Oracle Maximum
Availability Architecture

# Oracle Maximum Availability Architecture with Exadata Cloud at Customer and Exadata Cloud Service

ORACLE WHITE PAPER | NOVEMBER 2017

ORACLE

.

## Overview

The Oracle Exadata Database Machine is engineered to deliver dramatically better performance and cost effectiveness, with inherent high availability features for Oracle databases. Exadata features a modern cloud-based architecture with scalable high performance database servers, scalable intelligent storage servers with state-of-the-art PCI flash, and an ultra-fast InfiniBand internal fabric that connects all servers and storage. Unique software algorithms in Exadata implement database intelligence in storage, compute, and InfiniBand networking to deliver higher performance and capacity at lower costs than other platforms.

Leveraging all the benefits of Exadata, *Oracle Database Exadata Cloud Service* (ExaCS) is now offered in Oracle Public Cloud for customers who want to run mission critical databases in the public cloud and utilize the offsite nature of cloud for Disaster Recovery purposes. For customers who desire the benefits of the cloud, but require their databases to be located local in their datacenters, they can choose *Oracle Database Exadata Cloud at Customer* (ExaCC). ExaCC delivers the world's most advanced database cloud to customers and is ideal for customers who cannot move their databases to the public cloud due to sovereignty laws, industry regulations, corporate policies, or organizations that find it impractical to move databases away from other tightly coupled on-premises IT infrastructure.

Databases deployed on ExaCS or ExaCC are 100% compatible with existing on-premises databases, or databases that are deployed in Oracle's public cloud.

Oracle Maximum Availability Architecture (MAA) is a set of best practice blueprints for the integrated use of Oracle High Availability (HA) technologies (See Figure 1). MAA best practices are created and maintained by a team of Oracle developers that continually validate the integrated use of Oracle Database High Availability features. Real-world customer experience is also integrated into the validation performed by the MAA team, spreading lessons learned to other customers.

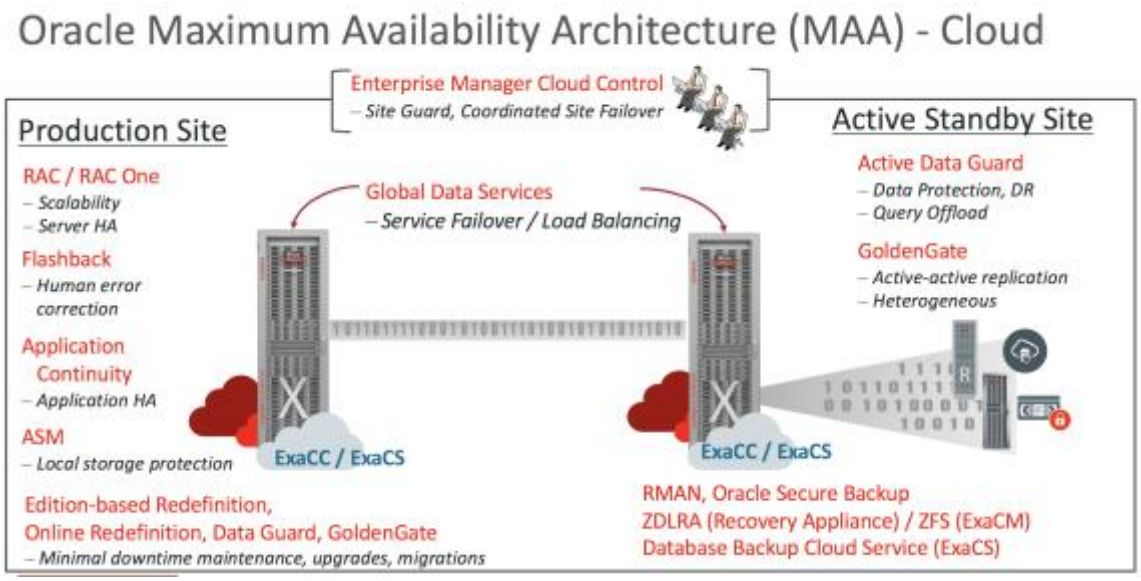


FIGURE 1: MAXIMUM AVAILABILITY ARCHITECTURE

The integration of Oracle Maximum Availability Architecture (Oracle MAA) with ExaCS and ExaCC provides the most comprehensive high availability solution for Oracle Databases in cloud deployments. In fact, Exadata in an MAA configuration is recognized by the analyst firm IDC as a system that delivers at least 99.999% availability and is categorized in the IDC AL4 fault-tolerant market segment.

Exadata 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 on Exadata MAA. Every Exadata system has gone through extensive availability testing both internal to Oracle and by mission critical customers worldwide. Global Exadata community improvements are channeled back into further enhancements that benefit every Exadata system.

This paper describes the inherent high availability ExaCC and ExaCS capabilities and how they can be expanded with Oracle's MAA reference architectures and best practices to deliver higher levels of availability. Database High Availability options such as Oracle Real Application Cluster, Oracle Active Data Guard, and Oracle Multitenant are included with ExaCC and ExaCS subscriptions.

Topics covered in this paper are:

- ExaCC and ExaCS MAA Reference Architectures
- ExaCC and ExaCS HA Benefits
- Addressing Unplanned Outages
- Addressing Planned Maintenance
- Cloud Backup and Recovery Options

This white paper is intended for CIOs, Enterprise Architects, Database Architects, Database Administrators, Technology Managers, Solution Architects, Application Architects, and those who are influential in the overall design of database architecture.

## ExaCS and ExaCC Administration

Before going into the details of MAA best practices, it is important to understand what operations in ExaCS or ExaCC customers can execute and what is managed by Oracle.

Customers have complete access to all Oracle Database and OS features to ensure smooth and simple migration from on-premises Oracle deployments to ExaCC or ExaCS. Each database instance is configured with Virtual Machine (DomU) in each database server of the Exadata system and is owned by the customer. Customers have root privileges in the domU and DBA privileges on the Oracle databases. Customers can configure the system as they like, and load additional agent software on the Exadata database servers to conform to business standards or security monitoring requirements.

Customers perform familiar database administration and OS administration tasks aided by cloud automation for backup, patching, and upgrades. Database and OS updates are initiated by customers on their preferred schedules. In ExaCS model, Oracle deploys and manages Exadata infrastructure. In ExaCC model, the underlying infrastructure for Exadata Cloud at Customer is deployed and managed by Oracle. Infrastructure includes Exadata InfiniBand network, physical Exadata Database and Storage servers, firmware, Dom0, and Exadata Storage Server Software. Since it is deployed in customer datacenter, power, space, cooling, and networking infrastructure is provided by the customer. This model allows customers to focus on business application requirements, and not on database infrastructure management.

## ExaCS and ExaCC MAA Reference Architectures

Oracle MAA best practices define three high availability reference architectures applicable to the Exadata platforms that address the complete range of availability and data protection required by enterprises of all sizes and lines of business. The MAA architectures, or tiers, are designated Platinum, Gold, and, Silver. They deliver the service levels described in Figure 2.

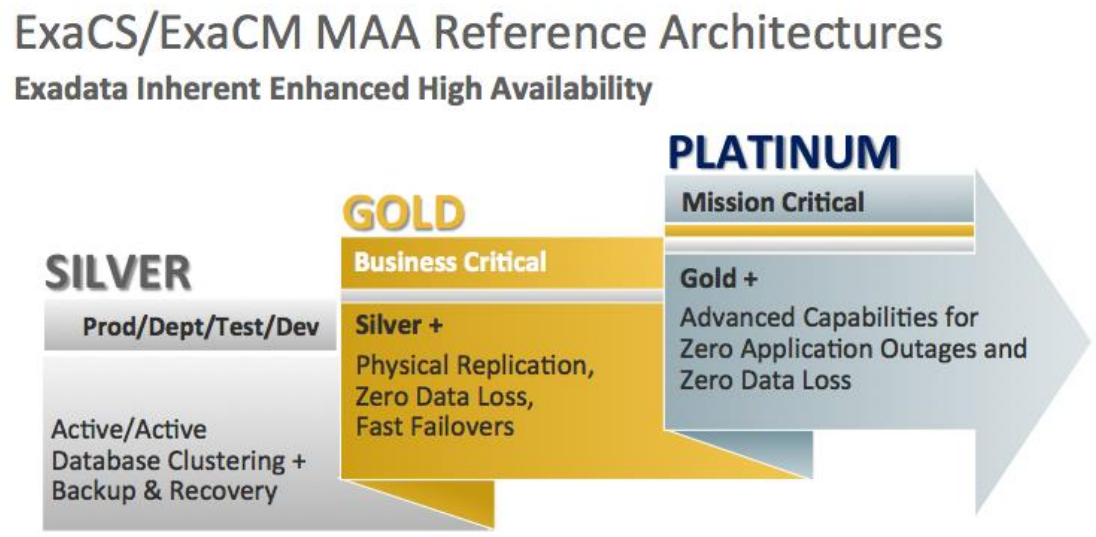


FIGURE 2: MAA REFERENCE ARCHITECTURES

Each tier uses a different MAA reference architecture to deploy the optimal set of Oracle high availability capabilities that reliably achieve a given service level at the lowest cost and complexity. The tiers explicitly address all types of unplanned outages including data corruption, component failure, system and site outages, as well as planned outages due to maintenance, migrations, or other purposes.

By integrating MAA reference architectures with ExaCC and ExaCS, customers can reap an even higher availability level with significantly less administration and cost when compared with other platforms. Though MAA reference architectures starts with "Bronze" level, there is no Bronze tier of reference architecture for Exadata based systems because each Exadata system has built-in high availability for localized outages.

***NOTE**: Unless explicitly called out, the MAA reference architectures discussed in this section are applicable to both ExaCS and ExaCC.*

## SILVER

Silver level MAA architecture provides a simple way to configure disaster recovery using backup & recovery. High Availability is achieved by deploying a clustered database using Oracle Real Application Clusters (Oracle RAC) or Oracle RAC One[1]. Oracle RAC provides server level protection for both planned and unplanned outages. Periodic backups are performed at the local site and the backups are replicated on a remote site for disaster recovery. For database consolidation with highest density and ROI, Oracle Multitenant Architecture using Container Database with one or more Pluggable databases is recommended. For ExaCC backups, it is recommended to use Zero Data Loss Recovery Appliance (ZDLRA or Recovery Appliance). Automated tooling is provided to use NFS storage for backups. Configure the storage to be replicated to the remote site for disaster recovery purposes. For ExaCS, Oracle Database Backup Cloud Service is configured as the backup destination. Automated tooling is in place for ExaCS and ExaCC to run daily and weekly backups.

[1] Oracle RAC One can be manually configured with ExaCC and will be supported with ExaCS in the future

Recovery time may be longer depending on the amount of data to be restored and recovered. Potential data loss depends on when the last backup was done. Figure 3 depicts Silver Reference Architecture.

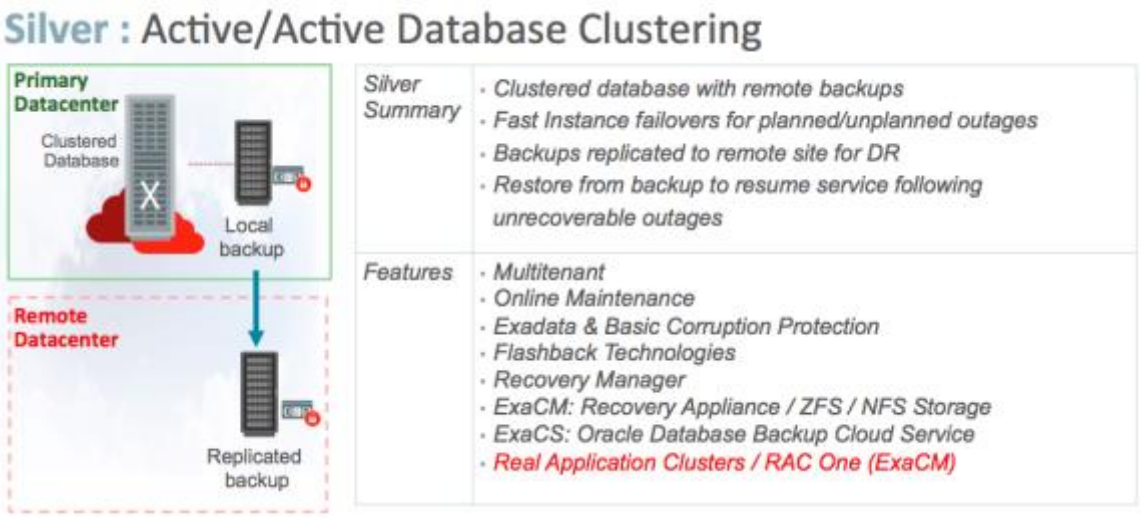


FIGURE 3: SILVER REFERENCE ARCHITECTURE

## GOLD

Gold enhances the Silver tier by adding real-time database replication to another data center using Oracle Active Data Guard. This architecture not only protects databases from site failure, but also enables faster failovers (as no recovery is required in general) with potentially zero data loss. You can add another ExaCC to the local data center for an additional copy of data for local failovers. If the distance between the two data centers is far enough to create higher latency in the network, and you still want zero data loss upon failovers, then the Active Data Guard Far Sync feature can be deployed. With ExaCS, you may choose to use either a compute or DBCS instance for deploying Far Sync. With ExaCC, you can deploy it on any x86 Linux-x86 64-bit server.

With ExaCC, you can choose to deploy a local standby using SYNC mode of Data Guard replication along with Data Guard Fast-Start Failover and integrated application failover so that zero data loss can be established. Local backups for all databases will be configured. This configuration provides complete isolation and protection. With Active Data Guard, read-only operations such as analytical queries, reporting, backups, and testing can be offloaded to the standby databases. Symmetric or similar Exadata configurations are recommended to achieve the same Service Level Agreements (SLA) upon failovers.

While deploying ExaCC Disaster Recovery across data centers, if the network latency between two customer data centers is low enough (single digits), a single control plane is enough to manage both the ExaCCs. For high latent network, it is recommended to use one control plane on each site. Far Sync instances and observer are to be deployed in another availability domain (physically separate location, power etc.) than the databases to isolate the fault. Figure 4 depicts Gold MAA reference architecture.

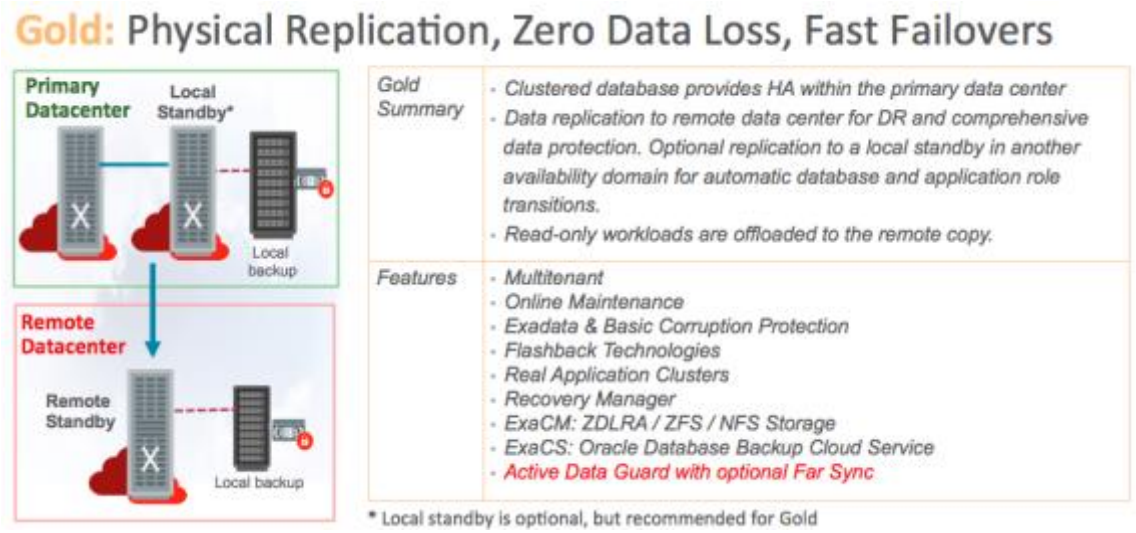


FIGURE 4: GOLD REFERENCE ARCHITECTURE

## PLATINUM

To achieve ultimate data protection from both planned and unplanned downtimes with zero application outage and zero data loss, additional components such as Oracle GoldenGate, Oracle 12c Application Continuity, Edition Based Redefinitions, and Oracle 12c Far Sync Data Guard are added to the configuration. Data is replicated to another availability domain locally in SYNC mode, and also replicated to the remote site in another data center using Oracle Active Data Guard. To achieve zero data loss, Far Sync Active Data Guard is deployed. Each site should have its own backup strategy. Oracle GoldenGate and Edition Based Redefinition are deployed to enable zero downtime application upgrades. Platinum deployments are generally custom configured to meet the highest availability requirements.

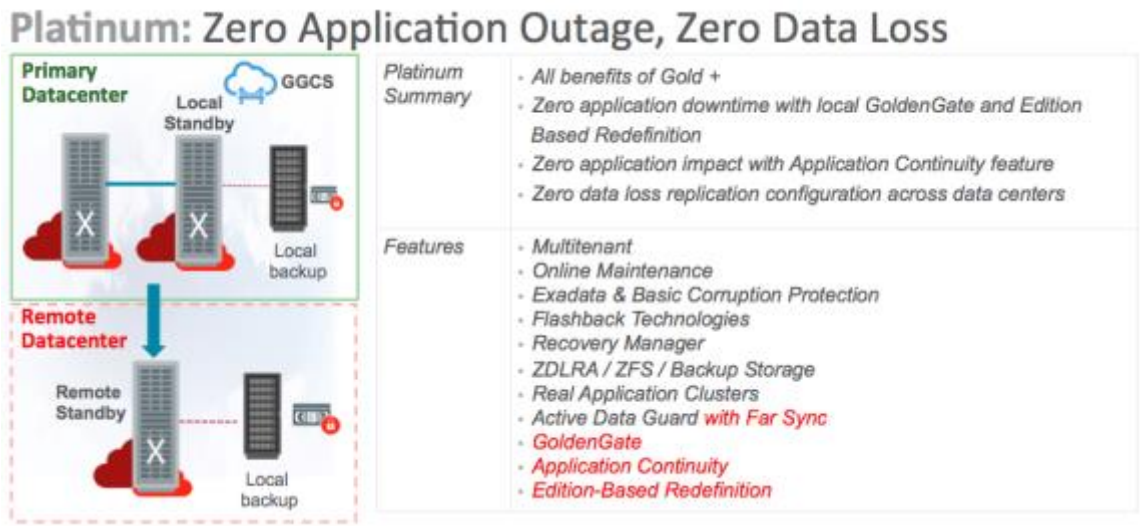


FIGURE 5: PLATINUM REFERENCE ARCHITECTURE

In summary, each reference architecture provides different level of protection against potential downtime and data loss. You can deploy a mix of these architectures based on the SLAs of each database. For example, mission critical production databases are generally deployed with Gold or Platinum level data protection, while test databases can be deployed using the Silver level of data protection.

Table 1 provides an overview of the achievable high availability Service Level Agreements (SLAs) for each Exadata MAA reference architectures. For all cases, Exadata test systems are recommended and are discussed in more detail in the Importance of a Test Environment section of this paper.

TABLE 1: MAA REFERENCE ARCHITECTURE SUMMARY

| MAA Architecture | Deployment Model | Potential Down Time (RTO) | Potential Data Loss (RPO) | High Availability | Disaster Recovery |
|---|---|---|---|---|---|
| Silver using Oracle RAC / RAC One | ExaCC/ExaCS Primary in one data center + Backup storage is configured with replication to another data center.<br>Multitenant provide largest consolidation density | Seconds for local failures<br>Hours for disasters | Hours to day<br>ExaCC: With Recovery Appliance, seconds. | Oracle RAC or RAC One<br>Inherent Exadata HA for server, network and storage failures | Remote Backup / Restore |
| Gold using Silver + Oracle Active Data Guard DR | ExaCC/ExaCS Primary in one data center is replicated to ExaCS/ExaCC in another data center. With ExaCC, data can also be replicated ted to another ExaCC in the same data center but with separate power etc. for isolation and fast zero data loss failovers.<br>Local backups on both sites. | Seconds | 0* – seconds | Oracle RAC<br>Oracle Active Data Guard (SYNC)<br>Inherent Exadata HA for network and storage | Oracle Active Data Guard |
| Platinum using<br>Gold + Oracle GoldenGate + Application Continuity + Editions Based Redefinition | ExaCC/ExaCS Primary in one data center is replicated to ExaCS/ExaCC in a remote data center. ExaCC Primary is also replicated to a local .ExaCC. Features such as Oracle GoldenGate and Editions Based Redefinitiion enables zero downtime application upgrade. Application Continuity for zero Application Impact.<br>Local backups on both sites. | Zero Application Impact | Zero | Oracle RAC or<br>Oracle Active Data Guard with FSFO<br>Inherent Exadata HA for network and storage | OracleActive Data Guard and<br>Oracle GoldenGate |

## Application Stack Protection

The MAA reference architectures address protection for the database layer. Protecting application layer is equally critical, but explaining application layer protection in detail is beyond the scope of this paper – as every application is different and exhibits varied I/O characteristics with the database. However, Oracle recommends co-locating applications along with the database tier. Database tier is protected using Data Guard replication. Any of the

following methods can be used for co-locating applications with the database tier during failover and switchover procedures.

1. Use symmetrical compute nodes for hosting applications in both primary and standby sites with application binaries pre-installed. Using Database File System (DBFS) on the database, you can copy application data periodically into DBFS using the RSYNC utility. Application data is then replicated to the standby site as part of Oracle Active Data Guard replication. You can then RSYNC the application data from DBFS into the compute node.
2. Use periodic or continuous RSYNC between compute nodes for application data sync. This requires timing co-ordination when a failover or switchover occurs.
3. In an ExaCC environment, if you already use Oracle Enterprise Manager Grid Control, then you can manually configure Oracle Site Guard for orchestrating failovers and switchovers between sites. The application stack replication can be done using the RSYNC method of storage replication such as ZFS replication.

## ExaCC and ExaCS High Availability Benefits

ExaCC and ExaCS inherit all the high availability benefits of Exadata Database Machines that are engineered and pre-configured to achieve end-to-end availability against hardware faults such as FANs, PDUs, batteries, switches, disks, flash, database server, motherboards and DIMMs and so on. Extensive engineering and integration tests are being performed on Exadata systems on a continuous basis. The high availability characteristics inherent in Exadata are described in the following sections and are fully applicable to ExaCC and ExaCS. Also, ExaCC and ExaCS specific enhancements are specifically called out.

### Hardware Components

#### Redundant database servers

Exadata contains preconfigured industry-standard Oracle Database servers running Oracle Database 12*c* Release 1 or Release 2 (12.1, 12.2) or Oracle Database 11*g* Release 2 (11.2). Oracle engineering and testing teams ensure that the firmware, software, and hardware configuration is tuned and pre-configured to provide high availability with scalability and the least brownout events. 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.

For example, a typical non-Exadata database server eviction caused by a database server failure will result in waiting on CSS misscount (defaulted to 30 or 60 seconds in most systems) before even declaring a database server is dead. During that time, the entire cluster freezes and there's an application brownout. Customers will observe 30 or 60 seconds of application blackouts and possibly minutes of application brownout before returning to steady state processing. In contrast, Exadata is pre-configured with fast node death detection that reduces brownout to 2 seconds or less, which means that the application brownout is reduced to 2 seconds or less. Refer to Figure 6 for an example an application outage after database server failures.

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

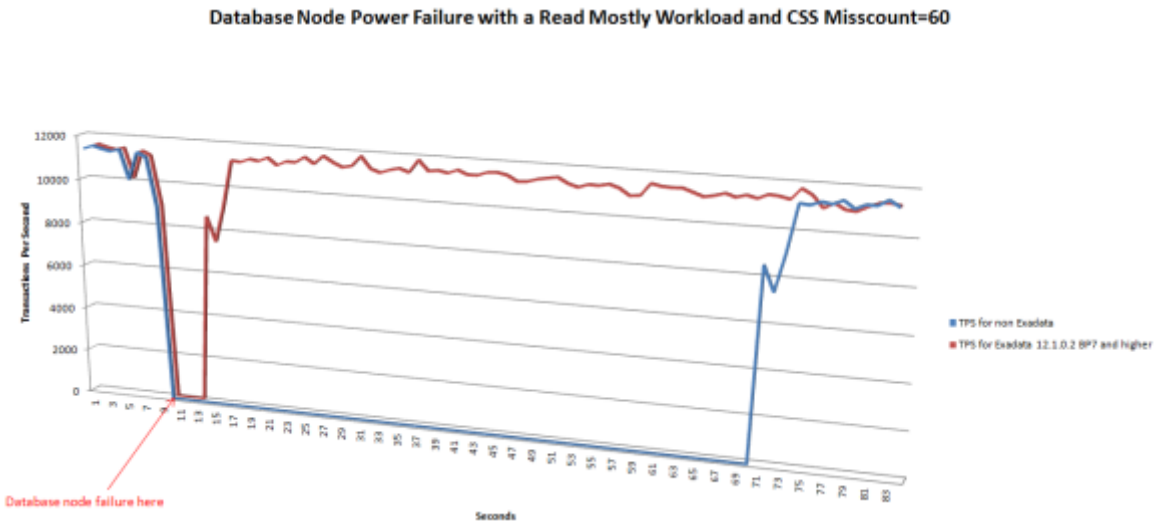


FIGURE 6: DATABASE SERVER POWER FAILURE

### Redundant storage

Exadata storage components such as database server disk drives, Exadata Storage Server disk drives, flash drives, and Oracle Exadata Storage Servers are all redundant. In ExaCS and ExaCC models, storage servers are fully managed by Oracle. Exadata Storage Servers are configured to tolerate hard disk, flash disk, flash card, and complete storage server failures. With ExaCS and ExaCC, DATA and RECO disk groups are pre-configured with ASM high redundancy, providing the highest form of data redundancy and storage protection. Database data blocks are mirrored across storage servers to ensure that the failure of an Exadata disk or Exadata storage server does not result in loss of data or availability. Also, disk drives are hot pluggable.

Exadata storage hardware and software are engineered for the lowest application brownout for storage failures and provide extensive data protection with Exadata HARD and Exadata disk scrubbing. Compared to 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 as compared to seconds and minutes with other storage running Oracle databases and applications.

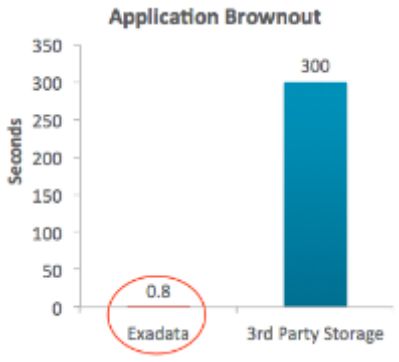


FIGURE 7: STORAGE FAILURE

### Redundant connectivity

ExaCC and ExaCS's InfiniBand network and switches are fully managed by Oracle. Redundant InfiniBand network connectivity using dual-ported Quad Data Rate (QDR), Host Channel Adapters (HCAs), and redundant switches are all pre-configured.

For network 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, and those are fully managed by Oracle. The PDUs accept separate power sources and provide a redundant power supply to:

- Exadata Database Servers
- Exadata Storage Servers
- InfiniBand switches
- Ethernet network switch

## Software Components:

The following are standard Oracle software components explicitly optimized and validated for ExaCS and ExaCC. Applying storage server updates or replacing any faulty components is managed by Oracle. Database updates, Grid Infrastructure and O/S updates are done by customers.

### Firmware and Operating System

All database and Exadata storage servers are packaged with validated firmware and operating systems preinstalled – including Dom0 of the database server. These are managed by Oracle.

### Database Server Tier

Grid Infrastructure (Oracle Clusterware and ASM) and Oracle RAC software are installed and patched to the recommended software version at deployment, enabling applications to tolerate and react to instance and node failures automatically with zero to near-zero application brownout. All Grid Infrastructure updates and most database updates can be applied in rolling fashion using Oracle Database Cloud Service Console or supplied cloud `exadbcpatchmulti` command. Database, Grid Infrastructure and the O/S within DomU is updated by customers.

### Storage Tier

Exadata storage servers 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. The HARD checks transparently handle all IO failure cases including ASM disk rebalance operations and disk failures. Refer to Appendix A for additional unique Exadata High Availability features and benefits. Storage tier updates and management is done by Oracle.

# Addressing Unplanned Outages

The MAA recommended solution for minimizing unplanned downtime is provided for each type of outage in Table 2 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 whether your application's performance SLAs are met, Oracle recommends simulating the key faults (for example, 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 ExaCC or ExaCS test system. Most outages should incur zero database downtime and a minimal application brownout for any connections. 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 following table includes links to detailed descriptions in Chapter 13, "Recovering from Unscheduled Outages" in Oracle Database 11g High Availability Best Practices[2]. Customers running Oracle Database 12c should refer to the outage matrix published in Oracle 12c High Availability Best Practices[3].

TABLE 2. UNPLANNED OUTAGE/SOLUTION MATRIX

| Unplanned Outage | ExaCS and ExaCC MAA | RTO | RPO |
|---|---|---|---|
| Site failure , clusterwide failure or production ExaCC or ExaCS failure | SILVER<br>Restore from backup and recover<br><br>GOLD and PLATINUM<br>Database Failover with a Standby Database<br>Complete Site Failover<br>Application Failover | Hours<br><br>Gold:<br>Seconds to min<br><br>Platinum:<br>Zero Application Downtime | Since the last backup<br>Seconds with Recovery Appliance<br><br>Gold:<br>Zero to Seconds<br><br>Platinum<br>Zero |
| Computer failure (Oracle RAC database server failure) or Database Instance failure (Oracle RAC database instance failure) | SILVER, GOLD, PLATINUM<br>Transparent Failover with Oracle RAC Recovery for Unscheduled Outages | Zero<br><br>For affected connections, seconds or for platinum, zero application impact. | Zero |
| Exadata Storage Server failure Including entire storage cell, disk, or flash failure | SILVER, GOLD, PLATINUM<br>Built-in redundancy and failover | Zero DB downtime<br>App brownout of secs in worse case. | Zero |
| Power failure or PDU failure or loss of power source or supply to any computer or Exadata cell storage server | SILVER, GOLD, PLATINUM<br>No application brownout due to redundant power failure. | Zero | Zero |
| Human error | < 30 minutes[4]<br>Recovering from Human Error | Varies | Varies |

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

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

Note: If there are sufficient system resources after an unplanned outage, the application impact can be very low as indicated by Table 2.

## Addressing Planned Maintenance

Oracle periodically applies software updates including security, network, or infrastructure updates automatically on ExaCC and ExaCS hardware and software with zero downtime. Customers are responsible for applying updates to the Exadata database servers from a supplied list of validated releases. Refer to Appendix B for details on addressing Planned Maintenance.

## Backup and Recovery Options for ExaCC and ExaCS

Both ExaCS and ExaCC has automated tooling in place for easier backup & recovery management. Since the deployment of ExaCS is in public cloud and ExaCC in customer datacenter, tooling is slightly different.

The following are the ExaCS backup options during service creation and deployment:

1. **Both Remote Storage and Exadata Storage**- enables two separate backup sets containing weekly full (RMAN level 0) backups and daily incremental backups. The backup to Exadata storage uses space in the RECO disk group using Incrementally updated backup strategy (Image Copy) and backup to the object storage using Oracle Database Backup Cloud Service for a default 30 days retention. This option requires subscription to the Oracle Database Backup Cloud Service.
2. **Remote Storage Only** - the remote storage is the cloud backup using Oracle Database Backup Cloud Service with a default of weekly full backups and daily incremental backups with 30 days retention. This option requires subscription to the Oracle Database Backup Cloud Service.
3. **None** - no automatic backups are configured. Users can perform on-demand backup.

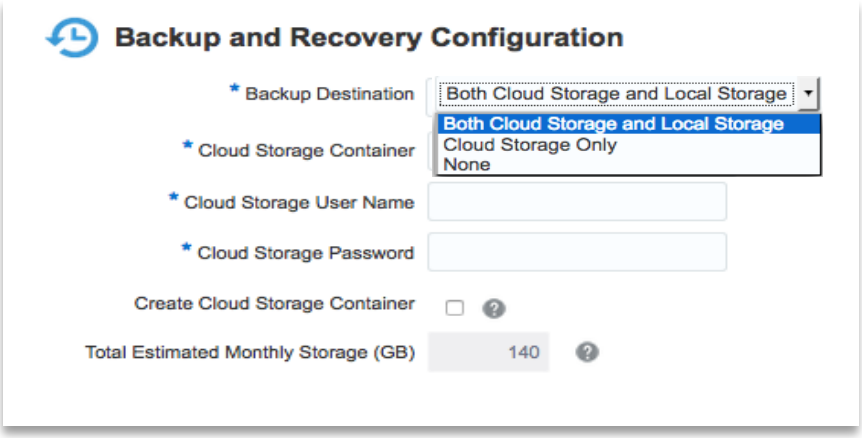


FIGURE 8: USER INTERFACE SAMPLE TO CONFIGURE BACKUP FOR EXA

4 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.

Since ExaCC is deployed on customer's datacenter, tooling is provided for taking backups to an Network File System (NFS) storage location. However, customers can choose to perform backups to other backup destinations – such as their on-premises Zero Data Loss Recovery Appliance or to Database Backup Cloud Service. Doing those non-NFS backup configurations currently involves few simple manual steps. To protect the data for disaster recovery, it is recommended to replicate or copy the backup data externally. MAA recommends standardizing on Recovery Appliance for backing up ExaCC and other Oracle databases in customer datacenter. Oracle Services would be able to help configuring manually that are not automated.

Figure 9 illustrates all of the possible backup configurations for ExaCS and ExaCC. For ExaCC, using 10GigE connectivity, customers can configure backups to a local NFS destination using a GUI or CLI.

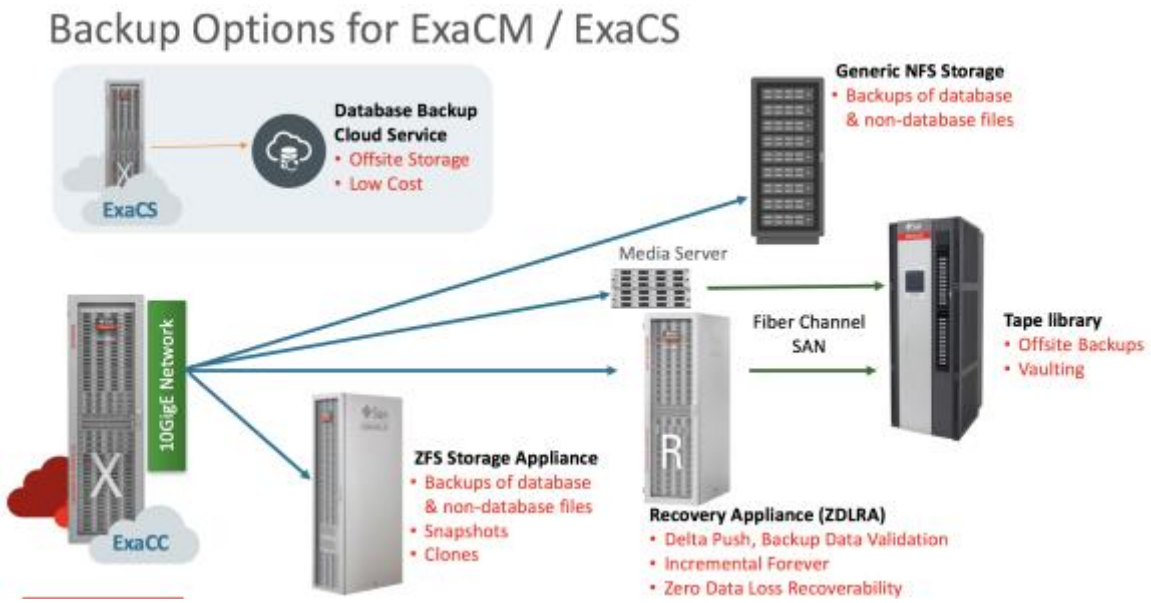


FIGURE 9: EXACS AND EXACC BACKUP CONFIGURATIONS

## Importance of a Test Environment

Similar to on-premises Exadata, investment in sufficient test system infrastructure is essential to ExaCS and ExaCC maximum availability. The benefits and trade-offs of various strategies for deploying test systems for Exadata are described in Table 3.

TABLE 3. TRADEOFFS FOR DIFFERENT TEST AND QA ENVIRONMENTS

| TEST ENVIRONMENT | BENEFITS AND TRADEOFFS |
|---|---|
| Full Replica of the Production ExaCC or ExaCS | Validate all Exadata Database Server updates and software changes.<br>Validate all functional tests.<br>Full performance validation at production scale<br>Full HA validation especially if the replica includes the standby system. |
| Standby Exadata, ExaCC, or ExaCS | Validate Exadata Database Server updates and software changes. Validate all functional tests. |

| TEST ENVIRONMENT | BENEFITS AND TRADEOFFS |
|---|---|
| | Full performance validation if using Data Guard Snapshot Standby but this can extend recovery time if a failover is required.<br>Role transition validation.<br>Resource management and scheduling is required. |
| Shared Exadata, ExaCC, ExaCS | Validate Exadata Database Server updates and software changes.<br>Validate all functional tests.<br>This environment may be suitable for performance testing if enough system resources can be allocated to mimic production.<br>Typically, however, a subset of production system resources, compromising performance testing/validation.<br>Resource scheduling is required. |
| Smaller Exadata system or Exadata with Exadata Snapshots | Validate Exadata Database Server updates and software changes.<br>Validate all functional tests.<br>No performance testing at production scale.<br>Limited full-scale high availability evaluations.<br>Exadata snapshots are extremely storage efficient. |
| Older Exadata system | Validate Exadata Database Server updates and software changes.<br>Limited firmware patching test.<br>Validate all functional tests unless limited by some new hardware feature<br>Limited production scale performance tests.<br>Limited full-scale high availability evaluations. |
| Non-Exadata system | Validate database and grid infrastructure software and updates only.<br>Validate database generic functional tests.<br>Limited testing of Exadata specific software features (e.g., HCC, IORM, Storage Index, etc.)<br>Very limited production scale performance tests<br>Limited high availability evaluations. |

## Conclusion

Exadata MAA is an integrated solution that provides the highest performing and most available platform for Oracle Database in the Cloud. This technical whitepaper has highlighted the high availability capabilities that are delivered pre-configured with every Oracle ExaCC and ExaCS system and describes how to enhance the platform with Exadata MAA reference architectures to protect from all failures and reduce downtime for planned maintenance activities.

## Appendix A: Additional Unique Exadata High Availability Features and Benefits

Refer to Table 4 below for additional Exadata high availability features and benefits. For a more detailed description of these capabilities, refer to Exadata documentation such as Oracle Exadata Database Machine System Overview, Exadata Machine Maintenance Guide and Exadata Storage Server Software User's Guide.

*Note: This information is provided for information purposes only. Features listed here are either handled automatically within Exadata or done by Oracle.*

TABLE 4: HA FEATURES AND BENEFITS

| AREA | FEATURE | HA BENFITS |
|---|---|---|
| REDUCED HA BROWNOUT | Fast Node Death Detection and failover | Reduced node failure detection from as many as 60 seconds to just 2 seconds or less. |
| | Automatic detection of Exadata storage failures with low application impact | Automatic detection and rebalance.<br>Application impact from 1 to 2 seconds delay |
| | Automatic detection of Exadata network failures with low application impact | Automatic detection and failover<br>Application impact from 0 to 5 seconds delay |
| | 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. |
| | Full high redundancy advantages for Oracle files and clusterware voting files with minimum of 3 storage Servers | 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. |
| 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.<br><br>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 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 |

| AREA | FEATURE | HA BENFITS |
|---|---|---|
| | | validation with Exadata is automatic. The HARD checks transparently handle all cases including ASM disk rebalance operations and disk failures |
| 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 I/O cases due to device driver, controller, or firmware issues or failing or dying disks, flash or bad storage sectors. |
| | 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 I/O cases. |
| | 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. |
| | 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. |
| | 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. Oracle Cloud Services or ACS will handle all replacements. |
| | 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. Customers can enable or adjust I/O resource plans via the Oracle Database Cloud Service Console or create custom IORM policies via logging a service request. |
| | 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.<br>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. |
| | 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 |

| AREA | FEATURE | HA BENFITS |
|---|---|---|
| | | 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.<br>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. |
| 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 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. |
| | Active Bonding Network | Exadata servers are 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 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. |
| | 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. |
| MANAGEMENT | Patching of Exadata Storage Servers, Exadata database Servers, and InfiniBand Switches | Patchmgr utility (and dbnodeupdate.sh) provides patching orchestration and automation for patching Exadata Storage Servers, Exadata database Servers and InfiniBand switches for both online and offline options. Oracle fully manages Exadata Infrastructure including Exadata Storage Servers and InfiniBand Switches. Oracle database cloud control plane provides `exadbcpatchmulti` command and REST APIs to patch Exadata database servers. |
| | 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 3+ times faster compared to previous releases. Even though most Exadata patching occurs with the application online, this enhancement |

| AREA | FEATURE | HA BENFITS |
|---|---|---|
| | | dramatically reduces the patching window. This is done by ACS . |
| | 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. Simplified status management. |
| | 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. |
| | 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 |
| | 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. |
| | 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. |
| | 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.<br><br>The Exadata Flash Log Statistics section in the AWR report now includes statistics for first writes to disk and flash. |

## Appendix B: Addressing Planned Maintenance

Table 5 shows the preferred approaches for performing scheduled maintenance on Exadata after testing and patching best practices have been followed. With ExaCS and ExaCC, planned maintenance is managed by Oracle. This information is provided to help you coordinate with Oracle for reducing downtime during planned maintenance.

To evaluate operational readiness and evaluate if your application's performance SLAs are met, Oracle recommends focusing on the key planned maintenance events (for example, 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 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 5. SOLUTIONS FOR SCHEDULED DOWNTIME 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).<br><br>See also:<br>• Oracle Database and Grid Infrastructure Patching<br>• Automatic Workload Management for System Maintenance<br>• Refer to ExaCS or ExaCC documentation. | 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<br>If Data Guard is not applicable or if less downtime is required by using active-active replication, consider Oracle GoldenGate.<br><br>See also:<br>• Database Upgrades | < 5 minutes with Data Guard;<br>No downtime with Golden Gate | 1+ years |
| Apply quarterly Exadata Database Bundle Patch (e.g. Database Patch | Oracle RAC rolling patch installation using opatch and Out-of-place patching. | No database downtime. | 3-12 months |

| PLANNED MAINTENANCE | PREFERRED ORACLE SOLUTION | PRIMARY DATABASE ESTIMATED DOWNTIME | FREQUENCY |
|---|---|---|---|
| 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 | See also:<br>• Oracle Database and Grid Infrastructure Patching.<br>• Automatic Workload Management for System Maintenance<br>• Refer to ExaCS or ExaCC documentation. | zero or minimum application impact with service relocation | |
| Apply Oracle interim patch or diagnostic patch to Oracle Grid Infrastructure and/or Oracle Database | Oracle RAC rolling patch installation using opatch or Online Patching.<br><br>See also:<br>• Oracle Database and Grid Infrastructure Patching, and Online Patching<br>• Automatic Workload Management for System Maintenance<br>• Refer to ExaCS or ExaCC documentation. | No database downtime; zero or minimum application impact with service relocation | As required |
| Database Server software quarterly update or new release | Oracle RAC rolling upgrade and service relocation<br><br>See also:<br>• Automatic Workload Management for System Maintenance<br>• Exadata Maintenance Guide<br>• Oracle Exadata Software Planned Maintenance presentation<br>• Refer to ExaCS or ExaCC documentation. | No database downtime; zero or minimum application impact with service relocation | Quarterly update: 3-12 months<br><br>New release: 1-2 years |
| Storage server software quarterly update or new release<br>- Handled by Oracle | Exadata storage server software rolling update with patchmgr<br><br>See also:<br>• Exadata Maintenance Guide | No downtime | Quarterly update: 3-12 months<br><br>New release: 1-2 years |
| InfiniBand Switch software<br>-Handled by Oracle | Exadata InfiniBand switch software rolling update with patchmgr<br><br>See also:<br>• Refer to Exadata Maintenance Guide | No database downtime; short application brownout | 1-2 years |
| Power distribution unit (PDU)<br>Keyboard, video, mouse (KVM)<br>-Handled by Oracle | Refer to Exadata Maintenance Guide | No database downtime; no application | 1-2 years (if necessary) |

| PLANNED MAINTENANCE | PREFERRED ORACLE SOLUTION | PRIMARY DATABASE ESTIMATED DOWNTIME | FREQUENCY |
|---|---|---|---|
| | | impact | |
| Site maintenance<br>- Hardware maintained by Oracle | Site, Hardware, and Software Maintenance Using Database Switchover<br>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 |
| Database platform or location maintenance | Database Platform or Location Migration | < 5 minutes | As required |

ORACLE

**Oracle Corporation, World Headquarters**
500 Oracle Parkway
Redwood Shores, CA 94065, USA

**Worldwide Inquiries**
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US


blogs.oracle.com/oracle


facebook.com/oracle


twitter.com/oracle


oracle.com

Integrated Cloud Applications & Platform Services



Deploying Oracle Maximum Availability Architecture with Exadata Cloud at Customer & Exadata Cloud Service
November 2017


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