Deploying Oracle Maximum Availability Architecture with Exadata Database Machine

ORACLE®

Table of Contents

.

Overview

The integration of Oracle Maximum Availability Architecture (Oracle MAA) operational and configuration best practices with Oracle Exadata Database Machine (Exadata MAA) provides the most comprehensive high availability solution for the Oracle Database.

Exadata Database Machine is a mature, integrated system of software, servers, storage and networking, all pre-configured according to Oracle MAA best practices to provide the highest database and application availability and performance. Mission critical applications in all industries and across both public and private sectors rely upon Exadata MAA. Every Exadata system – integrated hardware and software - has gone through extensive availability testing both internal to Oracle and by mission critical customers worldwide. The lessons learned from the experiences of this global community are channeled back into further enhancements that benefit every Exadata deployment.

This paper is intended for a technical audience: database, system and storage administrators and enterprise architects, to provide insight into Exadata MAA best practices for rapid deployment and efficient operation of Exadata Database Machine. The paper is divided into four main areas:

» Exadata MAA Architecture
» Initial Deployment
» Inherent Exadata HA Benefits
» Post Deployment: Exadata MAA Configuration
» Operational Best Practices for Exadata MAA

Exadata MAA best practices documented in this white paper are complemented by the following:

» My Oracle Support Note 757552.1 is frequently updated by Oracle development to provide customers the latest information gained from continuous MAA validation testing and production deployments.
» Exadata healthcheck (exachk) and its associated Oracle Exadata Assessment Report and MAA score card. This tool is updated quarterly.
» Additional MAA best practice papers that provide a deeper-dive into specific technical aspects of a particular area or topic published at www.oracle.com/goto/maa.

Exadata MAA Architecture

Exadata MAA architecture shown in Figure 1 is designed to tolerate unplanned outages of all types, providing both high availability (HA) and data protection. Exadata MAA also minimizes planned downtime by performing maintenance either online or in a rolling fashion. The range of potential outages and planned maintenance addressed by Exadata MAA are described in Appendix 1 of this paper: Exadata MAA Outage and Solution Matrix. For real world examples of how Exadata achieves end-to-end application availability and near zero brownout for various hardware and software outages, view the failure testing demonstrated in this Exadata MAA technical video[1]. Also refer to Oracle MAA Reference Architectures that provide blueprints that align with a range of availability and data protection requirements.
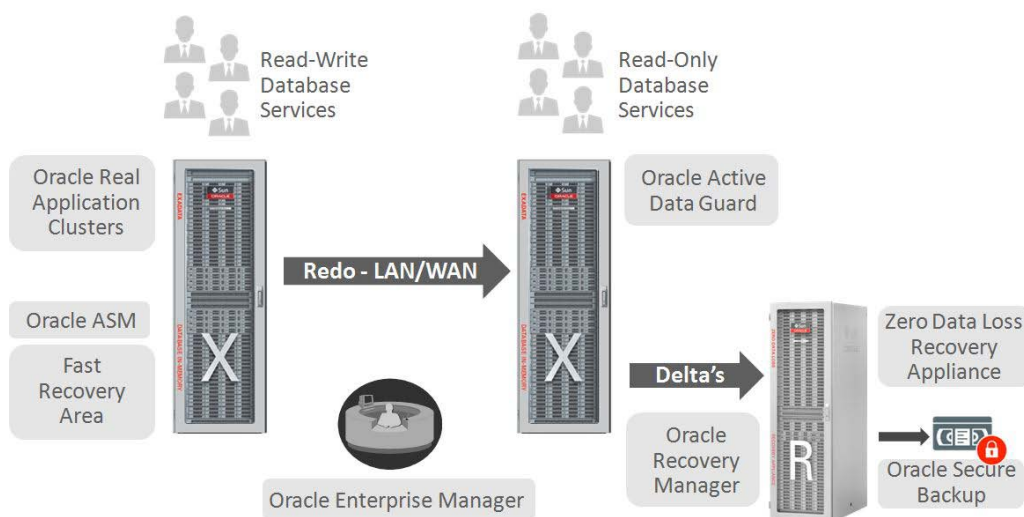

Figure 1. Basic Oracle Exadata Database Machine Configuration

Exadata MAA architecture "Gold Reference Architecture" consists of the following major building blocks:

» A production Exadata system (primary). The production system may consist of one Exadata elastic configuration or one or more interconnected Exadata Database Machines as needed to address performance and scale-out requirements for data warehouse, OLTP, or consolidated database environments.

» A standby Exadata system that is a replica of the primary. Oracle Data Guard is used to maintain synchronized standby databases that are exact, physical replicas of production databases hosted on the primary system. This provides optimal data protection and high availability if an unplanned outage makes the primary system unavailable. A standby Exadata system is most often located in a different data center or geography to provide disaster recovery (DR) by isolating the standby from primary site failures. Configuring the standby system with identical capacity as the primary also guarantees that performance service-level agreements can be met after a switchover or failover operation.

While the term 'standby' is used to describe a database where Data Guard maintains synchronization with a primary database, standby databases are not idle while they are in standby role. High return on investment is achieved by utilizing the standby database for purposes in addition to high availability, data protection, and disaster recovery. These include:

---

1 http://vimeo.com/esgmedia/exadata-maa-tests

» Oracle Active Data Guard enables users to move read-only queries, reporting, and fast incremental backups from the primary database, and run them on a physical standby database instead. This improves performance for all workloads by bringing the standby online as a production system in its own right. Oracle Active Data Guard also improves availability by performing automatic repair should a corrupt data block be detected at either the primary or standby database, transparent to the user. New Active Data Guard capabilities with Oracle Database 12c enable zero data loss failover to a remote location without impacting production database performance and new automation that greatly simplifies database rolling upgrades.

» Data Guard Snapshot Standby enables standby databases on the secondary system to be used for final pre-production testing while they also provide disaster protection. Oracle Real Application Testing can be used in conjunction with Snapshot Standby to capture actual production workload on the primary and replay on the standby database. This creates the ideal test scenario, a replica of the production system that uses real production workload – enabling thorough testing at production scale.

» Data Guard Standby-First patching (My Oracle Support Note 1265700.1) or Data Guard Database Rolling Upgrades are two methods of reducing downtime and risk during periods of planned maintenance. This is a key element of Exadata MAA Operational Best Practices discussed later in this paper.

Note that Data Guard is able to support up to 30 standby databases in a single configuration. An increasing number of customers use this flexibility to deploy both a local Data Guard standby for HA and a remote Data Guard standby for DR. A local Data Guard standby database complements the internal HA features Exadata by providing an additional layer of HA should unexpected events or human error make the production database unavailable even though the primary site is still operational. Low network latency enables synchronous replication to a local standby resulting in zero data loss if a failover is required and fast redirection of application clients to the new primary database

» A development/test Exadata system that is independent of the primary and standby Exadata systems. This system will host a number of development/test databases used to support production applications. The test system may even have its own standby system to create a test configuration that is a complete mirror of production. Ideally the test system is configured similar to the production system to enable:

  » Use of a workload framework (e.g. Real Application Testing) that can mimic the production workload.

  » Validation of changes in the test environment, including evaluating the impact of the change and the fallback procedure, before introducing any change to the production environment.

  » Validation of operational and recovery best practices.

With Exadata 12.1.2.1.0 and higher, Exadata also supports space-efficient database snapshots that can be used to create test and development environments.

Some users will try to reduce cost by consolidating these activities on their standby Exadata system. This is a business decision that trade-offs cost, operational simplicity and flexibility. In the case where the standby Exadata is also used to host other development and test databases, additional measures may be required at failover time to conserve system resources for production needs. For example, non-critical test and development activities may have to be deferred until failed system is repaired and back in production.

The Exadata MAA architecture provides the foundation needed to achieve high availability. Equally important are configuration and operational practices described in the sections that follow.

Initial Deployment

Exadata is a pre-optimized, pre-configured, integrated system of software, servers, and storage that comes ready-built to implement Exadata MAA. This section focuses on what is pre-configured and the HA best practices that apply to initial deployment.

Prior to delivery of Exadata it is important to run Oracle Exadata Deployment Assistant (OEDA) configuration tool (config.sh), enter configuration settings specific to your environment, generate the configuration files, and review the Installation Template in accordance with your data center and operational standards. Work with your network and database administrators to determine the proper network and database configuration settings to enter in OEDA configuration tool.

Oracle recommends that the following configurable deployment options be specified when completing the OEDA configuration tool:

1. On the Define Customer Networks screen, select Bonded for the Client network.
   » When channel bonding is configured for the client access network during initial configuration, the Linux bonding module is configured for active-backup mode (mode=1). If a different bonding policy is preferred, then you may reconfigure the bonding module after initial configuration. For additional details, refer to Oracle® Exadata Database Machine Installation and Configuration Guide.

2. On the Cluster configuration screen , in the Disk Group Details section, MAA recommends choosing Oracle Automatic Storage Management (ASM) HIGH redundancy for ALL (DATA and RECO) disk groups for best protection and highest operational simplicity from the following storage outage scenarios:
   » Double partner disk failure[2]: Protection against loss of the cluster and ASM disk group due to a disk failure followed by a second disk failure of a partner disk.
   » Disk failure when its partnered Exadata Storage Server is offline: Protection against loss of the cluster and ASM disk group when a storage server is offline and one of the storage server's partner disks fails. The storage server may be offline because of planned maintenance, such as rolling storage server patching, or because of a storage server failure.
   » Disk failure followed by disk sector corruption: Protection against data loss when latent disk sector corruptions exist and a partner storage disk is unavailable either due to planned maintenance or disk failure.

   Refer to the Exadata Storage Server Software User's Guide section "About Oracle ASM for Maximum Availability" and note that high redundancy disk groups require a minimum of three storage cells and, starting with Exadata Software 12.1.2.3.0, Oracle voting files can be placed in a high redundancy disk group with less than 5 storage servers by creating two extra quorum disks, each on a separate database server. Also in the case of disk failure or cell storage failure, ASM will automatically rebalance to restore redundancy if sufficient free space is available. An Exadata health check with exachk (described later) has a configuration check for minimum ASM free space.

   » Note: the default Exadata installation recommendation is DATA high redundancy and RECO normal redundancy. This configuration provides the best data protection and redundancy for your essential database files that reside in DATA and still provides plenty of usable capacity. The trade off is that additional operational steps are required to maintain high availability when the RECO disk group is lost or inaccessible. Refer to Configuration Prerequisites and Operational Steps for Higher Availability for a RECO disk group or Fast Recovery Area Failure (Doc ID 2059780.1).

3. On the Alerting screen, select Enable Email Alerting and/or Enable SNMP Alerting to receive storage server and database server alerts via SMTP, SNMP, or both.

4. On the Oracle Configuration Manager screen, select Enable Oracle Configuration Manager to collect configuration information and upload it to the Oracle repository. This enables faster analysis by Oracle Support Services when troubleshooting problems.

5. On the Auto Service Request screen, select Enable Auto Service Request (ASR) to automatically open service requests when specific Oracle Exadata Rack hardware faults occur..

---

2 Depending on the ASM redundancy level, each disk has a partner disk found in a separate Exadata storage cell. If a write or read errors occurs on one disk, then ASM can consult the Partnership Status Table (PST) to see whether any of the disk's partners are online. If too many partners are offline, ASM forces the dismounting of the disk group. Refer to *Oracle Automatic Storage Management Administrator's Guide* for more information about disk partnering.

6. On the Grid Control Agent screen, select Enable Oracle Enterprise Manager Grid Control Agent information on the Grid Control Agent page.

OEDA configuration tool creates the configuration files used with the OEDA deployment tool (install.sh). Upon delivery an Oracle systems engineer will validate the hardware and run a final check for any components that may have been damaged while in-transit. Next, Oracle Advanced Support Services or Consulting Services will execute the OEDA deployment tool using the supplied configuration files.

Within a few days after hardware arrival the system and database will be fully functional with:

» Validated hardware, InfiniBand network, and Exadata storage cells that are available and performing according to specifications.

» Recommended and supported operating system, firmware, and Oracle software and patches as described in My Oracle Support Note 888828.1. Oracle Advanced Support Services or Consulting Services may apply additional fixes as described in My Oracle Support Note 1270094.1 or there may be a slight deviation if the support note has recently changed.

» An Exadata Storage Grid consisting of DATA, RECO, and DBFS ASM disk groups. DATA and RECO disk groups are configured using the ASM redundancy option(s) previously selected through OEDA. More disk groups may be created if Exadata VM clusters are created.

» Network infrastructure that includes InfiniBand networks for all communication between database and Exadata storage servers. The network is configured for both client and management access.

» Oracle Real Application Clusters (Oracle RAC) database and Grid Infrastructure preconfigured using Exadata MAA configuration settings described in My Oracle Support Note 757552.1. If deploying Exadata Virtual Systems (Exadata OVM), each OVM RAC cluster (user domains) is under unique management domains. OEDA ensures this by default.   If you are creating new Exadata OVM clusters post deployment, refer to the Exadata Maintenance guide for additional instructions.

» Email alerts for storage server and database server alerts, Oracle Enterprise Manager Grid Control (initial configuration files), Automatic Service Request and Oracle Configuration Manager base setup will be automatically configured on the Exadata Database Machine if options checked as part of the OEDA configuration tool.

HA Benefits Inherent to Exadata

Exadata is engineered and preconfigured to enable and achieve end-to-end application and database availability with every hardware fault such FANs, PDUs, batteries, switch, disk, flash, database server, motherboards, and DIMMs for example. Extensive engineering and integration testing validates every aspect of the system, including hundreds of integrated HA tests performed on a daily basis. The HA characteristics inherent in Exadata are described in the following sections.

Hardware Components

The following hardware and component redundancy is common to all models of Exadata: X5-2, X4-2, X3-2, X3-8, X2-2, X4-8, X3-8 and X2-8 and future Exadata generations.

Redundant database servers

Exadata arrives at a customer site with multiple preconfigured industry-standard Oracle Database servers running Oracle Database 12c Release 1 (12.1) or Oracle Database 11g Release 2 (11.2) and Oracle RAC. Oracle engineering and testing teams ensure the firmware, software, and hardware configuration is tuned and pre-configured to provide high availability and scalability. Database servers are clustered, and they communicate with

each server using the high bandwidth, low latency InfiniBand network. With this configuration, applications can tolerate a database server or Oracle RAC instance failure with minimal impact.

Traditionally a typical database node eviction caused by a database node failure will result in waiting on CSS misscount (defaulted to 30 or 60 seconds in most systems) before even declaring a database node is dead. During that time the entire cluster freezes and there's an application brownout. With Grid Infrastructure 12.1.0.2 BP7 and higher and only on Exadata, the InfiniBand (IB) subnet manager and IB network are leveraged for an ultra fast and safe node eviction to reduce brownout to 2 seconds or less.

In the test results shown in Figure 2 there was no application brownout during the Exadata node failure with a 4 node RAC Exadata configuration but application throughput reduced and response time increased for 2 seconds. On non-Exadata systems, customers will observer 30 or 60 seconds of application blackouts.

Furthermore, with Exadata high bandwidth and low latency flash and storage GRID, customers can tune database initialization parameter FAST_START_MTTR_TARGET more aggressively reducing application brownout even further for instance and node failures overall. For any database parameter changes, it is still recommended to evaluate the performance impact on comparable test system prior to making production change.

Database Node Power Failure with a Read Mostly Workload and CSS Misscount=60
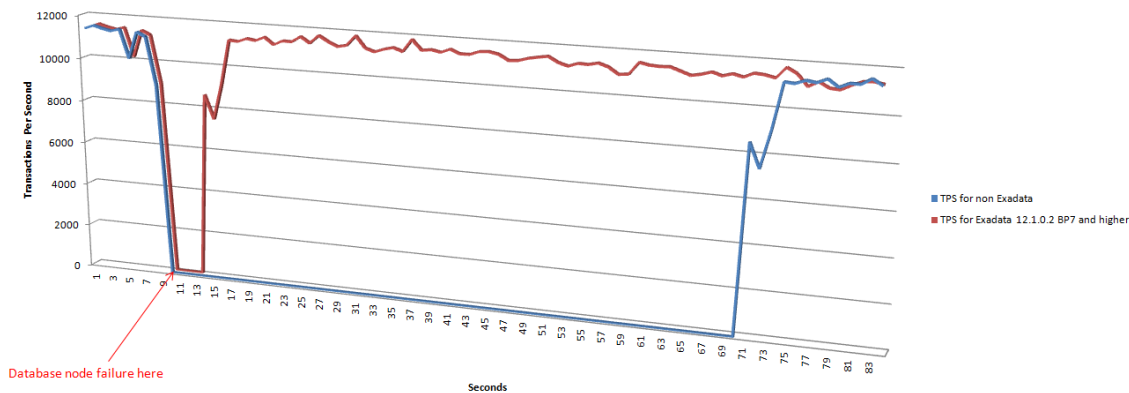
Figure 2: Database Node Power Failure

Redundant storage

Exadata storage components – database server disk drives, Exadata Storage Server disk drives, and Oracle Exadata Storage Servers (Exadata cell) are all redundant. Exadata Storage Servers are managed with ASM and configured to tolerate hard disk, flash disk, flash card, and complete storage server failures. Exadata Storage Servers are network-accessible storage devices with Oracle Exadata Storage Server Software pre-installed. Database data blocks and metadata are mirrored across cells to ensure that the failure of an Exadata disk or Exadata cell does not result in loss of data or availability. Disk drives are hot pluggable.

Exadata storage hardware and software have been engineered for the lowest application brownout for storage failures and provide extensive data protection with Exadata HARD and Exadata disk scrubbing. Compared to other traditional storage failures on other platforms, Exadata's application impact for disk, flash or storage server failure is

significantly lower. For example, Exadata storage failure can have less than 1 second application blackout and brownout versus seconds and minutes with other storage running Oracle databases and applications.

Published results from Oracle and EMC
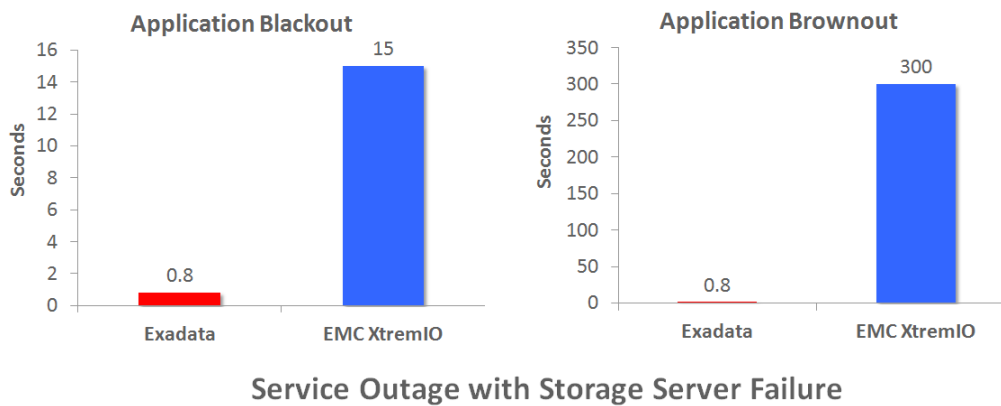
Figure 3. Storage Failure

Redundant connectivity

Redundant InfiniBand network connectivity using dual-ported Quad Data Rate (QDR) Host Channel Adapters (HCAs) and redundant switches are pre-configured. Configuring network redundancy for client access to database servers using Linux channel bonding is recommended and can be done at deployment time.

For network type failures within an Exadata system, the observed application brownout typically ranges from zero to single digit seconds.

Redundant power supply

Exadata has redundant power distribution units (PDUs) for high availability. The PDUs accept separate power sources and provide a redundant power supply to:

» Oracle Database nodes
» Exadata Storage Cells
» InfiniBand switches
» Cisco network switch

Power supplies for Oracle Database nodes, Exadata Storage Cells, InfiniBand and Cisco switches are all hot swappable.

Software Components:

The following are standard Oracle software components explicitly optimized and validated for Exadata Database Machine.

Firmware and Operating System

All database and Exadata storage servers are packaged with validated firmware and operating system software preinstalled.

Database Server Tier

Grid Infrastructure (Oracle Clusterware and ASM) and Oracle RAC software are installed and patched to recommended software version at deployment, enabling applications to tolerate and react to instance and node failures automatically with zero to near-zero application brownout. As described in Appendix 1, all Grid Infrastructure patches and most database patches can be applied in rolling fashion.

Storage Tier

» Tolerating hard disk, flash disk, flash card and Exadata cell failures

» Applying software changes in a rolling manner

» Exadata storage cells include Oracle Hardware Assisted Resilient Data (HARD) to provide a unique level of validation for Oracle block data structures such as data block address, checksum and magic numbers prior to allowing a write to physical disks. HARD validation with Exadata is automatic (setting DB_BLOCK_CHECKSUM is required to enable checksum validation). The HARD checks transparently handle all cases including ASM disk rebalance operations and disk failures.

High Performance

Oracle Development teams who focus on high performance for OLTP and Data Warehouse applications have optimized the configuration defaults set for Exadata. In some cases, there will be different default settings for different generations of Exadata systems. These settings are the result of extensive performance testing with various workloads, both in Oracle labs and in production deployments.

Validated configuration defaults have also undergone fault injection testing in a full MAA environment that includes Oracle RAC, ASM, RMAN, Flashback, and Data Guard.

Additional Exadata HA Features and Benefits

Refer to Table 1 for an overview of Exadata specific HA features and benefits. For a more detailed description of these capabilities please refer to Exadata documentation such as Oracle Exadata Database Machine System Overview, Exadata Machine Maintenance Guide and Exadata Storage Server Software User's Guide.

TABLE 1: HA FEATURES AND BENEFITS

AREA	FEATURE	HA BENFITS	DEPENDENTCIES
REDUCED HA BROWNOUT	Fast node detection and failover	Reduced node failure detection from as many as 60 seconds to just 2 seconds or less.	Grid Infrastructure 12.1.0.2 BP7 or higher
	Automatic detection of Exadata storage failures with low application impact	Automatic detection and rebalance. Application impact from 1 to 2 seconds delay	Continual improvements in each Exadata software release.
	Automatic detection of Exadata network failures with low application impact	Automatic detection and failover Application impact from 0 to 5 seconds delay	Continual improvements in each Exadata software release
	Reduce brownout for instance failures	With Exadata high bandwidth and low latency flash and storage GRID, customers can tune database initialization parameter, FAST_START_MTTR_TARGET, more aggressively without possible impact to the application reducing application brownout even further for instance and node failures.	Continual improvements in each Exadata database software release

AREA	FEATURE	HA BENFITS	DEPENDENTCIES
	Full high redundancy advantages for Oracle files and clusterware voting files with minimum of 3 storage cells	Oracle voting files can be placed in a high redundancy disk group with less than 5 storage server enabling all the data protection and redundancy benefits for both Oracle database and Oracle cluster. This will be done automatically through Oracle Exadata Deployment if you chose to create a high redundancy disk group.	Exadata 12.1.2.3.0 or higher
DATA PROTECTION	Automatic Hard Disk Scrub and Repair	Automatically inspects and repairs hard disks periodically when hard disks are idle. If bad sectors are detected on a hard disk, then Exadata automatically sends a request to ASM to repair the bad sectors by reading the data from another mirror copy. By default, the hard disk scrub runs every two weeks.	Database and GI 11.2. or 12c Exadata 11.2.3.3 or higher
		With Adaptive Scrubbing the frequency of scrubbing a disk may change automatically if bad sectors are discovered. If a bad sector is found on a hard disk in a current scrubbing job, Oracle Exadata Storage Server Software will schedule a follow-up scrubbing job for that disk in one week. When no bad sectors are found in a scrubbing job for that disk, the schedule will fall back to the scrubbing schedule specified by the $hardDiskScrubInterval$ attribute.	Exadata 12.1.2.3.0 or higher
	Exadata H.A.R.D.	Exadata Hardware Assisted Resilient Data (HARD) provides a unique level of validation for Oracle block data structures such as data block address, checksum and magic numbers prior to allowing a write to physical disks. HARD validation with Exadata is automatic. The HARD checks transparently handle all cases including ASM disk rebalance operations and disk failures	DB_BLOCK_CHEKSUM = TYPICAL or TRUE to enable all the Exadata HARD checks.
QUALITY OF SERVICE	I/O Latency Capping for Read Operations	Redirects read I/O operations to another cell when the latency of the read I/O is much longer than expected. This addresses the hung or very slow read IO cases due to device driver, controller, or firmware issues or failing or dying disks, flash or bad storage sectors.	Exadata 11.2.3.3.1 or higher Database and GI 11.2.0.4 BP8 or higher
	I/O Latency Capping for Write Operations	Redirects high latency write I/O operations to another healthy flash device. This addresses the hung or very slow write IO cases.	Exadata 12.1.2.1.0 or higher Database and GI 11.2.0.4 BP8 or higher Write-back flash cache enabled
	Exadata Cell I/O Timeout Threshold	Ability to set I/O timeout threshold that allows for long running I/O to be canceled and redirected to a valid mirror copy.	Exadata 11.2.3.3.1 or higher Database and GI 11.2.0.4 BP8 or higher
	Health Factor for Predictive Failed Disk Drop	When a hard disk enters predictive failure on Exadata Cell, Exadata automatically triggers an ASM rebalance to relocate data from the disk. The ASM rebalances first reads from healthy mirrors to restore redundancy. If all other mirrors are not available, then ASM rebalance reads the data from the predictively-failed disk. This diverts rebalance reads away from the predictively-failed disk when possible to ensure optimal rebalance progress while maintaining maximum data redundancy during the rebalance process.	Exadata storage 11.2.3.3 or higher

AREA	FEATURE	HA BENFITS	DEPENDENTCIES
	Identification of Underperforming disks and Automatic Removal	Underperforming disks affect the performance of all disks because work is distributed equally to all disks. When an underperforming disk is detected, it is removed from the active configuration. Exadata performs internal performance tests. If the problem with the disk is temporary and it passes the tests, then it is brought back into the configuration. If the disk does not pass the tests, then it is marked as poor performance, and an Auto Service Request (ASR) service request is opened to replace the disk. This feature applies to both hard disks and flash disks.	Exadata storage 11.2.3.2 or higher
	I/O Resource Management	I/O Resource Management (IORM) manages disk and flash IOPS and flash cache minimum and maximum flash cache size per pluggable database or physical databases.	For I/O Resource Management for Flash and Flash Cache space resource management: Exadata Storage 12.1.2.1.0 or higher and Exadata X2 generation or higher hardware
	Network Resource Management	Network Resource Management automatically and transparently prioritizes critical database network messages through the InfiniBand fabric ensuring fast response times for latency critical operations. Prioritization is implemented in the database, database InfiniBand adapters, Exadata Software, Exadata InfiniBand adapters, and InfiniBand switches to ensure prioritization happens through the entire InfiniBand fabric. Latency sensitive messages such as Oracle RAC Cache Fusion messages are prioritized over batch, reporting, and backup messages. Log file write operations are given the highest priority to ensure low latency for transaction processing.	Exadata Storage 11.2.3.3 Oracle Database 11.2.0.4 or higher IB switch firmware release 2.1.3-4 or higher
	Cell-to-Cell Rebalance Preserves Flash Cache Population	When a hard disk hits a predictive failure or true failure, and data needs to be rebalanced out of it, some of the data that resides on this hard disk might have been cached on the flash disk, providing better latency and bandwidth accesses for this data. To maintain an application's current performance SLA, it is critical to rebalance the data while honoring the caching status of the different regions on the hard disk during the cell-to-cell offloaded rebalance. The cell-to-cell rebalance feature provides significant performance improvement compared to earlier releases for application performance during a rebalance due to disk failure or disk replacement.	Exadata Storage 12.1.2.2.0 or higher Database and GI 12.1.0.2 BP11 or higher
PERFORMANCE	Exadata Smart Flash Logging	Exadata smart flash logging ensures low latency redo writes which is crucial to database performance especially OLTP workloads. This is achieved by writing redo to both hard disk and flash where the flash is used as a temporary store (cache) for redo log data to maintain	Exadata storage 11.2.2.4 and higher EF is only available for Exadata

AREA	FEATURE	HA BENFITS	DEPENDENTCIES
		consistently low latency writes and avoid expensive write outliers. Exadata smart flash logging is also needed for Extreme Flash (EF) configuration since flash devices can occasionally be slow. To avoid outliers for EF, redo writes are very selective in choosing and writing to multiple flash drives.	X5 generations or higher.
	Active Bonding Network	Exadata servers can be configured with active bonding for both ports of InfiniBand card. Active bonding provides much higher network bandwidth when compared to active passive bonding in earlier releases because both InfiniBand ports are simultaneously used for sending network traffic.	Exadata X4 generation or higher hardware Exadata storage 11.2.3.3 or higher
	Exadata Smart Write Back Flash Cache Persistent After Cell Restarts	Exadata Smart Flash Cache transparently and intelligently caches frequently-accessed data to fast solid-state storage, improving database query and write response times and throughput. If there is a problem with the flash cache, then the operations transparently fail over to the mirrored copies on flash. No user intervention is required. Exadata Smart Flash Cache is persistent through power outages, shutdown operations, cell restarts, and so on. Data in flash cache is not repopulated by reading from the disk after a cell restarts. Write operations from the server go directly to flash cache. This reduces the number of database I/O operations on the disks.	Exadata storage 11.2.3.2 or higher
	Data Guard Redo Apply Performance increase of 6+ X	Data Guard redo apply performance takes advantage of Exadata smart flash cache and overall I/O and network bandwidth enabling observed redo apply rates of up to 300 MB/sec for OLTP workloads and up to 800 MB/sec for batch and load workloads. Traditional storage tends to bottlenecked with network or storage IO bandwidth restricting redo apply performance typically below 50 MB/sec.	Rates observed in-house MAA testing and with real world customers. Rates may vary depending on the amount of database consolidation and available system bandwidth.
MANAGEMENT	Patching of Exadata Storage Cells, Exadata database nodes, and InfiniBand Switches	Patchmgr utility (and dbnodeupdate.sh) provides patching orchestration and automation for patching Exadata Storage Cells, Exadata database nodes and InfiniBand switches for both online and offline options.	Patchmgr supports Exadata Storage Cells Dbnodeupdate.sh supports Exadata Database Nodes Patchmgr extended to support InfiniBand Switches with Oracle Exadata Storage 11.2.3.3.0 and higher Patchmgr extended to support Exadata database node patching orchestration with Exadata Storage 12.1.2.2.0 software
	Performance improvements for Storage Server Software Updates	Updating Oracle Exadata Storage Server Software now takes significantly less time. By optimizing internal processing even further, the cell update process is now up to 2.5 times faster compared to previous releases. Even though most Exadata patching occurs with the application online, this enhancement dramatically reduces the patching window.	Oracle Exadata Storage Server Software release 12.1.2.3.0 and higher

AREA	FEATURE	HA BENFITS	DEPENDENTCIES
	Flash and Disk Life Cycle Management Alerts	Monitors ASM rebalance operations due to disk failure and replacement. Management Server sends an alert when a rebalance operation completes successfully, or encounters an error. Simplify status management.	Oracle Database release 12.1.0.2 BP4 or later, Oracle Exadata Storage Server Software release 12.1.2.1.0 or higher.
	Cell Alert Summary	Oracle Exadata Storage Server Software periodically sends out an e-mail summary of all open alerts on Exadata Cells. The open alerts e-mail message provides a concise summary of all open issues on a cell.	Oracle Exadata Storage Server Software release 11.2.3.3.0 or higher
	LED Notification for Storage Server Disk Removal	When a storage server disk needs to be removed, a blue LED light is displayed on the server. The blue light makes it easier to determine which server disk needs maintenance.	Oracle Exadata Storage Server Software release 11.2.3.2.0 or higher
	Drop Hard Disk for Replacement	Simple command for an administrator to remove hard disk from Exadata cell. The command checks to ensure that the grid disks on that hard disk can be safely taken offline from ASM without causing a disk group force dismount. If it is successful, service LED on the disk will be turned on for easy replacement	Oracle Exadata Storage Server Software release 11.2.3.3.0 or higher
	Drop BBU for Replacement	Simple command for an administrator to initiate an online BBU (battery backup unit) replacement. The command changes the controller to write-through caching and ensures that no data loss can occur when the BBU is replaced in case of a power loss.	Exadata X3 and X4 generations only. Exadata X5s disk controller HBAs come with 1 GB supercap-backed write cache instead of BBU. Oracle Exadata Storage Server Software release 11.2.3.3.0 or higher
	Minimize or eliminates false disk failures	I/Os are automatically redirected to healthy drives. The targeted unhealthy disk is power cycled. If the drive returns to normal status, then it will be re-enabled and resynchronized. If the drive continues to fail after being power cycled, then it will be dropped. Eliminates false-positive disk failures and helps preserve data redundancy, reduce operational management and avoids drop rebalance.	X5 Storage or higher since power cycle support required in chassis Only relevant for High Capacity Hard Disks and Extreme Flash SSDs
	Exadata AWR and Active Report	The Exadata Flash Cache Performance Statistics sections have been enhanced in the AWR report: 1) Added support for Columnar Flash Cache and Keep Cache. 2) Added a section on Flash Cache Performance Summary to summarize Exadata storage cell statistics along with database statistics. The Exadata Flash Log Statistics section in the AWR report now includes statistics for first writes to disk and flash.	Oracle Exadata Storage Server Software release 12.1.2.2.0 or higher Oracle Database release 12.1.0.2 Bundle Patch 11or later,

Post Deployment – Exadata MAA Configuration

Oracle strongly recommends that you DO NOT use initialization settings from previous Oracle configurations. Begin with the default configuration delivered with Exadata and then tune if required. Most customers will find minimal need for tuning database parameters in the post deployment phase except for SGA related settings. If more databases are added for consolidation, system resources may have to be adjusted as described in the MAA papers: MAA Best Practices for Database Consolidation and Oracle Multitenant with Oracle 12c[3] or Best Practices For Database Consolidation On Oracle Exadata Database Machine for 11g[4].

The following sections provide an overview of MAA configuration best practices. Refer to Oracle documentation for detailed MAA best practices for Oracle Database 11g[5] or Oracle Database 12c[6]. The sections below provide a high level review of the details included in the best practice documentation.

Database Archivelog Mode and Enable Database Force Logging

Archivelog Mode and Database Force Logging are prerequisites to insure all changes are captured in the redo and applied during database recovery operation. Enable database archive logging and database force logging to prevent any nologging operations.

```
SQL> ALTER DATABASE ARCHIVELOG;

SQL> ALTER DATABASE FORCE LOGGING;
```

Optionally you may set logging attributes at the tablespace level if it is possible to isolate data that does not require recovery into a separate tablespace. For more information refer to *"Reduce Overhead and Redo Volume During ETL Operations"* in the technical white paper, Oracle Data Guard: Disaster Recovery for Oracle Exadata Database Machine[7].

Fast Recovery Area

The Fast Recovery Area (FRA) is Oracle-managed disk space that provides a centralized disk location for backup and recovery files.

The FRA is defined by setting two database initialization parameters.

The `DB_RECOVERY_FILE_DEST` parameter specifies the location for the Fast Recovery Area. Use the RECO disk group, and MAA recommends leveraging Exadata storage grid instead of external storage for the best performance.

The `DB_RECOVERY_FILE_DEST_SIZE` parameter specifies (in bytes) a hard limit on the total space to be used by database recovery files created in the recovery area location.

Oracle Flashback Technologies

Oracle Flashback technologies provide fast point-in-time recovery to repair logical database corruptions that are most often caused by human error. A suite of Oracle Flashback technologies enable repair at an optimal level of granularity to achieve the fastest possible return to service with a minimum of data loss. The sections below provide an overview of best practices for each Flashback technology. Refer to Oracle Database High Availability

---

3 http://www.oracle.com/technetwork/database/availability/maa-consolidation-2186395.pdf

4 http://www.oracle.com/technetwork/database/features/availability/exadata-consolidation-522500.pdf

5 http://docs.oracle.com/cd/E11882_01/server.112/e10803/toc.htm

6 http://docs.oracle.com/database/121/HABPT/toc.htm

7 http://www.oracle.com/technetwork/database/features/availability/maa-wp-dr-dbm-130065.pdf

documentation for additional details to <u>recover from human error using Oracle Database 11g</u>[8] or <u>Oracle Database 12c</u>[9].

Flashback Database

Flashback Database uses flashback logs to 'rewind' an entire Oracle Database to a previous point in time. Flashback Database is used for fast point-in-time recovery from logical corruptions, most often caused by human error, that cause widespread damage to a production database. It is also used to quickly reinstate a failed primary database as a new standby database after a Data Guard failover, avoiding the costly and time-consuming effort of restoring from a backup. See *Data Guard Concepts and Administration* for more information on <u>fast reinstatement using Flashback Database using Oracle Database 11g</u>[10] or <u>Oracle Database 12c</u>[11]. Customers are also creating Oracle Flashback Database Guaranteed Restore Points prior to major changes including software changes for a quick fallback.

Issue the following command on both primary and standby databases to enable Flashback Database.

```
ALTER DATABASE FLASHBACK ON;
```

Oracle Database 11.2.0.3 and higher releases includes substantial performance enhancements that reduce the impact of flashback logging on the primary database, and in particular its impact on load operations and initial flashback allocations. This enables Flashback Database to be used with OLTP and data warehouse applications with minimal performance impact. Oracle MAA tests have achieved a load rate of 3 TB/hour using Oracle Database Release 11.2.0.2 and higher with Flashback Database enabled. This rate is possible due to the large I/O bandwidth of Exadata, 11.2.0.2 optimizations and MAA configuration best practices:

» Flashback Database best practices documented in <u>My Oracle Support Note 565535.1</u>.
» Use of local extent-managed tablespaces.
» Re-creation of objects instead of truncating tables prior to direct load.

Additional Oracle Flashback Technologies for More Granular Repair

More granular levels of repair are enabled by other Oracle Flashback technologies that include: Flashback Query, Flashback Version Query, Flashback Transaction, Flashback Transaction Query, Flashback Table, and Flashback Drop.

Flashback Drop requires configuration of a Recycle Bin. All of the other features use automatic undo management and require that you allocate sufficient disk space to achieve your desired undo retention guarantee – or how far back in the past you want to be able to recover. To configure the desired undo retention guarantee:

» The undo retention period (in seconds) is specified by setting the UNDO_RETENTION initialization parameter to a value that is at least double (2x) the desired detection period to ensure Oracle Flashback operations can operate.

Note that if you are using Oracle Active Data Guard to offload read-only workload to a standby database, please see additional guidance on setting undo retention period described in the section titled 'Avoiding ORA-1555 Errors on page 16 of <u>Oracle Active Data Guard Best Practices</u>[12].

---

8 https://docs.oracle.com/cd/E11882_01/server.112/e10803/outage.htm#HABPT5043

9 http://docs.oracle.com/database/121/HABPT/outage.htm#HABPT5043

10 http://docs.oracle.com/cd/E11882_01/server.112/e41134/scenarios.htm#SBYDB00910

11 http://docs.oracle.com/database/121/SBYDB/scenarios.htm#SBYDB00910

12 http://www.oracle.com/technetwork/database/features/availability/maa-wp-11gr1-activedataguard-1-128199.pdf

Backup, Restore, and Recovery

Oracle Database backup technologies and appliances such as Oracle Zero Data Loss Recovery Appliance (Recovery Appliance) and ZFS Storage Appliance, can be leveraged to backup and restore databases from Exadata Database Machine.

The Recovery Appliance provides comprehensive backup validation and data protection for Oracle Databases running on any platform and is Oracle's strategic solution for database backup and recovery. Customers using the Recovery Appliance benefit from reduced backup windows and database impact by using incremental backups forever and offloading deduplication, compression and validation from the Exadata database host to the Recovery Appliance. Most importantly, the Recovery Appliance enables reliable and fast database restore to any-point-in-time within a recovery window specified by the administrator.

ZFS Storage Appliance is a lowest cost local disk backup and restore solution with its ability to backup and restore Oracle databases at 10+ TB/hour depending on the configuration.

Backup and restore directly to Exadata storage provides the highest backup/restore rates.

A series of MAA technical whitepapers provide additional details:

» Backup and Recovery Performance and Best Practices for Exadata Database Machine - Oracle Database 11.2.0.2 and Later[13]
» MAA Best Practices for the Zero Data Loss Recovery Appliance
» Backup and Recovery Performance and Best Practices using Sun ZFS Storage Appliance with Oracle Exadata Database Machine[14]
» Oracle Exadata Database Machine - Backup & Recovery Sizing: Tape Backups[15]

Oracle Data Guard and Oracle Active Data Guard

Oracle Data Guard is the disaster recovery solution prescribed by the Maximum Availability Architecture (MAA) to protect mission-critical databases residing on Exadata. Data Guard is also used to maintain availability should any outage unexpectedly impact the production database, and to minimize downtime during planned maintenance. Data Guard is included in the Enterprise Edition license for Oracle Database.

Oracle Active Data Guard provides the following extensions to basic Data Guard functionality (Oracle Active Data Guard requires an additional license to access its advanced features).

» A physical standby database may be open read-only while it applies updates received from the primary. This enables offload of read-only workload to a synchronized standby database, increasing capacity and improving the performance of the primary database. It is significant to note that read-only queries on an Active Data Guard standby database have the same guarantee of read-consistency as queries executing on the primary database – no physical replication solution supplied by any other DBMS vendor can provide this level of read-consistency.
» RMAN block-change tracking can be enabled on the physical standby database, enabling fast incremental backups to be offloaded from the primary database.
» Automatic block repair is enabled for the Active Data Guard configuration. Should Oracle detect a physical corrupted block on either the primary or standby, it will automatically repair the physical block corruption using the good copy from the other database, transparent to the application and user.

---

13 http://www.oracle.com/technetwork/database/features/availability/maa-tech-wp-sundbm-backup-11202-183503.pdf

14 http://www.oracle.com/technetwork/database/features/availability/maa-wp-dbm-zfs-backup-1593252.pdf

15 http://www.oracle.com/technetwork/database/availability/maa-exadata-backup-methodology-495297.pdf

Data Guard is the disaster recovery solution used by Exadata MAA because:

» It provides data protection superior to storage remote-mirroring[16].

» It is a highly efficient physical replication process that supports all Oracle data types and Oracle Database features, providing greater simplicity and performance logical replication.

» It provides a simple zero-risk method with minimal downtime for installing Oracle patches, including Oracle Exadata Database Machine bundled patches, Oracle Exadata Storage Server Software patches, Patch Set Updates (PSU), Critical Patch Updates (CPU), or Patch Set Exceptions (PSE) using Standby-First Patch Apply. Standby-First Patch Apply is supported for certified software patches for Oracle Database Enterprise Edition Release 2 (11.2) release 11.2.0.1 and later as described in My Oracle Support Note 1265700.1.

» It supports database rolling upgrades when upgrading to new Oracle Patch-Sets (e.g. 11.1.0.7 to 11.2.0.4 or 12c) or future Oracle Database Releases using Transient Logical Standby from Oracle Database 11g onward[17] or the simpler DBMS_ROLLING capability with Active Data Guard 12c onward[18].

» Oracle Active Data Guard provides high return on investment on standby systems while they are in standby role by enabling read-only workload to be offloaded from primary systems, increased availability using automatic block repair, and zero data loss protection across any distance using Active Data Guard Far Sync.

For the complete set of best practices for optimal data protection and availability when configuring Data Guard, see the MAA documentation for Oracle Database 11g[19] ,Oracle Database 12c[20] or the latest Data Guard and MAA best practices on http://www.oracle.com/technetwork/database/features/availability/oracle-database-maa-best-practices-155386.html

Comprehensive Data Corruption Protection

Oracle Database includes comprehensive database-aware capabilities to either prevent or automatically detect and repair data corruption that could otherwise lead to substantial downtime and data loss. Data corruption occurs when a data block is not in a valid Oracle Database format, or whose contents are not internally consistent. Corruption can result from either hardware or software defects.

Listed below are the Exadata MAA recommended initialization settings for optimal data corruption protection. The recommended settings place a greater emphasis on protection compared to the default settings that favor minimizing performance impact. Where different from the default value, the recommended initialization settings will need to be configured manually. Note that a Data Guard standby database is necessary for complete protection.

Oracle recommends that you read My Oracle Support Note 1302539.1 for additional background, for new corruption prevention and detection tips and for important insight into the potential performance tradeoff. This tradeoff requires that you conduct performance testing for your application workload prior to production go-live.

On the Primary Database

» DB_BLOCK_CHECKSUM=FULL (default TYPICAL on Exadata)

» DB_BLOCK_CHECKING=FULL (default OFF on Exadata)

» Some applications are not able to set DB_BLOCK_CHECKING=MEDIUM or FULL on the primary due to performance impact. MAA recommends that you set DB_BLOCK_CHECKING=MEDIUM or FULL on the physical standby as a minimum practice to prevent the standby from various logical block corruptions.

» DB_LOST_WRITE_PROTECT=TYPICAL (default TYPICAL on Exadata)

---

16 http://www.oracle.com/technetwork/database/availability/dataguardvsstoragemirroring-2082185.pdf

17 http://www.oracle.com/technetwork/database/features/availability/maa-wp-11g-upgrades-made-easy-131972.pdf

18 http://docs.oracle.com/database/121/SBYDB/dbms_rolling_upgrades.htm#SBYDB5214

19 http://docs.oracle.com/cd/E11882_01/server.112/e10803/config_dg.htm#CEGEADFC

20 http://docs.oracle.com/database/121/HABPT/config_dg.htm#CEGEADFC

» Enable Flashback Technologies for fast point-in-time recovery from logical corruptions that are most often caused by human error and for fast reinstatement of a primary database following failover.

On the Data Guard Physical Standby Database

» DB_BLOCK_CHECKSUM=FULL

» DB_BLOCK_CHECKING=OFF

» DB_LOST_WRITE_PROTECT=TYPICAL

» Setting DB_BLOCK_CHECKING=FULL provides maximum data corruption detection and prevention on the standby; however redo apply performance can be reduced significantly. Because redo apply performance is so fast, especially with Exadata's write back flash cache, most customers can set DB_BLOCK_CHECKING=FULL or MEDIUM on the standby and still maintain the required redo apply rates and availability service levels.

» Enable Flashback Technologies for fast point-in-time recovery from logical corruptions that are most often caused by human error and for fast reinstatement of a primary database following failover.

» Use Oracle Active Data Guard to enable Automatic Block Repair (Data Guard 11.2 onward).

» Use Oracle Recovery Appliance for additional backup validation.

Application Clients - Failover Best Practices

High application availability is only achieved if an application reacts and fails over transparently when an unplanned outage or planned maintenance activity impacts the availability of the production database. For details on automating the detection refer to Client Failover Best Practices for Highly Available Oracle Databases for Oracle Database 11g[21] or Oracle Database 12c[22].

Automated Validation of MAA Best Practices - exachk

The Exadata MAA health check (exachk), automates the process of ongoing validation of your configuration as Exadata best practices continue to evolve in the future. Exachk is updated quarterly with new configuration best practices, software recommendations and information on any new critical alerts. Oracle recommends that you download the latest version of exachk from My Oracle Support Note 1070954.1 and respond to any additional configuration, software, or hardware recommendations made by the the exachk report.

Additional Exadata OVM HA Configuration Considerations

Exadata OVM High Availability has the following additional considerations:

1. Setup PingTarget to enable VIP and client failover for node failures

» The parameter pingtarget should be defined for the Grid Infrastructure resource ora.net1.network and set to the IP address of the client network gateway. If your client network is not based on ora.net1.network, use ora.netX.network resource where X >1 and set the pingtarget accordingly.

» For OVM environments which does not have pingtarget appropriately defined, a complete client network failure in a particular RAC cluster node where both the slave interfaces of the bonded interface bondeth0 fail, the VIP resource will not failover to a surviving node resulting in client connections hanging for extended periods of time. With pingtarget set, the VIP fails over to a healthy cluster node and the client connections to the faulty node get an immediate connection reset.

---

21 http://www.oracle.com/technetwork/database/features/availability/maa-wp-11gr2-client-failover-173305.pdf

22 http://www.oracle.com/technetwork/database/availability/client-failover-2280805.pdf

» Set pingtarget with the following command:

```
<$GI_HOME>/bin/srvctl modify network -netnum 1 -pingtarget <IP of the gateway of
the client network>
```

» Verify pingtarget with the following command:

```
<$GI_HOME>/bin/srvctl config network
```

2. For optimal and consistent backup performance for Exadata Software prior to 12.1.2.2.0, set rmem and wmem buffer settings.

» To ensure RMAN backups work optimally and avoid packet losses, the rmem and wmem buffer settings need to be set to 4M in sysctl.conf as shown below:

```
net.core.rmem_default = 4194304
net.core.wmem_default = 4194304
net.core.rmem_max = 4194304
net.core.wmem_max = 4194304
```

3. For Exadata OVM setup and maintenance, refer to Exadata Database Machine Maintenance Guide

Database Consolidation Best Practices

Exadata is optimized for Oracle Data Warehouse and OLTP database workloads, and its balanced database server and storage GRID infrastructure make it an ideal platform for database consolidation. Oracle database, storage, and network GRID architectures combined with Oracle resource management provide a simpler and more flexible approach to database consolidation than other virtualization strategies (e.g. hardware or O.S. virtualization). Exadata also supports virtual machines if additional isolation is required.

Refer to the following MAA white papers for database consolidation best practices for Exadata:

» Oracle Maximum Availability Architecture Best Practices for Database Consolidation[23]
» Best Practices for Database Consolidation on Oracle Exadata Database Machine[24]
» Oracle Exadata Database Machine Consolidation: Segregating Databases and Roles[25]

Operational Best Practices for Exadata MAA

The following operational best practices are required for a successful Exadata implementation:

» Document your high availability and performance service-level agreements (SLAs) and create an outage/solution matrix that maps to your service level agreements.

Understanding the impact to the business and the resulting cost of downtime and data loss is fundamental to establishing Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO). RTO measures your tolerance for downtime while RPO measures your tolerance for data loss. It is also likely that RTO and RPO will be different for different classes of outages. For example, server and disk failures usually have RTO/RPO of zero. A complete site failure may have larger RTO/RPO as well as less stringent performance SLAs. This is due to managing the trade-off between the potential frequency of an outage occurring and the cost or complexity of implementing HA/DR.

23 http://www.oracle.com/technetwork/database/availability/maa-consolidation-2186395.pdf

24 http://www.oracle.com/technetwork/database/features/availability/exadata-consolidation-522500.pdf

25 http://www.oracle.com/technetwork/database/availability/maa-exadata-consolidated-roles-459605.pdf

The range of outages that must be planned for in an Exadata MAA environment are described in Appendix 1, "Exadata MAA Outage and Solution Matrix". It is important to document your procedures for detection and repair for each type of outage. In some cases you will take advantage of Oracle software infrastructure to configure an automatic response. In other cases, Oracle software infrastructure will provide automatic notification but resolution will occur via human intervention. In all cases is it required to validate that each detection and repair procedure is able to meet SLAs. A similar process is followed to implement performance monitoring and rapid response to ensure that performance SLAs, such as throughput and response time requirements, are also achieved.

» Validate HA and Performance SLAs. Perform simple database node, database instance, database failure and cell storage fault-injection testing to validate the expected HA response including all automatic, automated, or manual repair solutions. Ensure that the application RTO and RPO requirements are met. Ensure that application performance is acceptable under different scenarios of component failures. For example, does the application continue to meet performance SLAs after node failure, Exadata Storage cell failure, and Data Guard role transition?

» Periodically (e.g. at least once a year) upgrade Exadata and database software as recommended in My Oracle Support Note 888828.1.

Exadata will be delivered and deployed with the then current recommended HA software and system components. Once deployed it is necessary to periodically run exachk and refer to the Exadata software maintenance best practices section of the MAA scorecard to evaluate if your existing Exadata software is within recommended range. The software maintenance checks within exachk will alert you to any critical software issues that may be relevant to your environment (be sure to download the latest version of exachk before running).

Between exachk releases a new Exadata critical issue that requires prompt attention may be identified, resolved, and information about the issue published in My Oracle Support Note 1270094.1. To receive proactive notification of newly published Alerts for Exadata critical issues from My Oracle Support, configure Hot Topics E-Mail for product Oracle Exadata Storage Server Software.

Pre-production validation and testing of software patches is one of the most effective ways to maintain stability. The high-level steps are:

  » Review the patch and upgrade documentation.

  » Evaluate any rolling upgrade opportunities in order to minimize or eliminate planned downtime.

  » Evaluate whether the patch qualifies for Standby-First Patching, described in My Oracle Support Note 1265700.1.

  » For all grid infrastructure software home and database software home changes, use Out-of-Place Software installation and patching for easier fallback, described in section 14.2.4.3 "Out-of-place Software and Installation Patching"[26] in *Oracle Database High Availability Best Practices*.

  » Validate the application in a test environment and ensure the change meets or exceeds your functionality, performance, and availability requirements. Automate the procedure and be sure to also document and test a fallback procedure.

  » If applicable, perform final pre-production validation of all changes on a Data Guard standby database before applying them to a production system.

  » Apply the change in your production environment.

» Execute the Exadata MAA health check (exachk), as described in My Oracle Support Note 1070954.1. Before and after each software patch, before and after any database upgrade, or minimally every month, download the latest release of exachk and run it in your test and production environments to detect any environment and configuration issues. Checks include verifying the software and hardware and warning if any existing or new MAA, Oracle RAC, or Exadata hardware and software configuration best practices need to be implemented. An MAA score card has been added with MAA configuration checks and best practices. An upgrade module has been added to proactively detect any configuration issues pre and post database upgrade.

---

26 http://docs.oracle.com/cd/E11882_01/server.112/e10803/schedule_outage.htm#sthref947

» Execute Data Guard role transitions and validate restore and recovery operations. Periodically execute Application and Data Guard switchovers to fully validate all role transition procedures. We recommend conducting role transition testing a minimum of once per quarter. In addition to Data Guard documentation, refer to Data Guard role transition best practices in MAA OTN website such as Role Transition Best Practices: Data Guard and Active Data Guard[27]. Similarly any restore and recovery operations should be validated to ensure backups and procedures are valid.

» Configure Exadata monitoring and Automatic Service Request[28]. Incorporate monitoring best practices as described in Enterprise Manager MAA OTN website.

» In addition to the various best practice references provided in this paper, Oracle maintains My Oracle Support Note 757552.1 as a master note for technical information pertaining to Exadata.

Importance of a Test Environment

Investment in sufficient test system infrastructure is essential to Exadata MAA. The benefits and trade-offs of various strategies for deploying test systems for Exadata are described in Table 2.

TABLE 2. TRADEOFFS FOR DIFFERENT TEST AND QA ENVIRONMENTS

TEST ENVIRONMENT	BENEFITS AND TRADEOFFS
Full Replica of the Production Exadata	Validate all patches and software changes. Validate all functional tests. Full performance validation at production scale Full HA validation especially if the replica includes the standby system.
Standby Exadata	Validate most patches and software changes. Validate all functional tests. Full performance validation if using Data Guard Snapshot Standby but this can extend recovery time if a failover is required. Role transition validation. Resource management and scheduling is required.
Shared Exadata	Validate most patches and software changes. Validate all functional tests. This environment may be suitable for performance testing if enough system resources can be allocated to mimic production. Typically, however, a subset of production system resources, compromising performance testing/validation. Resource scheduling is required.
Smaller Exadata system or Exadata with Exadata Snapshots	Validate all patches and software changes. Validate all functional tests. No performance testing at production scale. Limited full-scale high availability evaluations. Exadata snapshots are extremely storage efficient.
Older Exadata system	Validate most patches and software changes. Limited firmware patching test. Validate all functional tests unless limited by some new hardware feature Limited production scale performance tests. Limited full-scale high availability evaluations.

---

27 http://www.oracle.com/technetwork/database/availability/maa-roletransitionbp-2621582.pdf

28 http://www.oracle.com/us/support/auto-service-request/index.html

TEST ENVIRONMENT	BENEFITS AND TRADEOFFS
Non-Exadata system	Validate database and grid infrastructure software and patches only.
	Validate database generic functional tests.
	Limited testing of Exadata specific software features (e.g., HCC, IORM, Storage Index, etc.)
	Very limited production scale performance tests
	Limited high availability evaluations.

Conclusion

Exadata MAA is an integrated solution that provides the highest performing and most available platform for Oracle Database. This technical whitepaper has highlighted the HA capabilities that are delivered pre-configured with every Exadata Database Machine along with post-delivery configuration and operational best practices used by administrators to realize the full benefits of Exadata MAA.

Appendix 1: Exadata MAA Outage and Solution Matrix

Unplanned Outages

The following outage and solution matrix in Table 3 is an example of extensive high availability testing that Oracle conducts. The MAA recommended solution is provided for each type of outage along with the expected application recovery time (RTO) assuming sufficient system resources are still available to meet your application's performance SLAs and the application has been configured to transparently fail over to an available service..   To evaluate operational readiness and evaluate if your application's performance SLAs are met, Oracle recommends simulating the key faults (e.g. instance failure, node failure, disk failure, cell failure, logical failures and hangs) while running a real-world (using Real Application Testing and Database Replay) workload on an Exadata MAA test system. The priority column reflects suggested testing priority based on combination of probability of occurrence, importance of operational readiness, customer testing importance (not Oracle testing priority). Most outages should incur zero database downtime and a minimal application brownout for any connections.   If comparing with different hardware or storage vendor, inject the same equivalent fault and repeat the same workload for both environments.   For real world examples of how Exadata achieves end-to-end application availability and near zero brownout for various hardware and software outages, refer to this Exadata MAA video (http://vimeo.com/esgmedia/exadata-maa-tests).

Whether you are deploying manual or automatic failover, evaluate end-to-end application failover time or brownout in addition to understanding the impact that individual components have on database availability. The table includes links to detailed descriptions in Chapter 13, "Recovering from Unscheduled Outages" in Oracle Database 11g High Availability Best Practices[29]. Customers running Oracle Database 12c should refer to the outage matrix published in Oracle 12c High Availability Best Practices[30].

If there are sufficient system resources after an unplanned planned outage, the application impact can be very low as indicated by the table below.

TABLE 3. UNPLANNED OUTAGE/SOLUTION MATRIX

OUTAGE SCOPE	FAULT INJECTION PROCESS	EXADATA MAA	PRIORITY
site failure		Seconds to 5 minutes[31] Database Failover with a Standby Database Complete Site Failover Application Failover	LOW BUT WORTH TESTING FOR DR READINESS
clusterwide failure or production Exadata Database Machine failure		Seconds to 5 minutes Database Failover with a Standby Database Complete Site Failover Application Failover	LOW BUT WORTH TESTING FOR DR READINESS

---

29 http://docs.oracle.com/cd/E11882_01/server.112/e10803/outage.htm#i1005910

30 http://docs.oracle.com/database/121/HABPT/outage.htm#i1005910

31 Recovery time indicated applies to database and existing connection failover. Network connection changes and other site-specific failover activities may lengthen overall recovery time.

OUTAGE SCOPE	FAULT INJECTION PROCESS	EXADATA MAA	PRIORITY
computer failure (node) or RAC database node failure (simulating the impact of hardware failure, RAC node evictions, reboots or motherboard failure)	1. Unplug or forcefully power off database node 2. Wait 30 seconds or more 3. Restore power and power up database node, if needed 4. Wait for database node to be fully up	Small application downtime for cluster detection, cluster reconfiguration and instance recovery. For Exadata, cluster detection can be as low as 2 seconds for Grid Infrastructure 12.1.0.2 BP7 or higher. Managed automatically by Oracle RAC Recovery for Unscheduled Outages	HIGH
Database Instance failure or RAC database instance failure	Kill -11 PMON background process or shutdown abort the target instance	Small application downtime for affected connections. For the affected connections of the failed instance, brownout will consist of cluster reconfiguration (1sec) and instance recovery which is significantly faster on Exadata with Exadata write back flash cache. No database downtime[32]. Managed automatically by Oracle RAC Recovery for Unscheduled Outages	HIGH
Exadata Storage Server failure (simulating a storage head failure)	1. Unplug or forcefully power off storage cell 2. Wait longer than ASM disk repair timer	Small application impact with sub-second cell storage delay using our InfiniBand fabric fast detection mechanism.	LOW
Exadata disk pull and then push	1. Pull disk out Wait 10 seconds or more 2. Plug the same disk drive back in the same slot	Zero application brownout with Exadata write-back flash cache. Exadata and Oracle ASM tolerate storage failures and quickly redirect I/O to mirror(s) with minimum service level impact. Oracle can distinguish between a user pulling a good disk and a true failed disk. For a disk pull and push, a simple ASM resynchronization is done of the delta changes.	LOW
Exadata disk failure	Use the simulation commands: 1. alter physicaldisk <disk controller:disk slot #> simulate failuretype=fail 2. wait 1 minute 3. alter physicaldisk <disk controller:disk slot #> simulate failuretype=none	A true disk failure results in an immediate drop of the failed disk and subsequent ASM rebalance with no service level impact. Starting in Exadata cell 11.2.3.2.0 or higher, a blue LED light will indicate when the failed disk can be replaced.	HIGH

32Database is still available, but portion of application connected to failed system is temporarily affected.

OUTAGE SCOPE	FAULT INJECTION PROCESS	EXADATA MAA	PRIORITY
Exadata flash disk or flash DOM failure	1. Cannot physically pull the flash disk. Simulation command: alter physicaldisk <physicaldisk name of flash module> simulate failuretype=fail 2. Wait 1 minute 3. End Simulation command: alter physicaldisk <physicaldisk name of flash module> simulate failuretype=none	Small application impact with write back flash cache and fast repair of stale data.	MEDIUM
Power failure or PDU failure or loss of power source or supply to any computer or Exadata cell storage server	1. Pull power to one of the PDU	No application brownout due to redundant power failure.	LOW
human error		< 30 minutes[33] Recovering from Human Error	HIGH
hangs or slow down		See *Oracle Database High Availability Overview* documentation for solutions for unplanned downtime and for Application Failover	HIGH

Planned Maintenance

Table 4 shows the preferred approaches for performing scheduled maintenance on the Exadata after testing and patching best practices have been followed. The table includes links to detailed descriptions in Chapter 14, "Reducing Downtime for Planned Maintenance" in Oracle Database 11g High Availability Best Practices"[34]. Customer running Oracle Database 12c should refer to Oracle 12c High Availability Best Practices documentation[35]. For an overview of Exadata Planned Maintenance, start with Oracle Exadata Software Planned Maintenance presentation. Also see My Oracle Support Note 888828.1[36] or Exadata Maintenance Guide for additional information. To evaluate operational readiness and evaluate if your application's performance SLAs are met, Oracle recommends focusing on the key planned maintenance events (e.g. Oracle database or Grid Infrastructure maintenance upgrades, Exadata Platform software upgrades) while running a real-world (using Real Application Testing and Database Replay) workload on an Exadata MAA test system. The estimated downtime column reflects

---

33 Recovery times from human errors depend primarily on detection time. If it takes seconds to detect a malicious DML or DLL transaction, then it typically only requires seconds to flash back the appropriate transactions, if properly rehearsed. Referential or integrity constraints must be considered.

34 https://docs.oracle.com/cd/E11882_01/server.112/e10803/schedule_outage.htm#HABPT044

35 http://docs.oracle.com/database/121/HABPT/outage.htm#HABPT004

36 https://support.oracle.com/CSP/main/article?cmd=show&type=NOT&id=888828.1

the impact typically observed for the primary database for a tested and rehearsed maintenance activity.   It is a standard practice to validate any planned maintenance activity in your test environment first.

Standby-First Patching - Reduce Risk and Downtime with Data Guard

For all planned maintenance operations that involve software update or patch apply (for Exadata Platform, Oracle Grid Infrastructure, and Oracle Database), it is recommended to perform the update initially to a Data Guard Standby system first, using the guidelines and qualifications for Standby-First Patching, as described in My Oracle Support Note 1265700.1. Software updates performed in this manner have no impact to the primary database.

If there are sufficient system resources after a planned maintenance, the application impact can be very low as indicated by the table below.

TABLE 4. SOLUTIONS FOR SCHEDULED OUTAGES ON THE PRIMARY SITE

PLANNED MAINTENANCE	PREFERRED ORACLE SOLUTION	PRIMARY DATABASE ESTIMATED DOWNTIME	FREQUENCY
Oracle Grid Infrastructure (GI) patch set, maintenance, or major release upgrade	Oracle Grid Infrastructure rolling patch upgrade (see your platform-specific Oracle Clusterware Installation Guide for complete details). See also: • Oracle Database and Grid Infrastructure Patching • Automatic Workload Management for System Maintenance	No database downtime; zero or minimum application impact with service relocation	1+ years
Oracle Database patch set, maintenance, or major release upgrade	Oracle Database rolling upgrade with Data Guard (transient logical standby) or GoldenGate If Data Guard is not applicable or if less downtime is required by using active-active replication, consider Oracle GoldenGate. See also: • Database Upgrades	< 5 minutes with Data Guard; No downtime with Golden Gate	1+ years
Apply quarterly Exadata Database Bundle Patch (e.g. Database Patch for Engineered Systems and Database In-Memory for Oracle Database 12c, or Quarterly Database Patch for Exadata for Oracle Database 11g) to Oracle Grid Infrastructure and/or Oracle Database	Oracle RAC rolling patch installation using opatch and Out-of-place patching. See also: • Oracle Database and Grid Infrastructure Patching. • Automatic Workload Management for System Maintenance	No database downtime. zero or minimum application impact with service relocation	3-12 months

PLANNED MAINTENANCE	PREFERRED ORACLE SOLUTION	PRIMARY DATABASE ESTIMATED DOWNTIME	FREQUENCY
Apply Oracle interim patch or diagnostic patch to Oracle Grid Infrastructure and/or Oracle Database	Oracle RAC rolling patch installation using opatch orOnline Patching. See also: • Oracle Database and Grid Infrastructure Patching, and Online Patching • Automatic Workload Management for System Maintenance	No database downtime; zero or minimum application impact with service relocation	As required
Exadata Platform Software update: Database Server software quarterly update or new release	Oracle RAC rolling upgrade and service relocation See also: • Automatic Workload Management for System Maintenance • Exadata Maintenance Guide • Oracle Exadata Software Planned Maintenance presentation	No database downtime; zero or minimum application impact with service relocation	Quarterly update: 3-12 months New release: 1-2 years
Exadata Platform Software update: Storage server software quarterly update or new release	Exadata storage server software rolling update with patchmgr See also: • Exadata Maintenance Guide	No downtime	Quarterly update: 3-12 months New release: 1-2 years
Exadata Platform Software update: InfiniBand Switch software	Exadata InfiniBand switch software rolling update with patchmgr See also: • Refer to Exadata Maintenance Guide	No database downtime; short application brownout	1-2 years
Exadata Platform Software update: Power distribution unit (PDU) Keyboard, video, mouse (KVM)	Refer to Exadata Maintenance Guide	No database downtime; no application impact	1-2 years (if necessary)
Site maintenance	Site, Hardware, and Software Maintenance Using Database Switchover Complete Site Failover, Application Failover	< 5 minutes	As required
Database object reorganization or redefinition	Online object reorganization with DBMS_REDEFINITION (see Data Reorganization and Redefinition)	No downtime	As required

PLANNED MAINTENANCE	PREFERRED ORACLE SOLUTION	PRIMARY DATABASE ESTIMATED DOWNTIME	FREQUENCY
Database platform or location maintenance	Database Platform or Location Migration	< 5 minutes	As required

Integrated Cloud Applications & Platform Services

Deploying Oracle Maximum Availability Architecture with Exadata Database Machine
May 2016

Oracle is committed to developing practices and products that help protect the environment.