XC Family, which enables increased performance.

The Dell EMC XC Family is a hyper-converged solution that combines compute, networking, and storage resources provided by industry-leading Dell EMC PowerEdge™ servers and Nutanix AOS software. By combining the hardware resources from each appliance into a shared-everything model for simplified operations, improved agility, and greater flexibility. Dell EMC and Nutanix together deliver simple, cost-effective solutions for enterprise workloads.

This document provides guidelines for design, configuration, and optimization of Oracle RAC databases applications running on Dell EMC XC Family Nutanix infrastructure.

1 Introduction

1.1 Audience

This document is intended for decision makers, business leaders, architects, cloud administrators, Oracle database administrators, Hyper-converged infrastructure administrators, and technical administrators of IT environments that are responsible for deployment of Oracle RAC database on Dell EMC XC Family with ESXi hypervisors.

The reader of this document must have a working knowledge of Dell EMC XC Family, Nutanix AOS, VMware vSphere technologies, Oracle database technologies, and have a basic familiarity with storage, compute, and network technologies.

1.2 Terminology

The following table lists the terminology and acronyms that are used throughout this document:

Table 1 Terminology list

Term	Description
VM	Virtual Machine
CVM	Controller Virtual Machine
DSF	Distributed Storage Fabric
NFS	Network File System
iSCSI	Internet Small Computer Systems Interface
NVMe	Non-volatile Memory Express
PCIe	Peripheral Component Interconnect Express
SSD	Solid State Drive
SATA	Serial Advanced Technology Attachment
LACP	Link Aggregation Control Protocol
MTU	Maximum Transmission Unit
VG	Volume Group
AOS	Acropolis Operating System
ABS	Acropolis Block Service
RAC	Real Application Cluster
ASM	Automatic Storage Management
AU	Allocation Unit
OCR	Oracle Clusterware Registry
SGA	System Global Area
PGA	Program Global Area

2 Product overview

2.1 Dell EMC XC Family overview

The Dell EMC XC Family of hyper-converged appliances integrates our proven PowerEdge x86 server platform and Nutanix software into purpose-built enterprise-class solutions for virtualized environments. Backed by Dell EMC's Global Service and Support organization, these 1U and 2U appliances consolidate storage, compute, networking, and virtualization into a turnkey platform enabling application and virtualization teams to quickly and simply deploy new workloads. The Dell EMC XC Family Series, with pay-as-you-grow flexibility, is easily expanded with one node at a time for predictable scale-out expansion.

Dell EMC XC Family appliances simplify the deployment of virtual machines in any environment. The Nutanix Acropolis Operating System (AOS) runs in a Controller VM (CVM) on each node, aggregating storage resources (hard disk drives and flash storage) across all nodes. This pooled storage is made available to all hosts through a fault-tolerant architecture. With an unrivaled ability to run VMs out of the box, Dell EMC XC Family appliances deliver an easy, modular approach to building modern data centers.

The Dell EMC XC Family appliance is hypervisor agnostic and supports your choice of hypervisor. With a focus on ease-of-use and the ability to rapidly deploy multiple virtualized workloads, XC Family is the preferred platform for many enterprises and midmarket customers alike. The XC Family appliances use the Nutanix Distributed Storage Fabric (DSF) that delivers a unified pool of storage from all appliances across the cluster using techniques including striping, replication, auto-tiering, error detection, failover, and automatic recovery. XC Family delivers storage through multiple protocols such as NFS, SMB, and iSCSI while leveraging standard network infrastructure.

Following are the highlighted features that are provided by Dell EMC XC Family appliances:

- The Dell EMC XC Family are hyper-converged infrastructure appliances that integrate Dell EMC's industry-leading PowerEdge servers and Nutanix software in 1U and 2U purpose-built appliances.
- Multi-hypervisor support gives users freedom to choose their hypervisor from industry-leading hypervisors without lock-in.
- Start small with as few as three nodes and scale out in single node increments. Each node adds compute, storage, virtualization, and management resources, and can be joined to a cluster with zero downtime. Storage only nodes are also available.
- Simplified infrastructure management through Prism provides a highly intuitive management console. It has an easy-to-use user interface optimized for one-click simplicity.
- Single point of global 24x7 support for both the hardware and software is provided through Dell EMC ProSupport.

2.2 Dell EMC XC Family support for NVMe drives

NVM Express, or NVMe, is a standardized high-performance software interface for PCI Express solid-state drives (SSDs) that uses nonvolatile memory (NVM). The NVMe interface is designed to take advantage of the internal parallelism and low latency inherent to flash-based storage devices. NVMe has garnered industry attention as the next step in providing faster access to and from storage devices.

NVMe protocol supports multiple deep queues, which are an advancement to over traditional SAS and SATA protocols. Typical SAS devices support up to 256 commands and SATA devices support up to 32 commands in a single queue. These read and write commands were adequate for hard disk drive technologies but cannot take full advantage of current and next-generation NVM technologies.

In contrast, NVMe supports 64-K commands per queue and up to 64-K queues. These queues are designed so that I/O commands and responses to those commands operate on the same processor core and can take advantage of the parallel processing capabilities of multicore processors. Each application or thread can have its own independent queue, so no I/O locking is required.

The Dell EMC XC740xd-24 appliance provides scalable storage that enables you to adapt to your workload and operational demands. With comprehensive storage options, the Dell EMC XC740xd-24 appliance supports drive types of Express Flash PCIe SSDs that accelerate performance over previous drive technologies. Dell EMC XC740xd-24 appliance use PCIe lanes to connect directly to the processor and chipset and are accessible through a hot-pluggable drive bay.

NVMe performance considerations:

- Nutanix AOS has two components in the system, one is the oplog, which absorbs bursts of random writes to provide a low-latency front-end for writes and the other is the extent store which is the final resting place for all data. In a Dell EMC XC Family cluster with both NVMe and SSD drives, the oplog is contained entirely in the NVMe drives, while the extent store treats NVMe drives and SSDs uniformly without making any distinction among them. This way writes going to the oplog only engage the NVMe drives (until they are drained to the extent store) but reads and writes going to the extent store will engage all NVMe drives and SSDs uniformly.
- NVMe devices are highly parallelized; their maximum throughput is achieved when they are backing multiple virtual drives and/or multiple virtual machines.
- The high throughput of NVMe creates a correspondingly high CPU load. NVMe devices are thus a better fit for hardware configurations with many cores per socket (preferably at least eight) and multiple sockets (preferably at least two). High CPU frequencies are also desirable.

2.3 Dell EMC XC Family Acropolis Block Services

A feature called Acropolis Block Services (ABS) was released with Acropolis OS 4.7 (AOS). It enables DSF resources to be exposed directly to a virtualized guest operating system or physical hosts using the iSCSI protocol. This capability enables support for several use cases such as shared storage for Oracle RAC and other applications that require shared storage.

Dell EMC XC Family Nutanix storage configuration for ABS is handled through a construct that is called a volume group (VG). A VG is a collection of volumes that are commonly known as virtual disks (vDisks). ABS presents these vDisks to virtual machines and physical servers using iSCSI protocol. Multiple hosts can share the vDisks associated with a VG as shown in Figure 1. This is helpful for shared storage use cases such as Oracle RAC or Windows server clustering. To find out how to create VG and discover vDisks, see Appendix A.1.

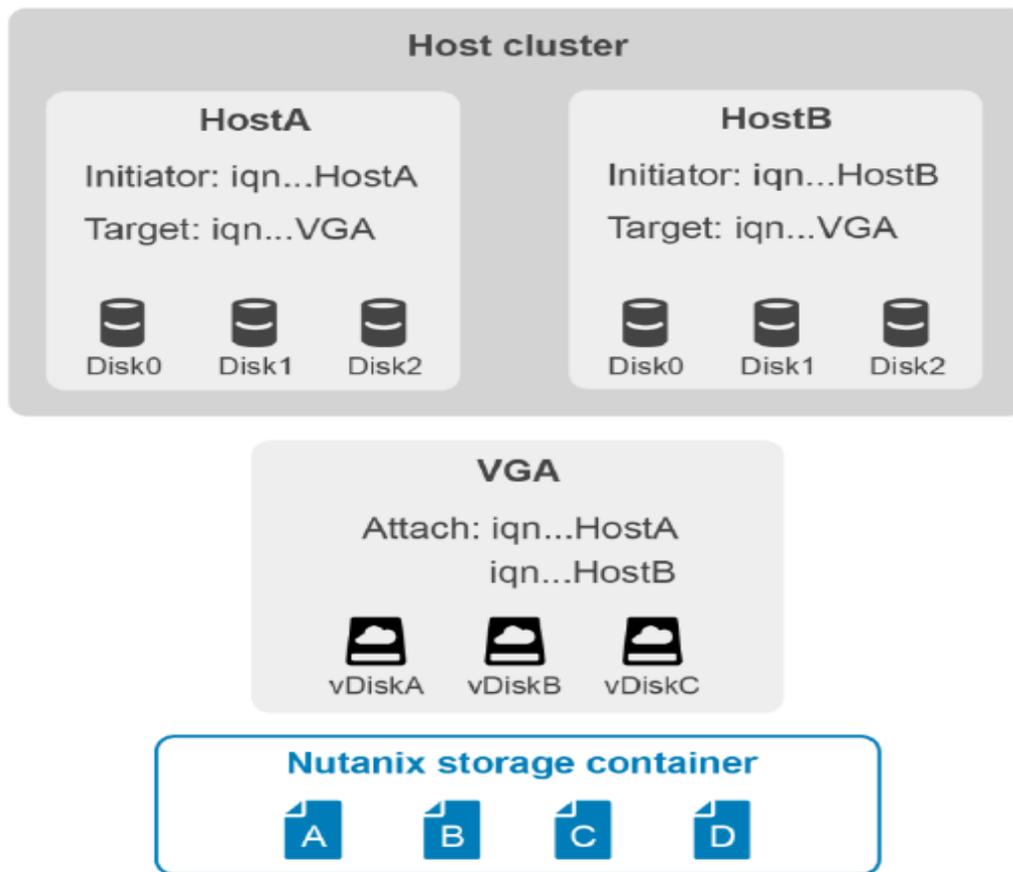


Figure 1 Acropolis Block Services architecture

Nutanix Volumes with Load Balancing (VGLB) functionality has been introduced with AOS 5.6. Load balancing of vDisks in a volume group enables IO-intensive VMs to use resources such as the CPU and memory from multiple Controller VMs (CVMs). vDisks belonging to a volume group are distributed across the CVMs in a cluster, which improves the performance and prevents bottlenecks, see [AHV-Admin-Guide-v56](#).

3 Solution infrastructure

This section outlines the architectural design goals for Oracle RAC with Dell EMC XC Family Series with VMware ESXi hypervisor.

- **Easy to deploy:** The Dell EMC XC Family Series appliance integrate server, storage and networking one chassis.
- **High Performance:** The solution designed and tuned for OLTP-style applications. The information life cycle management (ILM) process that automatically handles data placement. Applying ILM to the data life cycle in ORADB, the high-performance SSD tier automatically receives all hot data writes. Both hot and warm data reside in this tier to provide easy access and the highest performance.
- **High availability and data redundancy:** This allows the Oracle database and supporting applications to stay up and running even if the underlying hardware has issues.
- **Easy to scale:** Scale both performance and capacity, linearly and nondisruptively, as needs grow.
- **More storage.** Adequate storage capacity to support multiple multi-TB databases.
- **Simplified management:** Manage without having to deal with storage-related management tasks, such as multipathing, zoning, and masking.
- **Oracle Automatic Storage Management:** Support for Oracle Automatic Storage Management (ASM) without any caveats or restrictions.

In this solution, Oracle 12c RAC database uses a Nutanix DSF as the shared storage to store Oracle database files and the Oracle Clusterware Registry (OCR) and voting disk files of Oracle 12c Grid Infrastructure. The database nodes are also connected to private network for the Oracle cluster heartbeat and for the Oracle RAC data synchronization between the Oracle RAC database nodes. All the database nodes are connected to the public network, which allows applications and DBAs to connect to the database nodes. The public network is connected to the main corporate network.

While it is possible to support multiple Oracle database VMs on a same physical node, for high availability and fault tolerance, Dell EMC recommends spreading them out on multiple nodes and minimize the number of database instances running on the same node. For Oracle RAC, the RAC-instance VMs should run on different physical nodes. Affinity rules to be setup so that a RAC-instance VM always runs on a certain Dell EMC XC Family cluster host. Anti-affinity rules need to be setup so that two or more RAC-instance VMs never run on a single Dell EMC XC Family cluster host if there is a node failure.

3.1 Oracle RAC database configuration using Nutanix volumes

The Nutanix volumes feature enables DSF storage resources to be presented directly to VMs and physical servers using iSCSI. In this configuration, the database disk groups are presented to the VMs using iSCSI. A volume group consisting of multiple vDisks is created to store database-related files as shown in Figure 2.

Oracle RAC enables running multiple Oracle instances on multiple servers in the cluster against a single database. The database is hosted on a load balanced shared volume that is presented to Oracle instances using iSCSI. The end user applications access the database as traditional Oracle database. This architecture helps provide the highest availability and reliability to the Oracle database applications.

The architecture in Figure 2 shows four VMs on each host are grouped to form an Oracle RAC cluster.

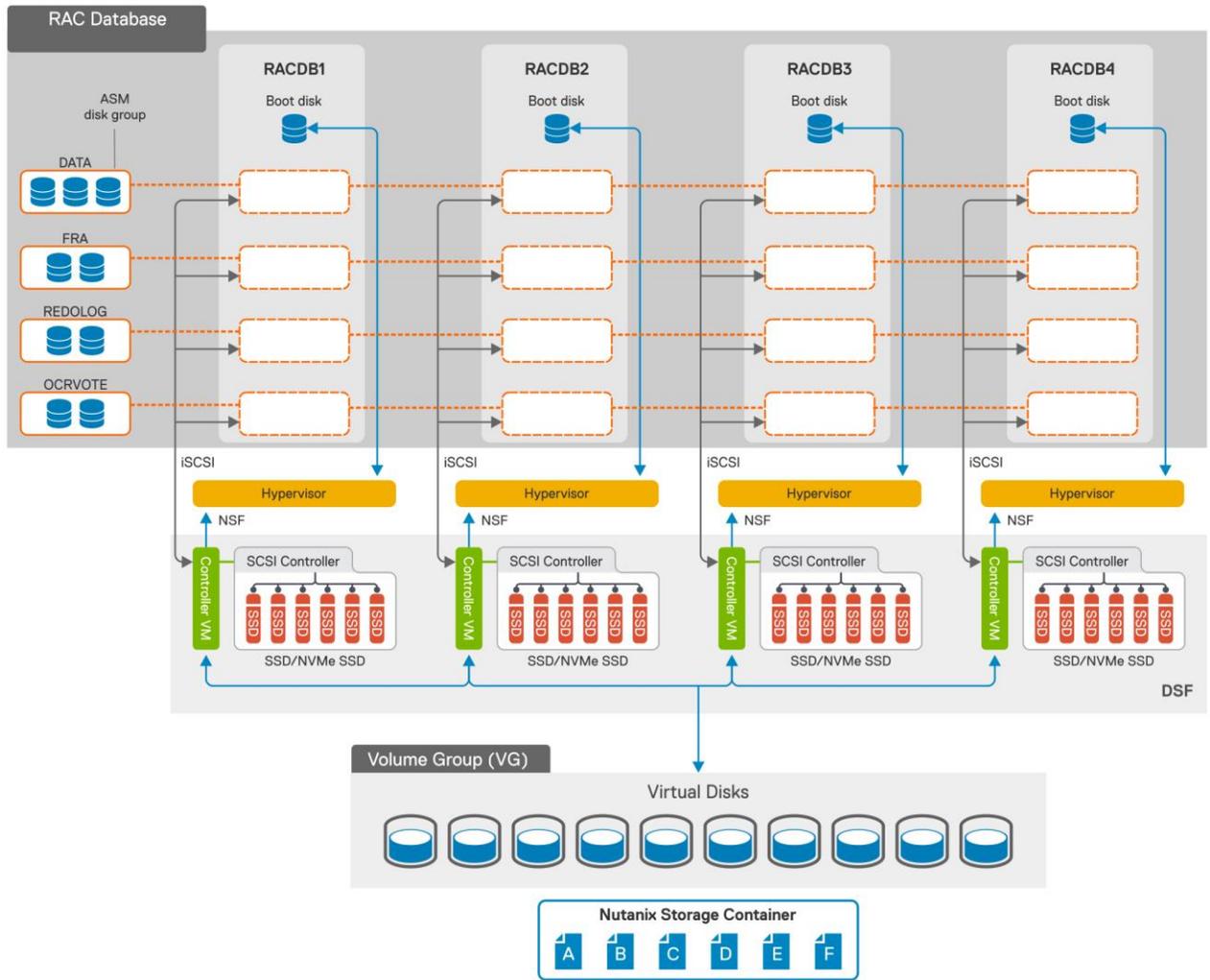


Figure 2 Oracle RAC configuration using ABS

Oracle Automatic Storage Management (ASM) is highly recommended for database-related files. The vDisks presented to each VM is mapped as an Oracle ASM disk and ASM disk groups are carved out using the ASM disks. The ASM disk groups are also shown in Figure 2.

3.2 Dell EMC XC Family hardware configuration

The following table lists the recommended hardware configuration per host for running Oracle RAC database on Dell EMC XC Family Series appliance:

Table 2 Dell EMC XC 740xd-24 hardware configuration

Dell EMC XC740xd-24 configuration details (NVMe + SSD)	
CPU	2 x Intel(R) Xeon(R) Gold 6134 (8C, 3.20 GHz)
Storage	<ul style="list-style-type: none"> T1: 4 x 745 GB NVMe SSDs 8 Gbps T2: 18 x 894 GB SAS SSDs 12 Gbps
Memory	12 x 32 GB DDR-4 2666 MHz (384 GB Total)
Network	1 x dual MLNX 25 GbE 2P ConnectX4LX RNDC

Dell EMC XC740xd-24 configuration details (NVMe + SSD)	
iDRAC	iDRAC9 Enterprise
Power	2 x 1100 W PSUs

The minimum number of Dell EMC XC Family hosts in a cluster is three. We used four Dell EMC XC Family Series nodes for this solution. When clustered together, the storage across four nodes is virtualized together to create a single storage pool. You can create one or more storage containers within a storage pool. The storage container is presented to all nodes as shared storage within the cluster.

The cluster attempts to keep virtual machines and their associated storage on the same cluster node for performance consistency. However, each cluster node is connected to, and communicates with, the other nodes on a 25GbE network. This communication enables virtual machines and their associated storage to reside on different cluster nodes as well.

The detailed hardware configuration specification of Dell EMC XC Family Series appliance, the cluster configuration, guest VM specifications, and network configuration details are provided in Appendix B.

3.3 Dell EMC XC Family network configuration (ESXi)

Figure 3 following exhibits dual switch ToR topology, two 25GbE switches are configured as VLT peers. The Dell EMC XC Family Series appliance hosts connect to both switches using a single NIC port connecting to each switch. This is the recommended topology for networking within a rack or cluster, it protects against a single switch failure.

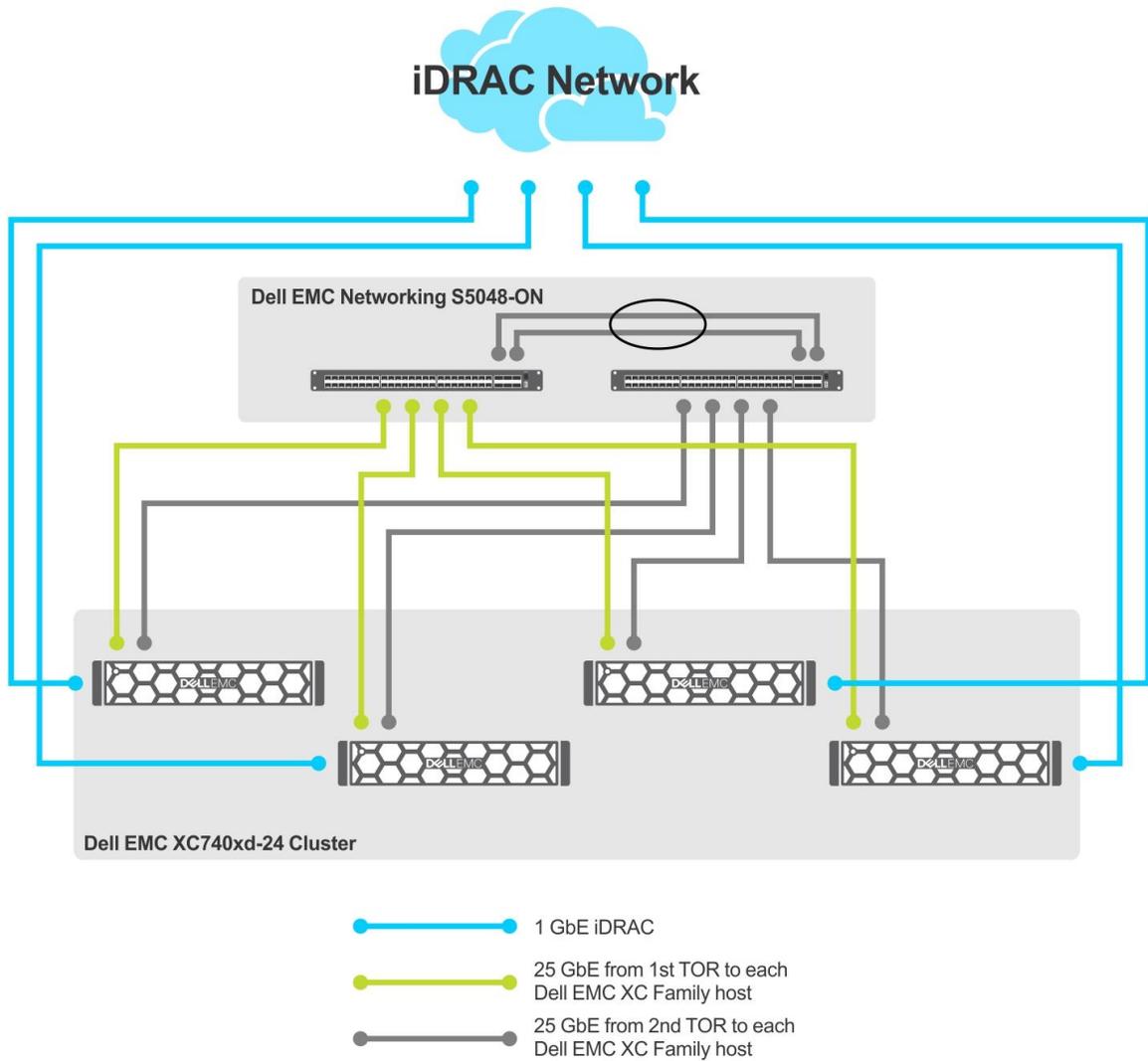


Figure 3 Dell EMC XC Family cluster network configuration

4 Sizing and best practices

4.1 Oracle database VM layout

You can run Oracle database in a VM; however, Oracle is resource-intensive by nature so virtualization of Oracle database must adhere to best practices. Each Oracle RAC node is installed with Red Hat Enterprise Linux Server release 7.4 operating system and configured with 16 vCPU and 256-GB memory, with 76 GB assigned to Oracle SGA (System Global Area) and volume group of disks are created using Nutanix volumes.

4.1.1 Processor

VMware vSphere® Virtual Symmetric Multiprocessing (Virtual SMP) enhances virtual machine performance by enabling a single virtual machine to use multiple physical processors simultaneously. The biggest advantage of an SMP system is the ability to use multiple processors to run multiple tasks concurrently, thereby increasing throughput. The Oracle architecture is multithreaded and includes multiple processes which make it a suitable candidate to take advantage of Virtual SMP. Sizing virtual CPUs of the virtual machine requires thorough understanding of Oracle workload. If the exact Oracle workload is not known, start with 8 virtual CPUs and increase the number later if necessary.

The following are virtual CPU-related best practices:

- Hyper-threading is a hardware technology on Intel processors that enables a physical processor core to act like two processors, enabling two independent threads to run simultaneously. Make sure hyper-threading enabled for Intel Core i7 processors.
- Avoid CPU over commitment to avoid adversely affecting virtual machine performance.
- Avoid CPU affinity on systems with hyper-threading.

4.1.2 Memory

You should be sure to allocate enough memory to hold the working set of applications you run in the virtual machine, thus minimizing thrashing. Because thrashing can dramatically impact performance, it is important not to under-allocate memory. Sizing the memory of the virtual machine should be large enough to accommodate the total sum of the size of the Oracle SGA, PGA, the Oracle background processes, and operating system used memory in environments to avoid kernel swapping between hypervisor and the guest operating system.

The following are virtual memory best practices:

- Use large memory pages.
- Do not turn off ESXI memory management mechanisms.
- Avoid overcommit of memory to avoid adversely affecting virtual machine performance.

4.1.3 HugePages

The HugePages feature enables Linux kernel to manage large pages of memory in addition to default 4k page size on x86 and x86_64. HugePages can give substantial benefits in virtual memory management by increasing the page size to 2 MB, thereby reducing the number of managed pages. HugePages feature must be enabled for a system with more than 16 GB of memory running Oracle database with total System Global Area (SGA) larger than 8 GB.

NOTE: The Automatic Memory Management (AMM) and HugePages features are not compatible in Oracle Database 11g and later. You must disable AMM to use HugePages.

4.2 VM storage adapter and virtual drives

Typically, an Oracle database spans across multiple LUNs to increase performance by allowing parallel I/O streams. In a virtualized environment, multiple virtual drives are used instead. Dell EMC recommends having at least four to six database virtual drives and add more disks, depending on the capacity requirements, to achieve better performance.

It is best practice to create multiple controllers and separate guest operating system virtual drive from database virtual drives. For most guest operating systems, the default virtual storage adapter in ESXi 6.5 is either LSI Logic Parallel or LSI Logic SAS, depending on the guest operating system and the virtual hardware version.

4.3 VM networking

A minimum of two 25GbE interfaces are recommended for each Dell EMC XC Family Series host. The actual number that is required depends on how many vSwitches and the total network bandwidth requirement. Each host should connect to dual redundant switches for network path redundancy as described in section 3.3. Table 3 shows number of vSwitches and their target usage.

Table 3 vSwitches and their target usage

vSwitch	Physical adapters	Network	Usage
vSwitchNutanix	NA	svm-iscsi-pg vmk-svm-iscsi-pg	Primary storage communication path
vSwitch0	vmnic0	Management network, VM network	Management traffic, VM public traffic, inter-node communication
vSwitch1	Vmnic1	VM network: Oracle	Private Oracle RAC interconnect

4.3.1 Other networking best practices

Additional networking best practices include the following:

- Dell EMC recommends using dedicated NICs on the hosts for management and RAC-interconnect traffic. Also, dedicated VLANs are recommended to segregate each type of traffic.
- For a standard virtual switch configuration, the default load-balancing policy is recommended: Route based on originating virtual port.

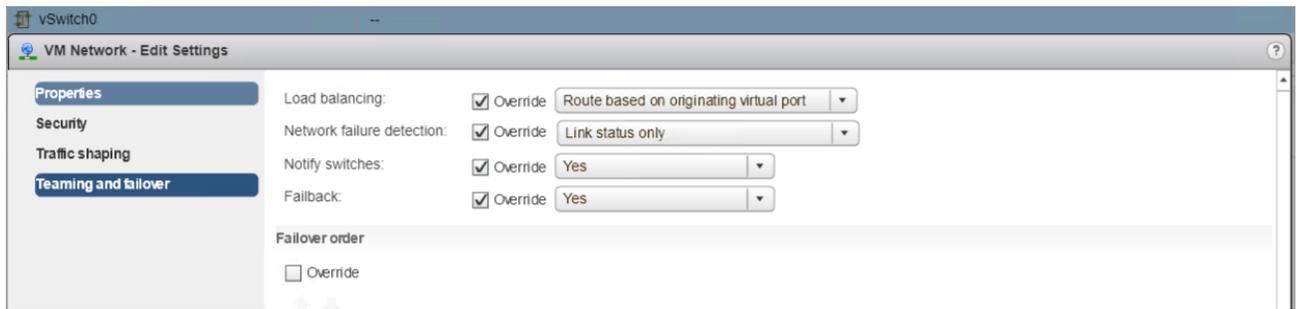


Figure 4 Standard virtual switch configuration

These attributes help simplify implementation for configurations such as LACP.

- The standard network packet size is 1500 MTU. Jumbo frames send network packets in a larger size of 9000 MTU. Increasing the transfer unit size enables more data to be transferred in a single packet which results in higher throughput, and lower CPU utilization and overhead. Use Jumbo frames only when all the network devices—including the network switches, CVMs, VMs, and ESXi hosts—on the network path can support the same MTU size.

4.4 Storage layout for Oracle database

Oracle Automatic Storage Management (ASM) is a volume manager and file system that is built into the Oracle Database server. Raw disk volumes are allocated to ASM for management and control in the same way that raw volumes are managed by a volume manager. Oracle ASM is the preferred storage management solution for either single instance database or RAC. Oracle ASM takes place of the traditional Linux volume manager and file system. When proper ownership (grid:asmadmin) and permission (660) are set on the disks, ASM takes over the management of the disks and create disk groups where data files reside and are managed automatically. ASM is highly integrated with, and highly optimized for, the Oracle Database.

Some of the benefits of ASM include:

- Automatic file management
- Stripes files rather than logical volumes
- Enables online disk reconfiguration and dynamic rebalancing
- Provides adjustable rebalancing speed
- Is automatically installed as part of the Grid Infrastructure

The most common way to consume Dell EMC XC Family Series storage is to mount as a datastore in the ESXi hypervisor through NFS. Create multiple volume groups for different types of Oracle database files (data files, redo log files, archive logs, OCR files) using Nutanix ABS. We attach the volumes to multiple Linux guest operating system virtual machines with iSCSI initiators. Virtual disks are then created, and Oracle ASM take control of these virtual drives like any hard drives. Thus, no extra configuration steps are required to make Dell EMC XC Family Series storage work with ASM.

The following Oracle ASM best practices:

- Configure ASM disk groups for external redundancy (except quorum groups). In External Redundancy mode, there is only one copy of the data. Since Dell EMC XC Family Series storage already offers data protection and HA by creating two replicas of the data on the storage level by default, it is sufficient to choose External Redundancy for the ASM disk groups.
- Use a 1 MB ASM allocation unit (AU) size for ASM disk groups. ASM disks are divided into Allocation Units. Data within a file residing on ASM is broken up into extents. As the AU size increases, the

extent count decreases, reducing the number of physical I/Os required when performing I/O operations. For OLTP, 1- 4 MB is sufficient.

- Use different ASM disk groups for OCR, database files, redo log, and archive log storage. Place each group of files in a different ASM disk group.
- ASM disk groups should contain uniform-sized disks, the amount of data on each disk is approximately the same. This uniform-sized disk ensures in reducing I/O bottleneck within a disk group and also rebalance will be faster whenever the storage configuration of a disk group changes such as when a disk is added.
- Oracle supports asynchronous I/O by default. The **disk_asynch_io** database parameter is set to **true**. For all-flash configurations, Dell EMC recommends the **NOOP** I/O scheduler because it can reduce I/O latency, improve throughput, and help reduce CPU cycles. The following shows an example of the udev rule to set IO scheduler as NOOP.

```
echo noop > /sys/block/$device/queue/scheduler
```

- Set the Linux maximum I/O size (max_sectors_kb) to match ASM AU size in rc.local or via UDEV.

```
echo 1024 > /sys/block/$device/queue/max_sectors_kb
```

- Set rotational value to 0 for SSDs to disable unneeded scheduler logic meant to reduce number of seeks.

```
echo 0 > /sys/block/$device/queue/rotational
```

- Use Linux device management facility, udev, to assign ownership and access permissions persistently on disk volumes. Without proper permission, ASM cannot manage and control the disk volumes. For example, the udev rule specifies a disk volume with the device id/WWN, 1NUTANIX NFS_3_0_850_9cb927d8_c44a_4eb1_962e_5e9e6d2567cf, creates a symlink in /dev/asmdata1, and assigns grid:oinstall and 0660 permission to the device file.

```
/etc/udev/rules.d/99-oracle-asm.rules:
KERNEL=="sd*", SUBSYSTEM=="block", PROGRAM="/lib/udev/scsi_id --page=0x83
--whitelisted --device=/dev/%k", RESULT=="1NUTANIX
NFS_3_0_850_9cb927d8_c44a_4eb1_962e_5e9e6d2567cf", SYMLINK+="asmdata1",
OWNER="grid", GROUP="asmadmin", MODE="0660"
```

Table 4 outlines the ASM disk group configuration.

Table 4 ASM disk group configuration

Disk	Protocol	Nutanix Volume Group	Number of volumes per ABS Volume Group	Size of Volumes	ASM disk group
Database Data	iSCSI	racdatavg	6	300 GB	DATA
Online Redo Logs	iSCSI	racredovg	4	20 GB	REDOLOG
Archive Logs and database Backup	iSCSI	racfravg	4	200 GB	FRA
OCR/Vote Disk	iSCSI	racocrvg	3	10 GB	OCR

4.5 Nutanix CVM

The Nutanix CVMs are virtual machines running on Dell EMC XC Family Series appliance. They have direct control of the SSDs, NVMe drives, and HDDs on the ESXi hosts through VMDirectPath I/O. They come preconfigured with modest settings so that customers can start using the storage platform without much startup time. However, for heavy workload applications such as databases, special attention should be paid to the CVM resources since they serve as the primary access points to the SSDs and HDDs. During internal database workload testing, the CVMs processor and memory resources were strained and limited the I/O performance. Dell EMC recommends increasing each CVM vCPU to a minimum of 12 and memory to 32 GB when heavy I/O workloads are planned.

4.6 Oracle database performance and scalability

Follow these guidelines at the start of an Oracle Database on Dell EMC XC Family Series cluster:

- Use the Oracle Validated Package (RPMs) for your Database version and Oracle standard operating system recommendations
- SGA size can be set in between 50 to 75 percent of allocated RAM for OLTP workload.
- PGA size depends on the number of connections and the required sort area. At the beginning, PGA can be set to 15 percent for OLTP workload.
- Utilize multiple disks for redo log, archive log, and database table spaces.
- Disable SELinux on all the Oracle RAC instance VMs
- Use Automatic Shared Memory Management (ASMM).
- Configure sufficient redo log groups, number of redo files, and redo file size to avoid any check-pointing errors.
- Make sure to set `filesystemio_options=setall` to facilitate parallel queuing of redo writes.
- Add additional disks to Oracle ASM disk groups (s) to spread the data files across more disks to avoid I/O subsystem contention.
- Dell EMC recommends the following Oracle database parameter settings:

```
archive_lag_target=900
db_block_size=8192 (or 16384 for OLAP or DW)
db_file_multiblock_read count - do not set!
```

```
db_writer_processes= (CPU count/4)
fast_start_mttr_target=120
parallel_threads_per_cpu=1
recyclebin=off
```

- Use Automatic Segment Space Management (ASSM) for Permanent Tablespace.
- Use Locally Managed Temporary Tablespace in RAC.
- Gather statistics for all objects to manage optimizer statistics.
- Use AWR (Automatic Workload Repository) baseline for comparative performance analysis.
- Use ADDM (Automatic Database Diagnostic Monitor) for database-wide performance diagnostic.

A Appendix

A.1 How to identify Dell EMC XC Family volume group in Linux Guest

Correctly identifying the database disks and setting the appropriate ownership and permission on them are critical steps to take to ensure that Oracle can properly take control of the disks. One of the most reliable methods to identify the correct disks to use is to query the disk ids or WWNs. In the Linux guest, the `scsi_id` command is commonly used to query the information. On the storage platform, the disks information is unzipped from the CVM. The following subsections describe the procedures to query the disk information.

In this solution, Dell EMC XC Family Series uses iSCSI redirection connectivity method to control target path management for vDisk load balancing and path resiliency.

NOTE: Beginning with the AOS 4.7 release, the preferred method for external connectivity to VGs does not use MPIO for storage load balancing or storage path resiliency. Instead of configuring host iSCSI client sessions to connect directly to CVMs, we use an external data services IP address. This data services IP acts as a discovery portal and initial connection point.

Perform the following steps:

1. Create volume group in Nutanix Prism. Click **Storage -> Volume Group**
2. Connect to one of the CVMs and invoke the **ncli** interface and issue the **volume-group** list command. A list of volume groups and the volumes within each volume group will be displayed.
3. Connect to Linux guest VM and install the iSCSI initiator package

```
%yum install install iscsi-initiator-utils
```

1. Check to ensure that the iSCSI daemon is running on the Linux guest VM

```
%systemctl status iscsid
```

2. Start the service if the status is not returned as running.
3. Get the iSCSI Initiator Name on the Linux Guest VM

```
%cat /etc/iscsi/initiatorname.iscsi
```

4. Attach the initiator IQN to the VG. Do this for each Linux guest VM participating in the cluster.

```
<acropolis> vg.attach_external vg_name_initiator_iqn
```

5. Discover the available iSCSI targets on the host.

```
%iscsiadm -m discovery -t st -p <external_data_services_ip_address>
```

Where -p is the external data services IP address.

6. Connect to the discovered targets on the host.

```
%iscsiadm -m node -login
```

7. Verify that the vDisk is attached.

```
%ls --l /dev/disk/by-path
```

The output shows the target name and LUN device mapping.

```
lrwxrwxrwx. 1 root root 9 Jul 23 01:20 ip-100.80.111.184:3260-iscsi-  
iqn.2010-06.com.nutanix:racdatavg-4fbeb5ea-cbe2-484a-a69f-e835dd3ab09b-  
tgt0-lun-0 -> ../../sdb
```

B Configuration details

Solution configuration: hardware components		Description
Dell EMC XC Family Series Cluster	4 x XC740xd-24: <ul style="list-style-type: none"> • ESXi 6.5 • AOS 5.8.1 • 2 x Intel(R) Xeon(R) Gold 6134 (8C, 3.20 GHz) • 384 GB RAM ((12 x 32 GB DIMMs) • 4 x 745 GB NVMe SSDs 8 Gbps • 18 x 894 GB SAS SSDs 12 Gbps • 1 x dual MLNX 25 Gb 2P ConnectX4LX RNDC 	A single VM is hosted on each server and Oracle RAC database will be configured using the four VMs.
Guest VM Configuration	4 x Linux VMs: <ul style="list-style-type: none"> • VM version: 13 • 16 vCPUs • 64 GB RAM 	The vCPU and memory configuration for Oracle instance VM
Nutanix CVM Configuration	<ul style="list-style-type: none"> • 12 vCPUs • 64 GB RAM • Nutanix Container: ORAC 	The vCPU and memory configuration for CVM. Used Non-default Nutanix storage container for Oracle RAC database.
Oracle database	Oracle 12c R1 RAC database: <ul style="list-style-type: none"> • Oracle 12c Clusterware and ASM • Oracle 12c EM Express • HammerDB tool • Silly Little Oracle Benchmark (SLOB) 	The HammerDB and SLOB tool used to simulate real-world Oracle database transactions.
Network	<ul style="list-style-type: none"> • 2 x Dell EMC Networking S5048-ON switches 	Dual redundant switches were used for high availability.
Performance monitoring	<ul style="list-style-type: none"> • Nutanix Prism • VMware vCenter • Oracle 12c EM Express 	Performance monitoring for the XC Family Series cluster was done using Nutanix Prism and vCenter. Oracle database I/O was monitored using Oracle 12c EM Express which is shipped as part of Oracle 12c.

C Technical support and resources

Dell.com/XCSeriesmanuals is focused on meeting customer needs with proven services and support.

[Dell EMC Technical Resource Center](http://DellEMC.com/TechnicalResourceCenter) on DellEMC.com provides expertise that helps to ensure customer success on Dell EMC XC Family Series platforms.

[Dell EMC XC Series Hyper-Converged Appliances](http://DellEMC.com/XCSeriesHyperConvergedAppliances) on DellEMC.com provides marketing materials and the latest product information on the XC Family Series.

C.1 Related resources

See the following referenced or recommended resources related to this document:

Oracle Best Practices Guide: http://go.nutanix.com/rs/nutanix/images/BP-2000_Oracle_on_Nutanix_Best_Practices.pdf

Dell EMC XC Family Series Networking Deployment and Best Practices Guide:
https://downloads.dell.com/solutions/networking-solution-resources/XC_Series_Networking_Guide_v2.0.pdf

Nutanix Bible: <http://nutanixbible.com/>

Nutanix blog: <http://longwhiteclouds.com/>

Acropolis Block Services: <http://go.nutanix.com/rs/031-GVQ-112/images/acropolis-block-services.pdf>

AHV Admin Guide <https://portal.nutanix.com/#/page/docs/details?targetId=AHV-Admin-Guide-v55:AHV-Admin-Guide-v55>

Oracle White Paper: <https://www.emc.com/collateral/TechnicalDocument/deploying-oracle12c-rac-db-on-xcseries-all-flash.pdf>

iSCSI settings on the Guest operating system for Nutanix Volume Groups configuration:
https://portal.nutanix.com/#/page/solutions/details?targetId=BP-2105-Linux-on-AHV:top_iscsi_settings.html

XC Family Series and XC Family Core Technical Resource Center
<http://dell EMC.com/xcseriesolutions>