Exadata Database Machine Administration Workshop

Student Guide - Volume II

D73668GC11
Edition 1.1
April 2012
D76623

ORACLE

Author

Peter Fusek

Technical Contributors and Reviewers

Akshay Shah
Alex Tsukerman
Amit Ganesh
Andrew Babb
Aslam Edah-Tally
Barb Lundhild
Bharat Baddepudi
Bill Hodak
Boris Erlikhman
Branislav Valny
Bruce Kyro
Caroline Johnston
Christian Craft
Dan Norris
Dave Winter
David Hitchcock
Douglas Utzig
Ed Gilowski
Eric Siglin
Georg Schmidt
Harald van Breederode
James He
Jean-Francois Verrier
Jia Shi
Jim Hall
Jim Spiller
Jim Viscusi
Joel Goodman
Juan Loaiza
Kam Shergill
Kevin Jernigan
Kodi Umamageswaran
Krishnanjani Chitta
Larry Justice
Lawrence To
Louis Nagode
Mahesh Subramaniam
Maria Billings
Mark Fuller
Mark Scardina
Mark Van de Wiel
Marshall Presser
Martin Jensen
Michael Cebulla
Michael Nowak
Nilesh Choudhury
Ravindra Dani
Raymond Dutcher
Richard Exley
Robert Carlin
Robert Pastijn
Roger Hansen
Sabyasachi Banerjee
Scott Gossett
Sriram Palapudi
Steven Lemme
Sue Lee
Sugam Pandey
Sumeet Lahorani
Sundararaman Sridharan
Tim Shelter
Umesh Panchaksharaiah
Uwe Hesse
Varun Malhotra
Vern Wagman
Vijay Sridharan
Vikram Kapoor
Vimala Jacob

Graphic Designer

Satish Bettegowda

Editors

Malavika Jinka
Smita Kommini
Aju Kumar

Publisher

Michael Sebastian Almeida

Contents

12

Bulk Data Loading using Oracle DBFS

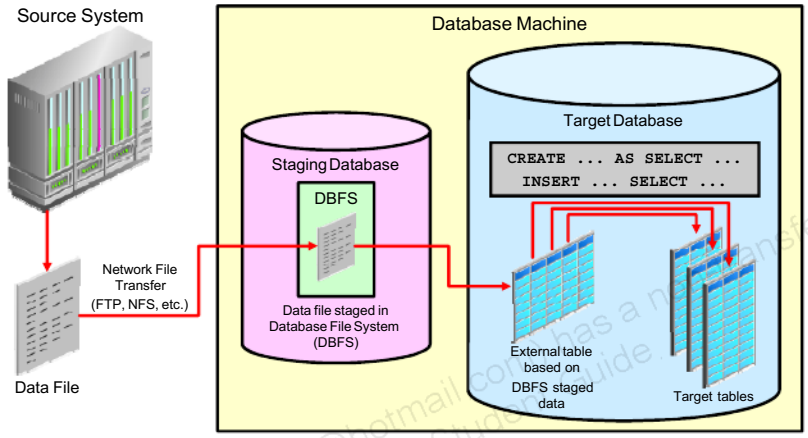
ORACLE

Objectives

After completing this lesson, you should be able to:

- Describe the process of using Oracle DBFS for bulk data loading into Database Machine
- Describe the process to configure the Database File System (DBFS) feature for staging input data files
- Use external tables based on input data files stored in DBFS to perform high-performance data loads

Bulk Data Loading using Oracle DBFS Overview



This lesson provides information on how to load databases on Database Machine with data from other systems using data files as the intermediate data store. The lesson focuses on the recommended approach for loading file-based data into Oracle using DBFS. This approach is illustrated on the slide. The remainder of this lesson describes this process in greater detail.

Note that DBFS can also be as a generalized shared file system on Database Machine. You can use the procedure for configuring DBFS contained later in this lesson to configure DBFS for purposes other than staging data files for bulk data loading.

Preparing the Data Files

- Data files can be in any format supported by the external table feature
- To facilitate high-performance parallel access:
  - Oracle automatically divides the files into 10MB granules
    - Exceptions: compressed files, data read from a pipe or a tape
  - If granules cannot be created then:
    - Each file is treated as a granule
    - The number of files determines the maximum degree of parallelism
    - You must use multiple files to manually enable parallelism
  - General rules of thumb
    - If using multiple files, then try to keep them similar in size
    - If the file sizes vary significantly, then list them in order from largest to smallest in the external table definition

The recommended approach for bulk data loading using Oracle DBFS relies on the external table feature of Oracle Database. The data files used for bulk data loading can be in any format supported by external tables. The process for creating the data files is outside the scope of this lesson and mostly depends on the facilities available in the source system.

However the data files are created, the following should be taken into account in order to facilitate high-performance parallel access to the data files while they are being loaded:

- Where possible, when accessing large data files through external tables Oracle automatically divides the files into 10 MB granules. These granules can be processed in separate parallel processing threads. Oracle is unable to use this approach with compressed files or data read from a pipe or a tape device.
- If granules cannot be used then each separate data file can be treated as a granule and the number of files determines the maximum degree of parallelism that is available. You can manually divide a large file into separate smaller files and use them to manually enable parallelism.
- If you are using multiple input data files in conjunction with a single external table, then you should try to keep the data files similar in size. If the file sizes do vary significantly, then list them in order from largest to smallest in the external table definition.

Staging the Data Files

- Data files should always be staged in DBFS
  - Not database server internal drives
- DBFS enables the database to be used as a file system
  - Shared storage for staging or storing data files, scripts, reports and other application files
  - Files are stored as SecureFiles LOBs inside database tables that are stored in Exadata
  - Files are exposed to the database servers using a POSIX-compatible file system interface
  - Files are protected like any Oracle data – ASM mirroring, Data Guard, Flashback, and so on

It is recommended to stage data files inside Database Machine using Database File System (DBFS). DBFS is an Oracle Database feature that enables the database to be used as a high-performance POSIX-compatible file system. Using the available space on internal database server disk drives for staging data files is highly discouraged.

Inside DBFS files are stored as SecureFiles LOBs. A set of PL/SQL procedures implement the file system access primitives, such as open, close, create, and so on. The `dbfs_client` utility enables the mounting of a DBFS file system as a mount point on the Database Machine database servers. It also provides the mapping from file system operations to database operations.

Note: The procedure detailed in this lesson applies to configuring DBFS on Linux database servers. ASM Cluster File System (ACFS) is not supported in conjunction with Exadata.

Configuring the Staging Area

DBFS should be configured in a separate staging database

- Use DBCA to create the staging database
  - See My Oracle Support note 1191144.1 for recommendations
- Create a bigfile tablespace for DBFS storage

```
SQL> CREATE BIGFILE TABLESPACE DBFS DATAFILE '+DBFS_DG'
     SIZE 32G AUTOEXTEND ON NEXT 8G MAXSIZE 300G
     NOLOGGING ONLINE PERMANENT EXTENT MANAGEMENT LOCAL
     AUTOALLOCATE SEGMENT SPACE MANAGEMENT AUTO;
```

- Create a DBFS user account

```
SQL> create user dbfs identified by dbfs
     quota unlimited on DBFS;
SQL> grant create session, create table, create
     procedure, dbfs_role to dbfs;
```

ORACLE

While DBFS is fully functional if it is co-located with the target database, it is recommended to configure DBFS in a separate staging database. Use the Database Configuration Assistant (DBCA) to create a database. If you wish to access DBFS across multiple Database Machine database servers create the staging database as a RAC database across the desired servers. Default settings are mostly suitable when configuring the staging database however My Oracle Support note 1191144.1 contains the following recommendations:

- Use the *General Purpose or Transaction Processing* template
- Use Oracle Managed Files and the DBFS_DG disk group to store the staging database
- De-select *Specify Flash Recovery Area*
- Set *Memory Size (SGA and PGA)* to 8192 MB
- De-select *Use Automatic Memory Management*
- Use `AL32UTF8` as the database character set
- Set the `PARALLEL_MAX_SERVERS` initialization parameter to 2

After the staging database is created, create a bigfile tablespace to use as the DBFS store. The slide shows an example `CREATE TABLESPACE` command containing the recommended tablespace options. Also, create a DBFS user account as suggested on the slide.

Configuring the Staging Area

- Additional database server operating system configuration:
  - Add the Oracle software owner, or user that will mount the DBFS file system, to the `fuse` group

```
# usermod -a -G fuse oracle
```

  - As `root`, create `/etc/fuse.conf` containing the entry: `user_allow_other`

```
# echo "user_allow_other" > /etc/fuse.conf
# chmod 644 /etc/fuse.conf
```

  - Create a mount point for DBFS with ownership and group permissions set to the Oracle software owner, or user that will mount the DBFS file system

```
# mkdir /data
# chown oracle:dba /data
```


To configure DBFS you must also perform the operating system configuration tasks listed on the slide. They should be performed on each database server where DBFS will be mounted. The examples show a typical configuration using the Oracle software owner as the DBFS mount point owner. You can create and use an alternative operating system user if you wish to separate DBFS access from the Oracle software owner and it's database administration privileges. You can also use the `dcli` utility to replicate the configuration steps on multiple database servers.

Configuring the Staging Area

- Creating the DBFS store

```
$ cd $ORACLE_HOME/rdbms/admin
$ sqlplus dbfs/dbfs
SQL> @dbfs_create_filesystem_advanced.sql DBFS mydbfs
     nocompress nodeduplicate noencrypt non-partition
```

- Mounting DBFS

```
$ nohup $ORACLE_HOME/bin/dbfs_client dbfs@<StagingDB>
  -o allow_other,direct_io /data < passwd.txt &
```

  - See My Oracle Support note 1054431.1 for automatic mounting configuration details

- Using DBFS
  - Access DBFS through the mount directory
  - Copy files to DBFS using methods such as FTP and NFS

After the staging database is created and the required operating system configuration is completed, you can create the DBFS store. Use the script located at `$ORACLE_HOME/rdbms/admin/dbfs_create_filesystem_advanced.sql` The script must be run by the DBFS database user (created earlier in the configuration process). The script accepts numerous parameters. In the example on the slide, `DBFS` represents the name of the tablespace you created to house the DBFS store, and `mydbfs` represents the name of the DBFS store. This name of the DBFS store is used later after DBFS is mounted to name the directory that appears under the DBFS mount directory.

To mount DBFS you can use `dbfs_client` as shown in the example on the slide. Note that the example references `/data` as the DBFS mount directory. Also, note that the DBFS user password is contained in a file called `passwd.txt`. The password file is only required when mounting DBFS and is not required while DBFS is being used. Alternatively you can configure an Oracle Wallet so that the DBFS client can mount a DBFS store without a password.

My Oracle Support bulletin 1054431.1 contains a procedure that can be used to automatically mount DBFS using Oracle Clusterware.

After DBFS is mounted you can access it through the mount directory (`/data` in the slide example). In the mount directory you will find a subdirectory named after the DBFS store; the contents of this subdirectory (`/data/mydbfs` in the example) is the contents of the DBFS store.

Note that it is possible to have multiple copies of `dbfs_client` accessing the same shared file system. The sharing and caching semantics are similar to NFS. Like NFS, the default mode caches writes on the client and flushes them after a timeout or when the file being modified is closed. Also like NFS, writes to a file are only visible to clients that open the file after the writer closed the file. This behavior is commonly referred to as close-to-open cache consistency.

For more information regarding DBFS on Database Machine refer to My Oracle Support bulletin 1054431.1. See also the chapters on DBFS located in the *Oracle Database SecureFiles and Large Objects Developer's Guide 11g Release 2 (11.2).*

Configuring the Target Database

- Prerequisites for data file access using external tables
  - Create an Oracle directory object that references the DBFS staging area directory
  - Grant the required permissions on the Oracle directory object
  - Create the required external tables
- Ensure efficient space management
  - Use bigfile tablespaces
  - Use 8MB initial extents for large segments
    - Set the `INITIAL` storage parameter to 8MB
    - Set the `CELL_PARTITION_LARGE_EXTENTS` initialization parameter to `TRUE` or `ALWAYS`
- Use unlimited quotas to bypass quota management
- Use the parallel clause to set the default degree of parallelism for the target tables

After the DBFS-based staging area is configured it can be populated with the required data files. To access the data files using external tables you must:

- Create an Oracle directory object that references the DBFS staging area directory.
- Grant the required permissions on the Oracle directory object so that the data files can be referenced by the required target database users.
- Create the required external tables referencing the data files stored in DBFS.

You should also prepare your target database more generally to optimize the load process. You should ensure efficient space management by following the general space management recommendations for use in conjunction with Exadata that are listed on the slide. In addition, you can optimize the performance of your loads if you bypass tablespace quota management and you set an appropriate default degree of parallelism for your target tables.

Loading the Target Database

- Recommended approach uses external tables
  - Parallel direct path loading for high-performance
  - In-flight processing using SQL
    - Transformations using SQL functions
    - Sort data while loading to optimize Exadata storage indexes
  - No need to re-stage the data
  - Other advanced features
    - For example: Input file preprocessing
- SQL*Loader can also be used
  - Parallel direct path loading can also be achieved
  - Less processing flexibility compared with external tables
  - Existing SQL*Loader scripts can be easily reused

The recommended method for loading the data files uses external tables. Loading data using external tables provides the following key benefits:

- You can use direct path loading and parallel processing to achieve high-performance.
- You can use the flexibility and power of SQL to efficiently process the data while it is being loaded. For example, you can use standard or user-defined SQL functions to transform the data during loading and you can sort the data while you load it to optimize the storage indexes that are automatically maintained by Exadata Storage Server.
- Because you can process the data while it is being loaded, it is unlikely that you will need to re-stage the data inside the target database.
- External tables provide some useful advanced features. A primary example is the ability to preprocess a data file using a user-defined program which provides additional flexibility and power to process the data using routines outside the database.

SQL*Loader can also be used to load data files. It can deliver comparable performance but does not provide the same level of processing flexibility when compared with external tables. SQL*Loader remains a good choice when customers already have SQL*Loader based scripts that they wish to reuse in a Database Machine environment.

Loading the Target Database

- Parallel loading using external tables
  - A `CREATE TABLE ... AS SELECT` statement automatically uses the defined degree of parallelism
  - An `INSERT ... SELECT` statement needs parallel DML to be enabled

```
SQL> alter session enable parallel dml;
```

- Direct path loading using external tables
  - A `CREATE TABLE ... AS SELECT` statement automatically uses direct path loading
  - An `INSERT ... SELECT` statement needs an APPEND hint to enable direct path loading

```
SQL> insert /*+ APPEND */ into my_table
     select * from my_external_table;
```


Bulk data loading using external tables is achieved using a `CREATE TABLE ... AS SELECT` statement or an `INSERT ... SELECT` statement.

To utilize parallel processing, it is recommended that you set the default degree of parallelism for the external tables used in the load and the target table being created or inserted into. Alternatively, you can set parallelism within the statement using a PARALLEL hint.

After parallelism is configured, a `CREATE TABLE ... AS SELECT` statement will automatically use the defined degree of parallelism. An `INSERT ... SELECT` statement requires parallel DML to be enabled before it can execute in parallel.

To utilize direct path lading from external tables, you must use an APPEND hint for an `INSERT ... SELECT` statement. `CREATE TABLE ... AS SELECT` statements automatically use direct path loading.

Remember that the bulk data loading approach outlined in this lesson can be used in conjunction with other techniques such as partition exchange loading.

Quiz

To facilitate parallel loading from typical flat files external table definitions must reference numerous smaller data files rather than one file containing all the data:

a. True

b. False

ORACLE

Answer: b

Typically, Oracle automatically divides typical flat files into 10MB granules to facilitate parallel processing. Oracle is unable to use this approach with compressed files or data read from a pipe or a tape device.

Quiz

While DBFS is fully functional if it is co-located with the target database, it is recommended to configure DBFS in a separate staging database:

a. True

b. False

Answer: a

Summary

In this lesson, you should have learned how to:

- Describe the process of using Oracle DBFS for bulk data loading into Database Machine
- Describe the process to configure the Database File System (DBFS) feature for staging input data files
- Use external tables based on input data files stored in DBFS to perform high-performance data loads

Additional Resources

- Lesson Demonstrations
  - Bulk Data Loading with Database Machine
- My Oracle Support Notes
  - Configuring a Database for DBFS on Oracle Database Machine
  - Configuring DBFS on Oracle Database Machine

Practice 12 Overview:
Bulk Data Loading using Oracle DBFS

In this practice you will perform a bulk data load on Database Machine. You will configure a database file system (DBFS) and use it to stage a CSV formatted file. You will then use the external table feature of Oracle Database to reference the CSV file. Finally, you will use a `CREATE TABLE AS SELECT` statement to copy the CSV file data into a table in your database.

13

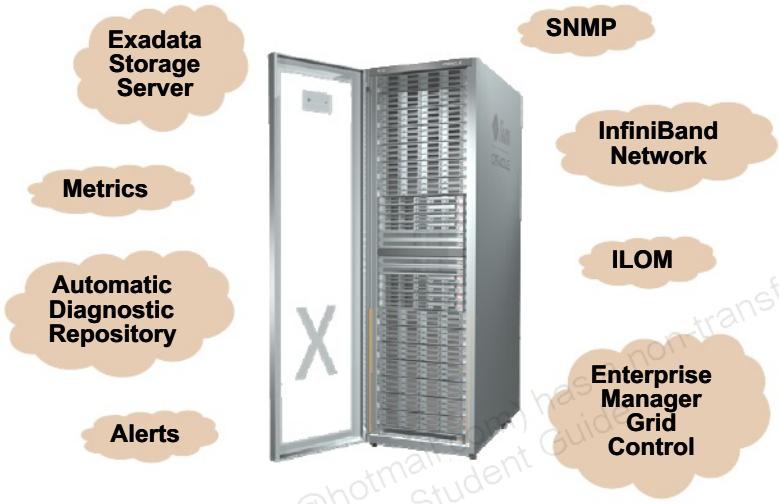
Exadata Database Machine Platform Monitoring Introduction

ORACLE

Objectives

After completing this lesson, you should be able to describe the key monitoring infrastructure technologies associated with Exadata Database Machine.

Monitoring Technologies and Standards



Embedded within the various Exadata Database Machine components are a variety of monitoring technologies. Some are based on widely adopted standards while others are specific to Database Machine. This lesson introduces the key monitoring technologies and standards used in conjunction with Database Machine. Most of the information presented in this lesson should be familiar to most students and it is presented to provide a consistent foundation for the following lessons.

Simple Network Management Protocol (SNMP)

- What is it?
  - A standard protocol for managing devices on a network
- What does it do?
  - Primarily used to propagate information about a device to a monitoring console somewhere on the network
    - Report alerts for hardware or software issues
  - Can be used to set device configuration parameters
- Where is it found?
  - Throughout Database Machine, including database servers, storages servers, InfiniBand switches, Ethernet switch, KVM (Keyboard Video and Mouse) switch and PDUs (Power Distribution Units)

Simple Network Management Protocol (SNMP) is an Internet-standard protocol for managing devices on IP networks. A wide array of devices support SNMP including many servers, network switches, routers, printers and workstations. SNMP is used mostly in network management systems to monitor network-attached devices for conditions that warrant administrative attention.

Typically, SNMP is used to propagate information about a device to a monitoring or management console located somewhere in the network. To do this, managed devices run an SNMP agent. SNMP agents expose management information about the managed systems as variables. The protocol also permits active management tasks, such as modifying and applying a new configuration through remote modification of these variables. SNMP agents can be separate processes or embedded into other software or hardware modules.

SNMP defines a variety of protocol data units (messages) that can be exchanged between agents and managers. One of the most important messages is an asynchronous notification from agent to manager known as a trap. An SNMP tap is often used by devices to report alert conditions.

SNMP is used extensively throughout Exadata Database Machine so that the various components of Database Machine can report monitoring information and alerts to network management systems such as Enterprise Manager Grid Control.

Intelligent Platform Management Interface (IPMI)

- What is it?
  - An open, industry-standard interface for server management
- What does it do?
  - Primarily used to perform server configuration and management independently of the server operating system
- Where is it found?
  - Exadata Database Machine database servers and Exadata Storage Servers

Intelligent Platform Management Interface (IPMI) is an open, industry-standard interface for the management of server systems over a number of different types of networks.

On face value it would seem that SNMP and IPMI perform essentially the same function however there are two main differences:

- IPMI is more focused on server management. IPMI functionality includes field-replaceable unit (FRU) inventory reporting, logging of system events and system recovery (including system resets, power on and power off).
- IPMI is associated with an architecture that allows administrators to remotely manage a system in the absence of an operating system or other system management software. The monitored system may be powered off, but the baseboard management controller (BMC) must be connected to a power source and the monitoring medium, typically a local area network connection. The BMC is a specialized microcontroller embedded in the server.

IPMI prescribes only the structure and format of the interfaces as a standard, while detailed implementations may vary.

Inside Database Machine, IPMI support is built into Integrated Lights Out Manager (ILOM) on each database server and Exadata Storage Server.

Integrated Lights Out Manager (ILOM)

- What is it?
  - Integrated service processor hardware and software
- What does it do?
  - Provides out-of-band server monitoring and management to:
    - Remotely control the power state of a server
    - View the status of sensors and indicators on the system
    - Provide a remote server console
  - Generates alerts for hardware errors and faults as they occur
- Where is it found?
  - Exadata Database Machine database servers and Exadata Storage Servers

Integrated Lights Out Manager (ILOM) provides advanced service processor hardware (baseboard management controller) and software that you can use to manage and monitor Database Machine servers. ILOM's dedicated hardware and software is preinstalled on the database servers and Exadata Storage Servers inside Database Machine. ILOM enables you to actively manage and monitor the server independently of the operating system state (out-of-band). With ILOM, you can:

- Learn about hardware errors and faults as they occur
- Remotely control the power state of your server
- View and use the graphical and non-graphical consoles for the host
- View the current status of sensors and indicators on the system
- Determine the hardware configuration of your system
- Receive alerts about important system events through notification methods such as SNMP traps and email alerts.

ILOM automatically initializes as soon as power is applied to the server. It provides a full-featured, browser-based web interface and has an equivalent command-line interface (CLI). There is also an industry-standard SNMP interface and IPMI support (ILOM is compliant with IPMI v1.5 and v2.0).

Exadata Storage Server Metrics, Thresholds and Alerts

- What are they?
  - Metrics, thresholds and alerts provide the foundation for monitoring Exadata Storage Server
- What do they do?
  - Metrics provide a measure relating to some aspect of storage server status or performance
  - Thresholds are metric levels, which if crossed automatically generate an alert notification
  - Alerts are automatically generated notifications of system events
- Where are they found?
  - Exadata Storage Server

Exadata Storage Server contains a system of metrics, thresholds and alerts which provide a foundation for storage server monitoring.

Metrics provide a measure relating to some aspect of storage server status or performance. Exadata Storage Server includes more than 150 metrics, many of which are associated with numerous separate measurements. For example, the current temperature of the server is a single-value metric. In contrast, the cumulative number of requests to read large blocks from a grid disk has a separate measurement for each grid disk. So on a typical Exadata cell this one metric can be associated with more than 20 separate observations.

Thresholds are definitions which allow administrators to define metric levels, which if crossed automatically generate an alert notification. Thresholds can contain two alert levels, warning and critical. For example, a threshold could be defined to generate a warning alert when cell memory utilization reaches 80%, and a critical alert when it reaches 90%.

Exadata Storage Server generates alerts for various system conditions. For example, an alert is generated if the cell server software terminates unexpectedly or if a sensor detects a potential imminent disk failure. Additional alerts can be generated for user-defined events through the use of thresholds.

Note that Enterprise Manager also contains a separate system of metrics, thresholds and alerts. Other software, such as Oracle Database, contains its own metrics and alerts.

Automatic Diagnostic Repository (ADR)

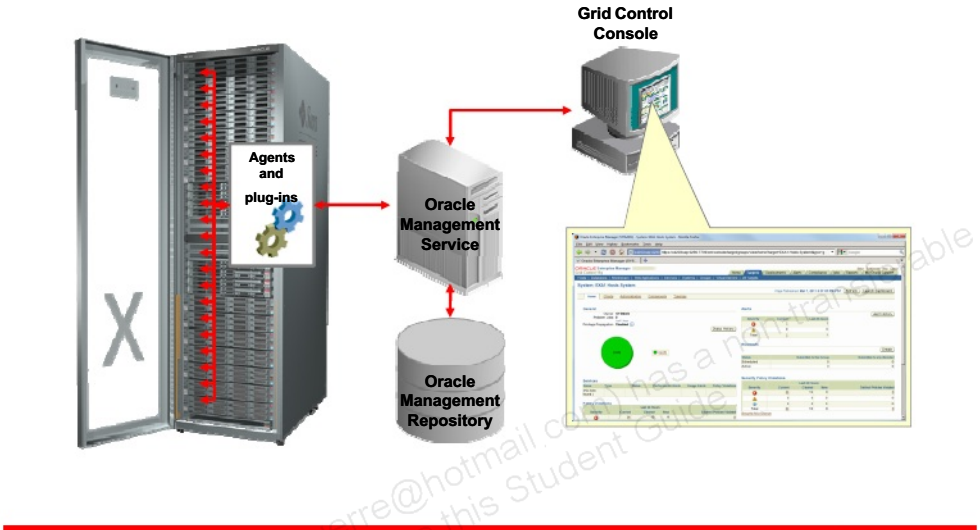
- What is it?
  - A file-based repository for diagnostic data such as traces, dumps and logs
- What does it do?
  - Provides a consistent organization for diagnostic data which enables administrators and Oracle Support to correlate and analyze diagnostic data more easily and effectively
- Where is it found?
  - Exadata Database Machine database servers and Exadata Storage Servers

The Automatic Diagnostic Repository (ADR) is a file-based repository for diagnostic data such as traces, dumps, the alert log, health monitor reports, and more. It has a unified directory structure across multiple products and multiple instances. Beginning with Release 11 *g*, the database, Automatic Storage Management (ASM), listener, and other Oracle products store diagnostic data in the ADR. Exadata Storage Server also uses the ADR structure to organize its diagnostic data.

Each instance of each product stores diagnostic data under its own home directory within the ADR. For example, in an Oracle Real Application Clusters environment with shared storage and Oracle ASM, each database instance and each ASM instance have separate ADR home directories.

ADR's unified directory structure, consistent diagnostic data formats across products and instances, and a unified set of tools enable customers and Oracle Support to efficiently correlate and analyze diagnostic data across multiple instances.

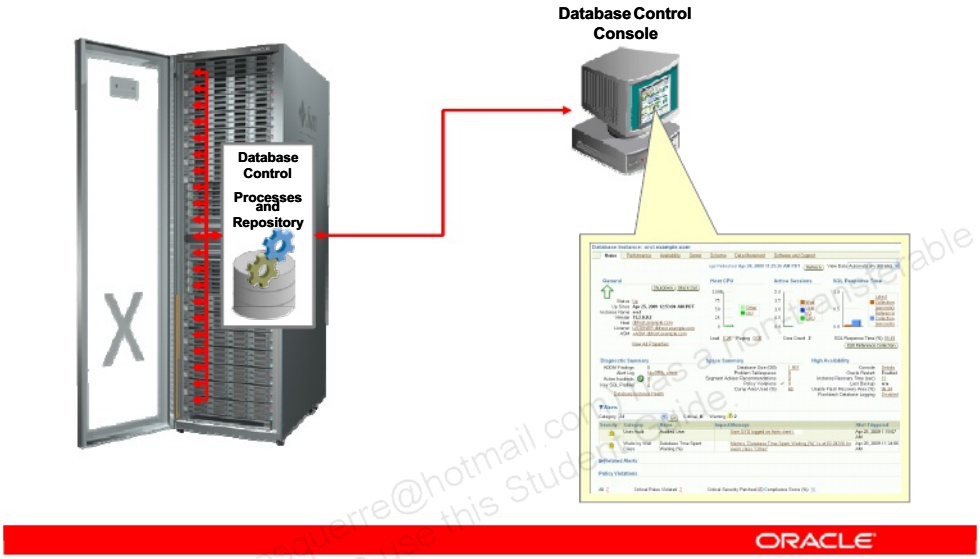
Enterprise Manager Grid Control



Enterprise Manager Grid Control is a system management software platform that delivers centralized monitoring, administration, and life cycle management functionality for information technology infrastructure, including systems running Oracle and non-Oracle technologies.

When used in conjunction with Exadata Database Machine, Enterprise Manager Grid Control should be configured with a set of system-specific plug-ins to enable monitoring of the various Database Machine components. Additionally, there are best practice recommendations regarding how to configure Database Machine targets inside Grid Control in order to facilitate monitoring of Database Machine as a unified whole rather than as a collection of discrete components. The next lesson in this course describes how to configure Grid Control for monitoring Exadata Database Machine.

Enterprise Manager Database Control



Enterprise Manager Database Control is a database administration environment built into Oracle Database. Database Control, can perform administrative tasks such as creating schema objects (tables, views, indexes, and so on), managing user security, managing database memory and storage, performing database backup and recovery, and bulk data movement (export and import). Performance and status information about the database can also be viewed.

In addition to its core database administration functions, Database Control also provides monitoring and management facilities for Exadata Storage Servers. However, Database Control does not provide the breadth of monitoring capabilities offered by Grid Control. For this reason, Grid Control is the recommended monitoring environment for Exadata Database Machine. In addition, Database Control is not considered further during this course.

Quiz

Which of the following monitoring infrastructure technologies can be used in conjunction with Exadata Storage Servers:

a. Simple Network Management Protocol (SNMP)
b. Integrated Lights Out Manager (ILOM)
c. Exadata metrics, thresholds and alerts
d. Automatic Diagnostic Repository (ADR)
e. Enterprise Manager Grid Control

Answer: a, b, c, d, e

Summary

In this lesson, you should have learned how to describe the key monitoring infrastructure technologies associated with Exadata Database Machine.

14

Configuring Enterprise Manager Grid Control 11*g* to Monitor Exadata Database Machine

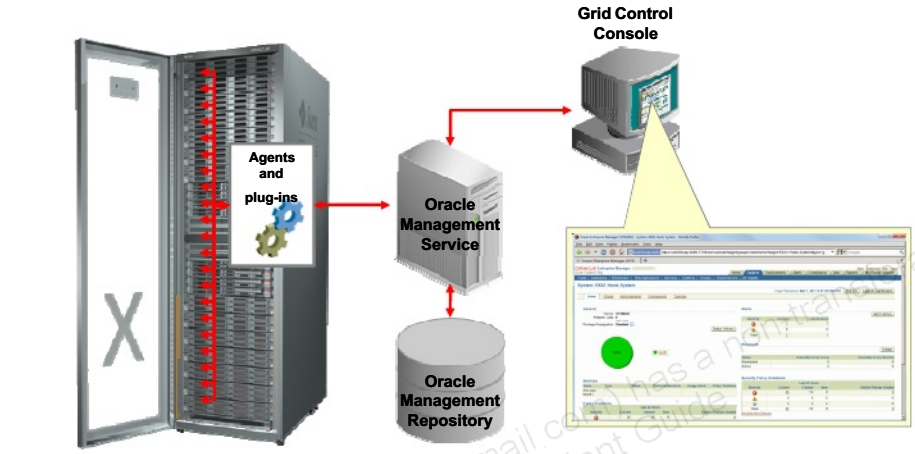
ORACLE

Objectives

After completing this lesson, you should be able to:

- Describe the Enterprise Manager Grid Control architecture as it specifically applies to Exadata Database Machine.
- Describe the placement of agents, plug-ins and targets
- Describe the recommended configuration for high availability
- Describe the plug-ins associated with Exadata Database Machine and how they are configured.
- Describe how to configure a Dashboard for Exadata Database Machine.

Enterprise Manager Grid Control Architecture Overview

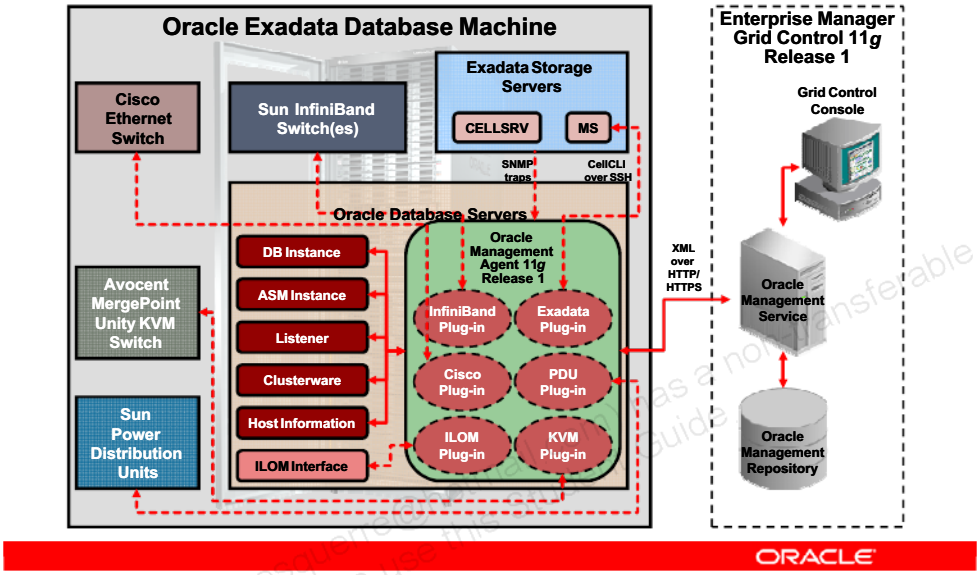


Enterprise Manager Grid Control contains four major elements:

1. Oracle Management Service (OMS) is the processing heart of the system.
2. Oracle Management Repository is the persistent store of monitoring, management and configuration information.
3. Grid Control Console provides a web-based management interface.
4. Agents and plug-ins are deployed to managed targets to collect information which is processed by the OMS.

This fundamental architecture does not change when Grid Control is used to monitor Exadata Database Machine.

Grid Control Monitoring Architecture for Exadata Database Machine



When Enterprise Manager Grid Control is used to monitor Exadata Database Machine, an Oracle Management Agent should be deployed to every database server. The agent is used to monitor and maintain Oracle software targets on the database server including database instances, ASM instances, listeners, clusterware resources and other host information.

In addition, a series of plug-ins are available which extend Grid Control to enable monitoring of other Database Machine components such as the Exadata Storage Servers, InfiniBand switches, Ethernet switch, KVM switch and Power Distribution Units (PDUs).

Each plug-in connects to its monitoring targets using different methods. For example, the Exadata Storage Server plug-in gathers monitoring data from the cells by making CellCLI calls over SSH, however Exadata cell alerts are delivered directly to the agent using SNMP traps. Each plug-in is described in greater detail later in this lesson.

Note that some of the plug-ins only support Enterprise Manager Grid Control 11 *g* Release 1 (or later) and Oracle Management Agent 11 *g* Release 1 (or later). Hence these are the recommended minimum versions to use in conjunction with Database Machine.

Configuring Grid Control to Monitor Exadata Database Machine

1. Deploy Oracle Management Agent to all database servers
2. Configure ASM and Oracle Database targets
3. Deploy the system monitoring plug-ins for Database Machine
   - Exadata Storage Server
   - Integrated Lights Out Manager (ILOM)
   - InfiniBand Switch(es)
   - Cisco Switch
   - Avocent MergePoint Unity (KVM) Switch
   - Power Distribution Units
4. Configure user-defined metrics for additional network monitoring
5. Configure the plug-ins for high availability
6. Configure an Exadata Database Machine aggregate service and dashboard

The list in the slide outlines the recommended steps to configure Enterprise Manager Grid Control for monitoring Exadata Database Machine. The assumed starting point is after the initial installation and configuration of both Database Machine and Grid Control.

The remainder of this lesson describes each step in greater detail.

Recorded demonstrations showing an example of each process step are also provided in conjunction with this lesson. A full list of the recorded demonstrations is provided in the additional resources section at the end of the lesson.

Deploying the Oracle Management Agent

- Where to deploy:
  - All Database Machine database servers
- Recommended version:
  - 11*g* Release 1 (11.1) or later
- How to deploy:
  - Grid Control Agent Installation Wizard
  - See demonstration Agent Installation and Configuration
  - Manual installation of some OS packages may be required

It is recommended to install Oracle Management Agent 11*g* Release 1 (or later) on all the Database Machine database servers.

Assuming that Enterprise Manager Grid Control 11*g* Release 1 is already running and there is network connectivity between it and the Database Machine, the easiest way to deploy the agent is to use the Grid Control agent installation wizard. An example of this process is provided in the demonstration entitled *Agent Installation and Configuration*.

Some earlier versions of the Database Machine database servers were shipped without all the packages required by the Oracle Management Agent. It is acceptable to modify the database servers to install the missing packages.

In the event that the Oracle Management Agent fails to install as a result of missing packages, install the required packages and retry the agent installation. A list of packages required for the agent is provided in the Enterprise Manager documentation.

Configuring ASM and Oracle Database Targets

- Management targets are automatically discovered after the Oracle Management Agent is started
- ASM instances and Oracle databases are discovered but are listed as unconfigured
- Administrators must supply monitoring credentials to complete configuration
- The procedures are essentially the same regardless of whether or not Exadata Database Machine is involved
- See demonstration Configuring ASM and Database Targets

After the Oracle Management Agent is installed and running on each database server, management targets are automatically discovered. This includes ASM instances, cluster databases, database instances, clusters, clusterware resources, network listeners, host information and of course the agents themselves.

After the initial target discovery phase, the ASM and Oracle Database instances are discovered but are listed as unconfigured. To complete the configuration of the ASM and Oracle Database instances, administrators must supply monitoring credentials.

For each ASM instance, connectivity between Grid Control and the ASM instance must be configured using the credentials of an administrative user having either the `sysdba` or `sysasm` role.

For each database, the `dbsnmp` user account must be enabled and its password must be supplied to enable connectivity from Grid Control. Note that to enable the `dbsnmp` account requires `sysdba` access to the database.

Note that the procedure for configuring the ASM instance targets is the same regardless of whether or not the instances reside on Exadata Database Machine. Likewise, the same procedure is used to configure database instance targets on Database Machine as on other clustered database environments.

An example of how to configure the ASM and database targets is provided in the demonstration entitled *Configuring ASM and Database Targets*.

Deployment Overview for System Monitoring Plug-ins

- Plug-ins extend Grid Control to enable monitoring of additional hardware and software components
- Each plug-in contains:
  - Logic that defines how Grid Control communicates with the component and what metrics are monitored
  - User Interface definitions that extend the Grid Control Console
- Plug-in implementation process:
  1. Perform prerequisites
  2. Import the plug-in into the Grid Control environment
  3. Deploy the plug-in to one or more agents
  4. Add a target for each instance of the component being monitored

System monitoring plug-ins extend Enterprise Manager Grid Control, enabling it to monitor additional hardware and software component types. For Exadata Database Machine, a pack of suitable plug-ins is available. Each plug-in contains logic which enables a management agent to communicate with the type of system component being monitored. It also contains user interface elements which extend the Grid Control Console and provide administrators with a view of the information associated with the plug-in.

The implementation process for each plug-in contains the following steps:

1. Perform prerequisites. Because plug-ins communicate with different system components in different ways, they each have different prerequisites which must be configured. The prerequisites are documented in the Installation Guide which accompanies each plug-in.
2. Import the plug-in. Plug-ins are packaged as Java archives (`jar` files). Before a plug-in can be used, its package must be imported into the Grid Control environment.
3. Deploy the plug-in. Deployment configures the plug-in on one or more agents. This enables the agent to communicate with the type of hardware or software component associated with the plug-in.
4. Add targets. After a plug-in is deployed, each component instance must be registered as a target.

Details and additional information for each plug-in follows later in this lesson.

Recommended Plug-in Deployment Strategy

- Recommendation:
  - Associate all plug-in targets with a primary agent on one of the Database Machine database servers
  - Configure target failover using a secondary agent on another database server for high availability
- Rationale:
  - At any time, all plug-in targets are associated with one agent which simplifies management and problem diagnosis
  - Automatic target failover ensures continuation of monitoring if the primary agent is down
  - The performance impact of monitoring is low and confined to one database server
  - Storage Server monitoring can leverage the InfiniBand interconnect

Subject to version requirements and network connectivity, the various Exadata monitoring plug-ins can be deployed to any management agents associated with the Grid Control environment. For example, it is possible (though not recommended) for the Exadata Storage Server plug-in to be deployed against an agent running on the OMS server.

The best practice recommendation for Database Machine plug-in deployments is to deploy all the plug-ins to at least two agents running on different Database Machine database servers. Administrators are advised to configure all the plug-in targets so that they are all initially associated with just one of the management agents, the primary agent. After the initial round of targets is configured, there is a recommended process to configure the plug-ins for high availability. The resulting configuration automatically migrates all the plug-in targets to a nominated secondary agent in cases where the primary agent goes down. The slide lists the rationale behind this recommendation. Configuring the plug-ins for high availability is described later in this lesson.

The plug-in configuration examples and descriptions which follow throughout this lesson are all based on the above recommendation. Customers that choose an alternative plug-in deployment strategy will need to adjust their configurations accordingly.

Deploying the System Monitoring Plug-in for Oracle Exadata Storage Server

- Prerequisites:
  - Preferred credentials for the agent targets
  - SSH user equivalence between the agent user and the `cellmonitor` user on all the cells
- Recommendations:
  - Configure a realm to enable aggregated reporting

```
CellCLI> alter cell realmName = <Realm Name>
```

  - Configure the cells to deliver alerts to Grid Control using SNMP
- Demonstration:
  - Configuring the Exadata Storage Server Plug-in

The following prerequisites are required to successfully deploy the System Monitoring Plug-in for Oracle Exadata Storage Server:

- Preferred credentials must be set in the Grid Control preferences. The host username and password for the user running the Management Agent must be specified. Preferred credentials must be set for each agent host where the plug-in is deployed.
  Note that Agent Preferred Credentials are also required to successfully deploy other system monitoring plug-ins. Typically, Preferred Credential settings are made once prior to deploying the first plug-in and are reused for subsequent plug-ins.
- The storage server plug-in communicates with Exadata cells using the `cellmonitor` user account on the Exadata Storage Servers. SSH user equivalence must be configured between the user running the Management Agent on the agent host (typically `oracle`) and the `cellmonitor` user on all the cells that will be monitored by the plug-in. Because the `cellmonitor` user account has limited access to the storage server operating system, a special procedure is required to establish SSH user equivalence.

  This procedure is documented in the Installation Guide that accompanies the storage server plug-in. An example is also shown in the demonstration that accompanies this topic.

In addition, the following recommendations apply:

- Exadata Storage Servers can be collected into a group known as a realm. The Exadata Storage Server plug-in contains numerous reports that aggregate monitoring information for a realm. Typically all of the cells in one Exadata Database Machine will be grouped into a realm however this is not an absolute requirement. It is recommended that you define realms based on logical groupings that make sense in your environment.
- In conjunction with the plug-in, it is recommended that you configure all your cells to deliver alerts to Grid Control using SNMP. This involves enabling the SNMP notification method on each cell along with adding the agent host name and port as an SNMP subscriber.

An example of how to deploy the storage server plug-in is provided in the demonstration entitled *Configuring the Exadata Storage Server Plug-in*.

Deploying the System Monitoring Plug-in for Oracle Exadata ILOM

- Recommendation:
  - Use the ILOM plug-in only for monitoring ILOM on Database Machine database servers
    - Not required for Exadata Storage Server because ILOM monitoring is already managed by Exadata Storage Server software.
- Prerequisites:
  - `ipmitool` version 1.8.10.1 or later
  - ILOM user credentials
    - A dedicated ILOM user account for the plug-in is highly recommended
- Demonstration:
  - Configuring the ILOM Plug-in

The ILOM plug-in monitors the ILOM service processor in a server. It enables hardware events and sensor data to be monitored using Grid Control. For Database Machine, it is recommended to use the ILOM plug-in only for monitoring the ILOM server processor of the Database Machine database servers. The ILOM plug-in is not required to monitor Exadata Storage Servers since ILOM metrics and alerts are already managed by the Exadata Storage Server software.

The ILOM plug-in requires `ipmitool` software version 1.8.10.1 or later to be present on the servers being monitored by the plug-in.

Though not an absolute requirement, it is highly recommended to create a specific ILOM user account for the ILOM plug-in to use. The procedure is documented in the Installation Guide that accompanies the ILOM plug-in.

An example of how to deploy the ILOM plug-in is provided in the demonstration entitled *Configuring the ILOM Plug-in*.

Deploying the System Monitoring Plug-in for Oracle Exadata InfiniBand Switch

- Prerequisites:
  - Switch firmware version 1.1.3-2 or later
  - SNMP configuration:
    - Enable SNMP
    - Set SNMP protocol to `v1` (version 1)
    - Set the SNMP community string if you plan to use an SNMP community string other than the default (`public`)
- Demonstration:
  - Configuring the InfiniBand Switch Plug-in

The following prerequisites are required to successfully deploy the System Monitoring Plug-in for Oracle Exadata InfiniBand Switch:

- The InfiniBand Switch must be running firmware version 1.1.3-2 or later.
- SNMP must be configured on the InfiniBand Switch to allow the plug-in to obtain its information. Here is a summary of the required settings:
  - SNMP must be enabled
  - The SNMP protocol must be set to `v1` (version 1)
  - If you plan to use an SNMP community string other than the default (`public`), then you must make the appropriate setting and make sure you use the same setting when you later create the InfiniBand switch targets.

An example of how to deploy the InfiniBand switch plug-in is provided in the demonstration entitled *Configuring the InfiniBand Switch Plug-in*.

Deploying the System Monitoring Plug-in for Exadata Cisco Switch

- Prerequisites:
  - Switch configuration:
    - Enable access to allow the plug-in to poll the switch
    - Configure the switch to deliver SNMP traps to the plug-in host
    - Configure the switch to send only environmental monitor SNMP traps
    - Verify and save the settings
  - An SNMP trap forwarder on each plug-in host to catch Cisco switch traps and forward them to the Management Agent
- Demonstration:
  - Configuring the Cisco Ethernet Switch Plug-in

The following prerequisites are required to successfully deploy the System Monitoring Plug-in for Exadata Cisco Switch:

- The Cisco switch must be configured to enable SNMP communications between the switch and the plug-in. The slide provides and overview of the required configuration steps. Note that the Cisco Ethernet switch provided with Database Machine does not include SSH support. Oracle recommends connecting to the switch management interface using an alternative such as telnet.
- The Cisco switch has no documented way to send SNMP traps to ports other than UDP 162. However the Oracle Management Agent runs as the `oracle` user and cannot listen on UDP ports under 1024. A trap forwarder must be configured so that the SNMP traps generated by the switch are forwarded to the Oracle Management Agent.

Detailed instructions for both prerequisite tasks are provided in the Installation Guide that accompanies the plug-in. An example of how to deploy the Cisco Ethernet switch plug-in is provided in the demonstration entitled *Configuring the Cisco Ethernet Switch Plug-in*.

Deploying the System Monitoring Plug-in for Avocent MergePoint Unity (KVM) Switch

- Prerequisites:
  - Switch software version 1.2.8 or later
  - Switch configuration:
    - Confirm that SNMP is enabled on the KVM switch
    - Configure the switch to deliver SNMP traps to the plug-in host
  - An SNMP trap forwarder on each plug-in host to catch the KVM traps and forward them to the Management Agent
    - Same configuration as required for the Cisco switch plug-in
- Demonstration:
  - Configuring the Avocent KVM Switch Plug-in

The following prerequisites are required to successfully deploy the KVM switch plug-in:

- The KVM plug-in requires that the application software version on the KVM switch is 1.2.8 or later. Confirm you have the required software version.
- Confirm that SNMP is enabled on the KVM switch.
- Configure the KVM switch to send SNMP traps to the host where the KVM switch plug-in will be deployed.
- The KVM switch sends SNMP traps to the default port (UDP 162). However the Oracle Management Agent runs as the `oracle` user and cannot listen on UDP ports under 1024. Configure a trap forwarder to send SNMP traps generated by the KVM on to the Oracle Management Agent. If you have already configured a trap forwarder for the Cisco switch plug-in, no additional configuration is required.

Detailed instructions for all prerequisite tasks are provided in the Installation Guide that accompanies the plug-in. An example of how to deploy the KVM switch plug-in is provided in the demonstration entitled *Configuring the Avocent KVM Switch Plug-in*. An example showing the required SNMP trap forwarder configuration is contained in the Cisco switch plug-in deployment demonstration.

Deploying the System Monitoring Plug-in for Oracle Exadata Power Distribution Unit

- Prerequisites:
  - PDU connected and configured for management network access
  - PDU firmware version 1.02 or later
  - SNMP enabled on PDU
  - PDU configured to communicate with the plug-in host
  - PDU alarm thresholds set. Settings based on:
    - Database Machine Version:
      - Exadata Database Machine X2-2 (using X4170 and X4275 servers)
      - Exadata Database Machine X2-2 (using X4170 M2 and X4270 M2 servers)
      - Exadata Database Machine X2-8 (using X4800 and X4270 M2 servers)
    - Database Machine Size:
      - Full Rack, Half Rack or Quarter Rack
    - PDU Type:
      - Low-voltage 15 kVA Single Phase
      - High-voltage 15 kVA Single Phase
      - Low-voltage 24 kVA Three Phase
      - Low-voltage 15 kVA Three Phase
      - High-voltage 15 kVA Three Phase
      - High-voltage 24 kVA Three Phase

The following prerequisites are required to successfully deploy the System Monitoring Plug-in for Oracle Exadata Power Distribution Unit (PDU):

- By default, the PDUs inside Exadata Database Machine are not physically connected to any network. To monitor PDUs using the plug-in, connect each PDU to the management network and configure the PDU network port with IP addressing information. Note that each PDU should be connected directly to your network, not to the Cisco Ethernet switch in the Database Machine rack. This configuration helps to ensure that alerts generated by the PDUs are delivered to Grid Control even if a power problem causes a fault in the Database Machine management network.
- Confirm that the PDU firmware version is 1.02 or later. The current version can be obtained by logging into the web interface of the PDU. On the left side of the screen, click Module Info to view the PDU firmware version.
- Confirm that SNMP is enabled on the PDU and configure the PDU to enable SNMP communications with the host where the PDU plug-in will be deployed.
- Configure the PDU with alarm threshold values. Appropriate threshold values depend on the Oracle Exadata Database Machine version, the Oracle Exadata Database Machine size, and the PDU type used. The *Oracle Exadata Database Machine Owner's Guide* contains the recommended threshold values for each situation.

Refer to the PDU User's Guide and the Installation Guide that accompanies the PDU plug-in for further details.

Configuring User-Defined Metrics for Additional Network Monitoring

- The health of InfiniBand network ports should be monitored on all Database Machine servers
  - InfiniBand monitoring is automatically performed by MS on storage servers
- User Defined Metric (UDM) scripts are available from My Oracle Support note 1110675.1:
  - `emudm_ibconnect.sh` monitors InfiniBand connectivity to other Database Machine servers
  - `emudm_netif_state.sh` monitors network interface state for Ethernet and InfiniBand network interfaces
- See demonstration Configuring User Defined Metrics for Additional Network Monitoring

InfiniBand port monitoring checks the health of InfiniBand network ports and interfaces on database servers and Exadata Storage Servers. InfiniBand port monitoring on storage servers is performed by the Exadata Storage Server Management Server (MS). No additional monitoring is required on storage servers. If a port is not functioning correctly, MS creates an alert, and delivers the alert via SNMP to Grid Control.

By default, Database Machine database servers have no built-in InfiniBand monitoring. Checks using commands such as `ibstatus` and `perfquery` can be run periodically, however this is a manual process. These checks can be automated using User Defined Metrics (UDMs) in Grid Control. There are two UDM scripts provided for database servers:

- `emudm_ibconnect.sh`- This UDM script monitors connectivity over the InfiniBand network to other database servers and storage servers. This list of database servers is built from the `ocrdump SYSTEM.crs.e2eport` key. The list of cells is built from `cellip.ora`. The script does not validate connectivity to additional devices on the InfiniBand network such as media servers.

  Note that the `ibhosts` command is not used because it requires `root` permissions. The approach used in this script is preferred as scope is limited to servers that are relevant to the Exadata Database Machine, and not the whole InfiniBand network. Using this script it is recommended to create one UDM for each Database Machine database server.

- `emudm_netif_state.sh`- This UDM script can monitor the network interface state for Ethernet or InfiniBand interfaces. For InfiniBand interfaces, it checks the underlying InfiniBand port, state, rate, and 3 error counters; `RcvErrors`, `SymbolErrors`, and `LinkIntegrityErrors`. If the interface is the master for Linux bonding, then the state of the slave interfaces is verified. The error counters are reported only when they increase since the last invocation of the script and only when `LinkDowned` has not increased. Previous values are cached in the temporary file `/var/tmp/ibCounterCache_<device><port>`. Oracle recommends using this script to create one UDM for each network interface on each Database Machine database server.

The UDM scripts are available from My Oracle Support note 1110675.1. An example of how to deploy the UDM scripts is provided in the demonstration entitled *Configuring User Defined Metrics for Additional Network Monitoring*.

Configuring Grid Control for Exadata Database Machine: Alternative Approach

- There is an alternative method for rapidly installing Grid Control and the Exadata plug-ins. It can:
  - Install Grid Control on a standalone server
  - Install the management agent and Exadata plug-ins on:
    - A newly established Grid Control installation
    - An existing Grid Control environment with other existing targets
  - Remove the management agent from database servers
  - Remove Grid Control from a standalone server
- The approach is integrated with the Database Machine installation and initial configuration processes
- Linux and Solaris database server OS options are both supported

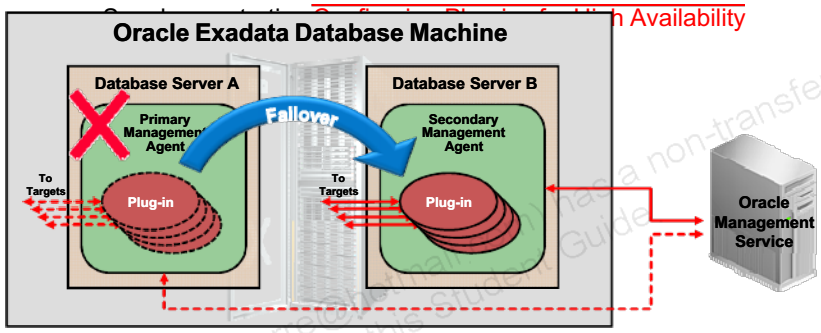
An alternative procedure exists for rapidly installing Grid Control and the Exadata plug-ins. The procedure is integrated with the Database Machine installation and initial configuration process and supports both the Linux and Solaris database server OS options. The procedure is outlined in the updated *Oracle Exadata Database Machine Owner's Guide* and requires patch 11852882.

The procedure provides the ability to install a fresh Grid Control environment on to a standalone server using a one-step installation program called `setupem.sh`. In addition, `setupem.sh` provides functionality to install the Exadata plug-ins on a fresh Grid Control installation or on an existing Grid Control environment which may already be monitoring other targets. This includes performing the prerequisites for each plug-in, importing the plug-in into Grid Control and deploying the plug-in to all of the Database Machine management agents. Removal of the Exadata plug-ins and the Grid Control software can also be performed.

Note that while `setupem.sh` can deploy the Exadata plug-ins, administrators must still configure all the Grid Control targets which represent the various Database Machine components.

Configuring the Plug-Ins for High Availability

- Normally a plug-in target is bound to a specific agent
  - If the agent is down the target cannot be monitored
- A procedure exists to facilitate target failover to a secondary agent
  - See My Oracle Support note 1110675.1



By default, a target monitored using a plug-in is bound to the agent used during initial deployment. A procedure exists which allows the monitoring of a plug-in target to be automatically moved to a predefined secondary agent in cases where the primary agent is down. This is done by configuring each target to have a default host and a failover host.

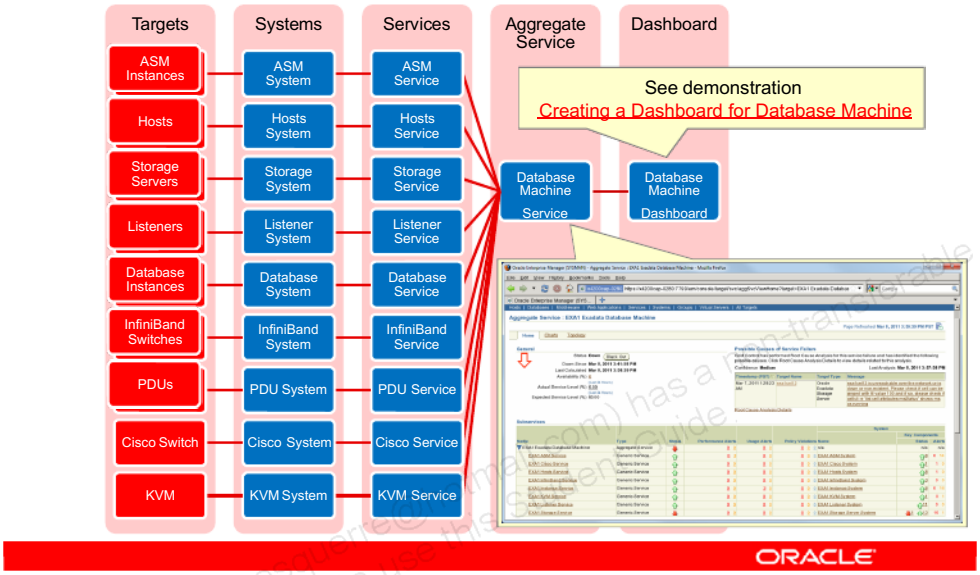
When the agent on the default host, the primary management agent, becomes unavailable for any reason, monitoring of the affected targets is automatically moved to the agent on the failover host, the secondary management agent. When the primary management agent becomes available again, the targets are automatically moved back to it.

To enable target failover and failback you must install the failover extension PL/SQL package into the management repository. The failover configuration process is described in the document entitled *Installation of plug-ins and dashboard configuration*. This document and the failover extension are available from My Oracle Support note 1110675.1.

Note that when plug-ins are configured for high availability, the prerequisites for each plug-in must be met for both the primary agent host and the secondary agent host. For example, you must define Agent Preferred Credentials for both the primary and secondary management agent hosts. Also, you must deploy the plug-ins to both hosts.

An example of how to configure the plug-ins for high availability is provided in the demonstration entitled *Configuring Plug-ins for High Availability*.

Configuring an Exadata Database Machine Aggregate Service and Dashboard



After all the Exadata system monitoring plug-ins are installed and configured, Grid Control can be used to monitor all the Exadata Database Machine components (targets). For a Full-Rack Database Machine there are typically more than 60 targets.

Naturally, administrators prefer to monitor Database Machine as a single entity rather than 60 separate components. A recommended approach has been developed to facilitate monitoring of Database Machine as a single entity. It is based on best-practice fundamentals however there is scope in some areas for customization to suit different needs. The diagram in the slide illustrates the resulting hierarchy of targets inside Grid Control. Following is an outline of the high-level methodology involved:

1. For each Database Machine target type in Grid Control, create a collection containing all the targets of that type using the System target type. For example, group the ASM instances into one System, group the Exadata Storage Servers into another System, and so on.
2. Create a Service based on each of the previously created Systems. A Service is an abstraction of a System that can have administrator-defined performance and usage metrics associated with it. Different performance and usage metrics can be created to suit different situations. For example, an administrator of a Data Warehouse might be interested in I/O throughput as a key measure of performance while an administrator of an OLTP database might be more interested in an average response time indicator.

3. Create an Aggregated Service to group all the Services into a single Grid Control entity. The screen shot on the slide shows an example of the home page for an Aggregate Service in Grid Control.
4. Create a monitoring Dashboard based on the Aggregate Service. The Dashboard provides a single screen overview of the health, performance and usage of the Database Machine. The Dashboard can be customized to suit varying requirements by displaying different performance and usage metrics.

The process outlined above is detailed in the document entitled *Installation of plug-ins and dashboard configuration*. This document is available from My Oracle Support note 1110675.1. An example of how to configure a dashboard for Exadata Database Machine is provided in the demonstration entitled *Creating a Dashboard for Database Machine*.

Quiz

What is the recommended implementation approach for Exadata Database Machine system monitoring plug-ins?

a. Configure all the plug-in targets using different agents to distribute the monitoring load

b. Configure each plug-in target using all the available agents for maximum availability

c. Configure all the plug-in targets using just one agent for the sake of simplicity

d. Configure all the plug-in targets using one primary agent and configure target failover to a secondary agent to facilitate high availability and simplify deployment

Answer: d

Quiz

You can customize the Exadata Database Machine target in Enterprise Manager Grid Control 11*g* by displaying different performance and usage metrics:

a. TRUE

b. FALSE

Answer: b

There is no Enterprise Manager Grid Control 11*g* target type for Database Machine. Rather you aggregate the component targets into Systems and Services to provide a consolidated view of the entire Database Machine.

Summary

In this lesson, you should have learned how to:

- Describe the Enterprise Manager Grid Control architecture as it specifically applies to Exadata Database Machine.
- Describe the placement of agents, plug-ins and targets
- Describe the recommended configuration for high availability
- Describe the plug-ins associated with Exadata Database Machine and how they are configured.
- Describe how to configure a Dashboard for Exadata Database Machine.

Additional Resources

- Lesson Demonstrations
  - Agent Installation and Configuration
  - Configuring ASM and Database Targets
  - Configuring the Exadata Storage Server Plug-in
  - Configuring the ILOM Plug-in
  - Configuring the InfiniBand Switch Plug-in
  - Configuring the Cisco Ethernet Switch Plug-in
  - Configuring the Avocent KVM Switch Plug-in
  - Configuring User Defined Metrics for Additional Network Monitoring
  - Configuring Plug-ins for High Availability
  - Creating a Dashboard for Database Machine
- My Oracle Support Notes
  - Oracle Database Machine Monitoring Best Practices

15

Monitoring Exadata Storage Servers

ORACLE

Objectives

After completing this lesson, you should be able to:

- Describe Exadata Storage Server metrics, alerts and active requests
- Identify the recommended focus areas for Exadata Storage Server monitoring
- Describe how to monitor the recommended Exadata Storage Server focus areas

Lesson Overview

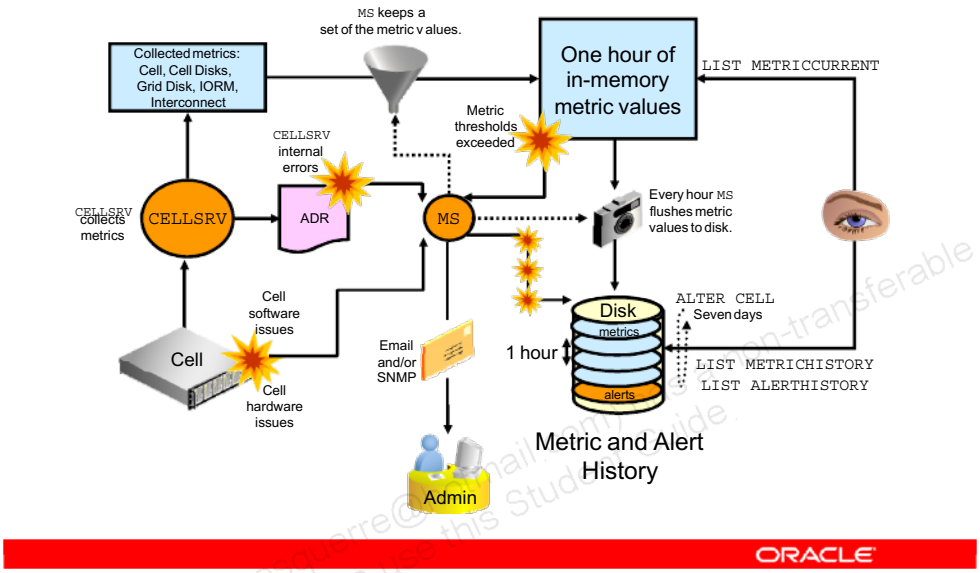
- Exadata Metric and Alerts Architecture and Overview
- Monitoring Exadata Storage Server with Grid Control Overview
- Monitoring Hardware Failure and Sensor State
- Monitoring Exadata Storage Server Availability
- Checking for Undelivered Alerts
- Checking for Disk I/O Errors
- Checking for Network Errors
- Monitoring Filesystem Free Space
- Comparing Metrics Across Multiple Storage Servers
- Monitoring Metrics Within a Storage Server
- Third Party Monitoring Tools

What to Monitor

This lesson focuses on best practice recommendations for Exadata Storage Server regarding what should be monitored from a general administrative perspective, and how it should be monitored.

The list in the slide summarizes the topics contained in the lesson. The shaded area highlights the recommended areas to monitor.

Exadata Metrics and Alerts Architecture



The diagram in the slide illustrates the architecture for Exadata Storage Server metrics and alerts. It shows how CELLSRV periodically records important run-time properties, called metrics, for cell components such as CPUs, cell disks, grid disks, flash cache, and IORM statistics. These metrics are recorded in memory. Based on its own metric collection schedule, the Management Server (MS) gets the set of metric data accumulated by CELLSRV. MS keeps a subset of the metric values in memory, and writes a history to an internal disk-based repository every hour. This process is conceptually similar to database AWR snapshots.

The retention period for metric and alert history entries is specified by the `metricHistoryDays` cell attribute. You can modify this setting with the CellCLI `ALTER CELL` command. By default, it is seven days. You can view the metric value history by using the CellCLI `LIST METRICHISTORY` command, and you can view the current metric values by using the `LIST METRICCURRENT` command.

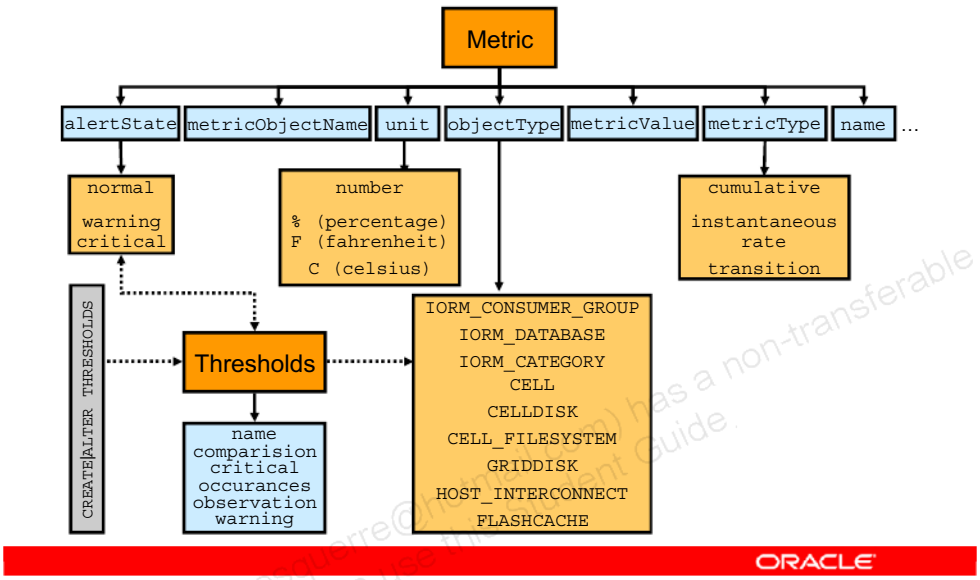
At the Exadata cell level, you can define thresholds for a selection of metrics prior to Exadata Storage Server software release 11.2.2.3.1. Commencing with Exadata Storage Server software release 11.2.2.3.1, thresholds can be defined for all cell metrics.

In addition to collecting metrics, Exadata Storage Server can generate alerts. Alerts represent events of importance occurring within the cell, often indicating that an Exadata cell function is compromised. MS generates an alert when it discovers a:

- Cell hardware issue
- Cell software or configuration issue
- CELLSRV internal error
- Metric that has exceeded a threshold defined in the cell

You can view previously generated alerts using the `LIST ALERTHISTORY` command. In addition, you can configure the cell to automatically send an email and/or SNMP message to a designated set of administrators.

Monitoring Exadata Storage Server with Metrics



Metrics are recorded observations of important run-time properties or internal instrumentation values of the storage cell and its components, such as cell disks or grid disks. Metrics are a series of measurements that are computed and retained in memory for a period of time, and stored on a disk for a more permanent history.

The graphic in the slide describes some of the important metric attributes. Each metric:

- Has a name and description
- Is associated with a `metricObjectName` that is the name of the object being measured, such as a specific cell disk, grid disk, or consumer group
- Belongs to a group that is defined by its `objectType` attribute. The possible groups are shown in the slide.
- Has a `metricType`, which is an indicator of how the statistic was created or defined. Possible values and their meanings are:
  - `cumulative`: Cumulative statistics since the metric was created
  - `instantaneous`: Value at the time that the metric is collected
  - `rate`: Rates computed by averaging statistics over observation periods
  - `transition`: Are collected at the time when the value of the metrics has changed, and typically captures important transitions in hardware status
- Has a measurement `unit`. Possible units are shown in the slide.

Understanding the composition of the metric name provides a good insight into the meaning of the metric. The value of the name attribute is a composite of abbreviations. The attribute value starts with an abbreviation of the object type on which the metric is defined:

- CL_ (cell)
- CD_ (cell disk)
- GD_ (grid disk)
- FC_ (flash cache)
- DB_ (database)
- CG_ (consumer group)
- CT_ (category)
- N_ (interconnect network)

After the abbreviation of the object type, many metric names conclude with an abbreviation that relates to the description of the metric. For example, CL_FANS is the instantaneous number of working fans on the cell.

I/O-related metric name attributes continue with one of the following combinations to identify the operation:

- IO_RQ (number of requests)
- IO_BY (number of MB)
- IO_TM (I/O latency)
- IO_WT (I/O wait time)

Next in the name could be _R for read or _W for write. Following that, there might be _SM or _LG to identify small or large I/Os, respectively. At the end of the name, there could be _SEC to signify per second or _RQ to signify per request. For example:

- CD_IO_RQ_R_SM is the number of requests to read small blocks on a cell disk.
- GD_IO_BY_W_LG_SEC is the number of MB of large block I/O per second on a grid disk.

If a metric value crosses a user-defined threshold, an alert will be generated. Metrics can be associated with warning and critical thresholds. Thresholds relate to extreme values in the metric, which might indicate a problem or other event of interest to an administrator.

Thresholds are supported on cell disk and grid disk I/O error count metrics (CD_IO_ERRS_MIN and GD_IO_ERRS_MIN), along with the cell memory utilization (CL_MEMUT) and cell filesystem utilization (CL_FSUT) metrics. In addition, you can set thresholds for I/O Resource Management (IORM) related metrics. The CellCLI LIST ALERTDEFINITION command lists the metrics for which thresholds can be set.

For further details regarding Exadata cell metric and threshold attributes, refer to the *Oracle Exadata Storage Server Software User's Guide*.

Monitoring Exadata Cell Metrics: Examples

```
CellCLI> LIST METRICDEFINITION WHERE objectType ='CELL' DETAIL
         name: CL_CPUT
         description: "Cell CPU Utilization is the percentage of time over
                       the previous minute that the system CPUs were not
                       idle (from /proc/stat). "
         metricType: Instantaneous objectType: CELL
         unit: %
         ...
```

```
CellCLI> LIST METRICHISTORY WHERE name like 'CL_.*' -
         AND collectionTime > '2009-10-11T15:28:36-07:00'
         CL_RUNQ    cell03_2    6.0       2009-10-11T15:28:37-07:00
         CL_CPUT    cell03_2    47.6 %    2009-10-11T15:29:36-07:00
         CL_FANS    cell03_2    1         2009-10-11T15:29:36-07:00
         CL_TEMP    cell03_2    0.0 C     2009-10-11T15:29:36-07:00
         CL_RUNQ    cell03_2    5.2       2009-10-11T15:29:37-07:00
         ...
```

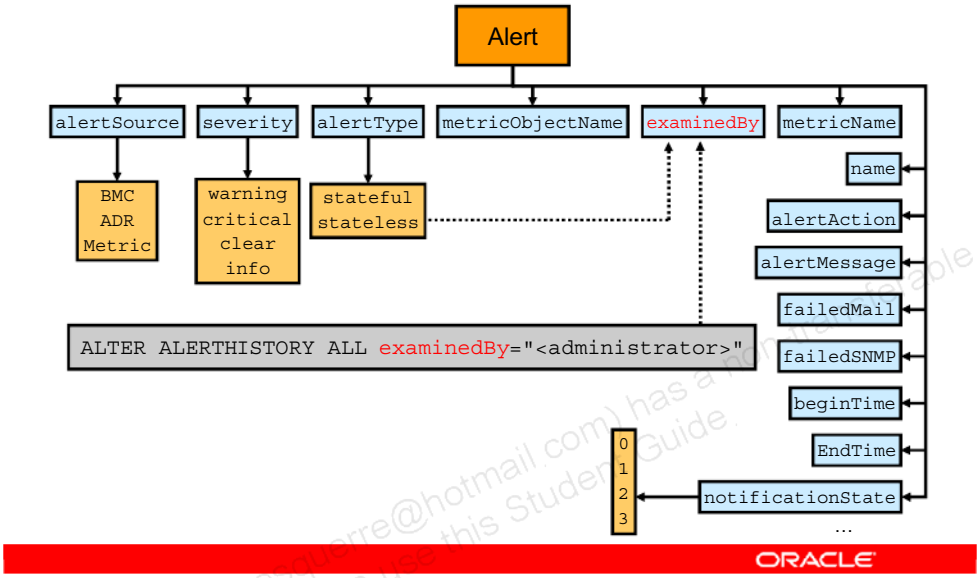
```
CellCLI> LIST METRICCURRENT WHERE objectType = 'CELLDISK'
         CD_IO_TM_W_SM_RQ CD_1_cell03 205.5 us/request
         CD_IO_TM_W_SM_RQ CD_2_cell03  93.3 us/request
         CD_IO_TM_W_SM_RQ CD_3_cell03   0.0 us/request
         ...
```

ORACLE

The slide shows some basic commands that you could use to display metric information:

- Use the `LIST METRICDEFINITION` command to display the metric definitions for the cell. A metric definition describes the configuration of a metric. The example does not specify any particular metric, so all metrics corresponding to the `WHERE` clause are printed. In addition to the `WHERE` clause, you can also specify the metric definition attributes you want to print. If the `ATTRIBUTES` clause is not used, a default set of attributes is displayed. To list all the attributes, you can add the `DETAIL` keyword at the end of the command.
- Use the `LIST METRICHISTORY` command to display the metric history for the cell. A metric history describes a collection of past metric observations. Similar to the `LIST METRICDEFINITION` command, you can specify attribute filters, an attribute list, and the `DETAIL` keyword for the `LIST METRICHISTORY` command. The above example lists metrics having names that start with `CL_` that were collected after the specified time.
- Use the `LIST METRICCURRENT` command to display the current metric values for the cell. The above example lists all cell disk metrics. The metric values shown in the slide correspond to the average latency per request of writing small blocks to a cell disk. For this metric there is a metric observation for every cell disk.

Monitoring Exadata Storage Server with Alerts



Exadata Storage Server alerts represent events of importance occurring within the storage cell, typically indicating that storage cell functionality is either compromised or in danger of failure. An administrator should investigate alerts, because they might require urgent corrective or preventative action. Notification of cell alerts using email or SNMP, can be configured during the initial configuration of Database Machine. An existing configuration can be adjusted, or a new configuration can be established, at any time using the `ALTER CELL` CellCLI command.

Cell alerts can be stateful or stateless. Stateful alerts represent observable cell states that can be subsequently retested to detect whether the state has changed, indicating that a previously observed alert condition is no longer a problem. Stateless alerts represent point-in-time events that do not represent a persistent condition; they simply show that something occurred.

Alerts can have one of the following severities: `warning`, `critical`, `clear`, or `info`. Examples of possible events that trigger alerts are physical disk failure, disk read/write errors, cell temperature exceeding recommended value, cell software failure, and excessive I/O latency.

Metrics can be used to signal stateful alerts using `warning` or `critical` threshold values. When the metric value crosses the threshold value, an alert is signaled. An alert with a `clear` severity indicates that a previous critical or warning condition has returned to normal. For threshold-based alerts, a `clear` alert is generated when the measured value crosses back over the threshold value.

Alerts with an `info` severity are stateless and log conditions that might be informative to an administrator but for which no administrator action is required. Informational alerts are not distributed by email or SNMP notifications.

The slide illustrates some of the important alert attributes. Each alert has the following attributes:

- `name` provides an identifier for the alert.
- `alertSource` provides the source of the alert. Some possible sources are listed in the slide.
- `severity` determines the importance of the alert. Possible values are `warning`, `critical`, `clear`, and `info`.
- `alertType` provides the type of the alert: `stateful` or `stateless`. Stateful alerts are automatically cleared on transition to normal. Stateless alerts are never cleared unless you change the alert by setting the `examinedBy` attribute. This attribute identifies the administrator who reviewed the alert and is the only alert attribute that can be modified by the administrator using the `ALTER ALERTHISTORY` command.
- `metricObjectName` is the object for which a metric threshold has caused an alert.
- `metricName` provides the metric name if the alert is based on a metric.
- `alertAction` is the recommended action to perform for this alert.
- `alertMessage` provides a brief explanation of the alert.
- `failedMail` is the intended email recipient when a notification failed.
- `failedSNMP` is the intended SNMP subscriber when a notification failed.
- `beginTime` provides the timestamp when an alert changes its state.
- `endTime` provides the timestamp for the end of the period when an alert changes its state.
- `notificationState` indicates progress in notifying subscribers to alert messages:
  - `0`: never tried
  - `1`: sent successfully
  - `2`: retrying (up to 5 times)
  - `3`: five failed retries

Note: Some I/O errors may result in an ASM disk going offline without generating a cell alert. You should continue to perform I/O monitoring from your databases and ASM environments to identify and remedy these kinds of problems.

Monitoring Cell Alerts and Creating Thresholds: Examples

```
CellCLI> LIST ALERTDEFINITION ATTRIBUTES name, metricName, description
    ADRAlert "CELL Incident Error"
    HardwareAlert "Hardware Alert"
    StatefulAlert_CG_IO_RQ_LG CG_IO_RQ_LG "Threshold Based Stateful Alert"
    StatefulAlert_CG_IO_RQ_LG_SEC CG_IO_RQ_LG_SEC "Threshold Based …Alert"
    StatefulAlert_CG_IO_RQ_SM CG_IO_RQ_SM "Threshold Based Stateful Alert"
    ...
```

```
CellCLI> LIST ALERTHISTORY WHERE severity = 'critical' -
         AND examinedBy = '' DETAIL
```

```
CellCLI> ALTER ALERTHISTORY 1671443814 examinedBy="JFV"
```

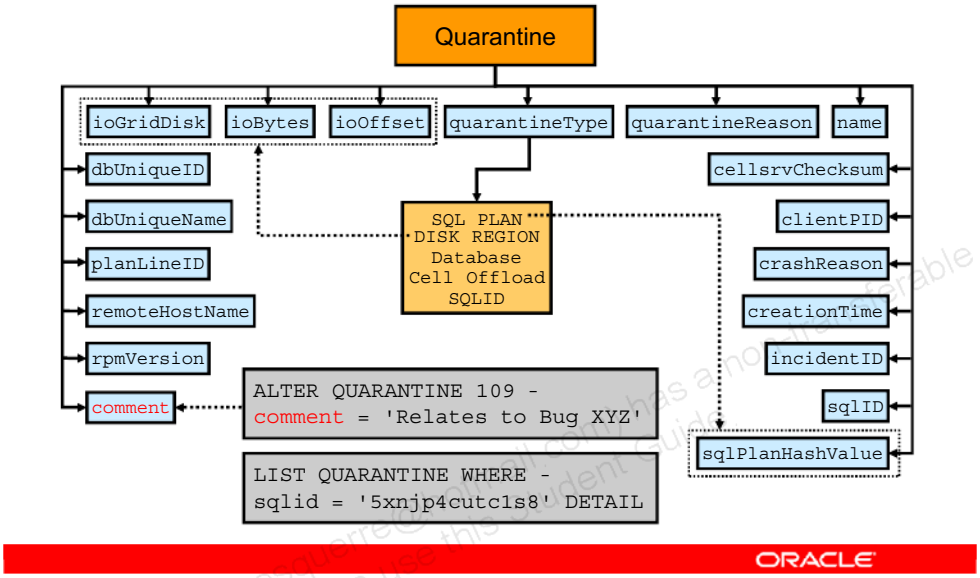
```
CellCLI> CREATE THRESHOLD ct_io_wt_lg_rq.interactive -
         warning=1000, critical=2000, comparison='>', -
         occurrences=2, observation=5
```

The slide shows some examples of CellCLI commands that display alert information. The commands for displaying alerts are very similar to the ones used for displaying metric information:

- Use the `LIST ALERTDEFINITION` command to display the definition for every alert that can be produced on the cell. The example in the slide displays the alert name, metric name, and description. The metric name identifies the metric on which the alert is based. `ADRAlert` and `HardwareAlert` are not based on any metric and, therefore, do not have metric names.
- Use the `LIST ALERTHISTORY` command to display the alert history for a cell. The example in the slide lists in detail all critical alerts that have not been reviewed by an administrator.
- Use the `ALTER ALERTHISTORY` command to update the alert history for the cell. The above example shows how to set the `examinedBy` attribute to the user ID of the administrator that examined the alert. The `examinedBy` attribute is the only `ALERTHISTORY` attribute that can be modified. The example uses the alert sequence ID to identify the alert. `alertSequenceID` provides a unique sequence ID number for the alert. When an alert changes its state, another occurrence of the alert is created with the same sequence number but with a different timestamp.

- The `CREATE THRESHOLD` command creates a threshold that specifies the conditions for generation of a metric alert. The example creates a threshold for the `CT_IO_WT_LG_RQ` metric associated with the `INTERACTIVE` category. This metric specifies the average number of milliseconds that large I/O requests issued by the category have waited to be scheduled by IORM in the past minute. A large value indicates that the I/O workload from this category is exceeding the allocation specified for it in the IORM plan. The alert is triggered by two consecutive measurements (`occurrences=2`) over the threshold values: one second for a warning alert (`warning=1000`) and two seconds for a critical alert (`critical=2000`). The `observation` attribute is the number of measurements over which measured values are averaged.

Isolating Faults with Exadata Storage Server Quarantine



In addition to metrics and alerts, when prescribed faults are detected within Exadata Storage Server, a quarantine object is automatically created. By means of this, the action that caused of the fault can be quarantined so that the fault can be avoided in the future. Quarantine reduces the chance of storage server software crashes, and improves storage availability.

For example, if the cell crashes while performing Smart Scan for a SQL statement, Exadata Storage Server quarantines the SQL statement. Later, when the same SQL statement occurs again, the cell will not allow the SQL statement to use Smart Scan.

The following types of automatic quarantine are available:

- `SQL PLAN`: Created when the cell crashes while performing Smart Scan for a SQL statement. The SQL Plan for the SQL statement is quarantined, and Smart Scan is disabled for the SQL plan.
- `DISK REGION`: Created when the cell crashes while performing Smart Scan of a disk region. The 1 MB disk region being scanned is quarantined and Smart Scan is disabled for the disk region.
- `Database`: Created when the cell detects that a particular database causes instability. Instability detection is based on the number of SQL Plan Quarantines for a database. Smart Scan is disabled for the database.

- `Cell Offload`: Created when the cell detects some offload feature has caused instability. Instability detection is based on the number of Database Quarantines for a cell. Smart Scan is disabled for all databases.

When a quarantine is created, an alert is generated to notify administrators about what was quarantined, why the quarantine was created, when and how the quarantine can be dropped manually, and when the quarantine is dropped automatically. All quarantines are automatically removed when a cell is patched or upgraded.

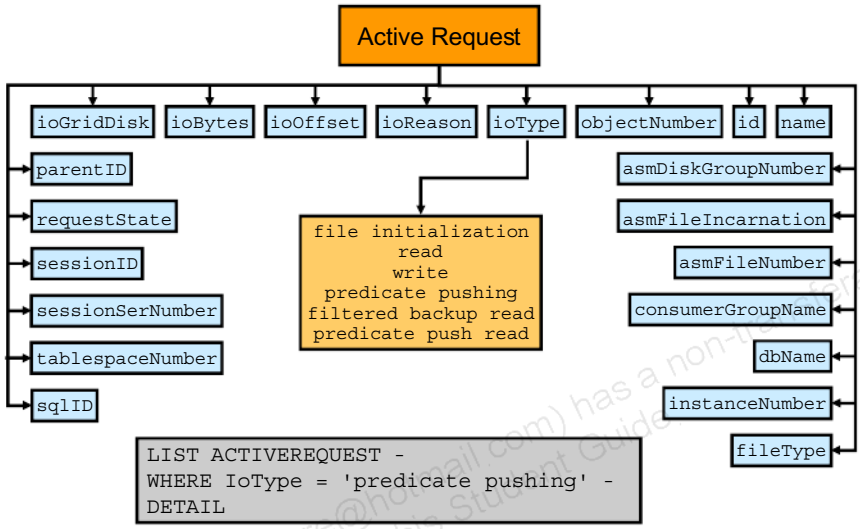
The Exadata Storage Server quarantine mechanism is designed to be automated and self-contained. However, CellCLI commands are available to manually manipulate quarantines.

- Use the `LIST QUARANTINE` command to show the quarantines currently on the cell.
- Use the `ALTER QUARANTINE` command to set the comment attribute. The comment attribute is the only quarantine attribute that can be modified.
- Use the `DROP QUARANTINE` command to manually remove a quarantine. In general, a quarantine can be removed if the quarantined entity is not expected to cause further problems. Refer to the alert message associated with the quarantine for more details before manually removing a quarantine.
- Use the `CREATE QUARANTINE` command to manually create a quarantine object. Manual creation of quarantines should be done in coordination with Oracle Support Services. Manual quarantines are created to proactively isolate SQL statements which are known to cause problems. Following is an example of manual quarantine creation:

```
CELLCLI> CREATE QUARANTINE  quarantineType="SQLID",
sqlid="5xnjp4cutc1s8"
```

The slide lists the attributes associated with the quarantine object. Note that the `sqlPLanHashValue` attribute is only applicable when the `quarantineType` is `SQL PLAN`. Likewise, the IO attributes (`ioGridDisk`, `ioBytes` and `ioOffset`) are only applicable when the `quarantineType` is `DISK REGION`.

Monitoring Exadata Storage Server with Active Requests



An active request provides a client-centric or application-centric view of client I/O requests that are currently being processed by a cell.

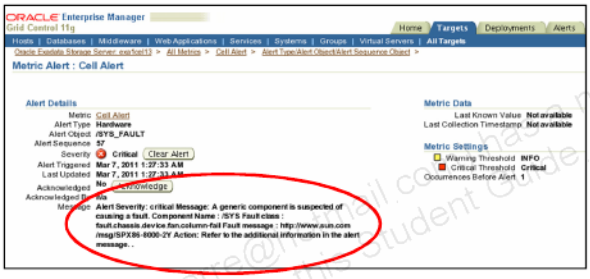
The slide shows the most important attributes of an active request. You can see that an active request is characterized at all levels: instance, database, ASM, and cell. Most of the attributes have self-explanatory names. Here is a brief explanation of some of the attributes:

- `ioReason` is the reason for the I/O activity, such as a control-file read.
- `ioType` identifies the type of active request. Possible values are listed in the slide.
- `requestState` identifies the state of the active request. Possible values include:
  - Accessing Disk
  - Computing Result
  - Network Receive
  - Network Send
  - Queued Extent
  - Queued for Disk
  - Queued for File Initialization
  - Queued for Filtered Backup Read
  - Queued for Network Send
  - Queued for Predicate Pushing
  - Queued for Read
  - Queued for Write
  - Queued in Resource Manager

Use the `LIST ACTIVEREQUEST` command to display active request details for the cell. The syntax is very similar to other `LIST` commands. You can specify which attributes to display or you can display them all using the `DETAIL` clause. You can also filter the output using a `WHERE` clause.

Monitoring Exadata Storage Server with Grid Control Overview

- Each storage server is a separate target in Grid Control
  - Storage servers can be grouped together in a System
- Metrics inside Grid Control are mostly based on cell metrics
- Additional thresholds can be set in Grid Control
- Alerts generated inside the cell are displayed inside Grid Control
- Additional alerts can be generated inside Grid Control



An Exadata Storage Server is monitored as a single Grid Control target, which covers the hardware, operating system and Exadata Storage Server software. Exadata Storage Servers are independent units that are each identified as separate targets in Grid Control. However, storage servers can be grouped together in a System to facilitate monitoring as a group.

Grid Control in conjunction with the Exadata Storage Server plug-in leverage and extend the system of metrics and alerts within each Exadata Storage Server.

When Grid Control is used to monitor storage servers, there are two different types of related metrics: storage server metrics and Enterprise Manager metrics. In most cases there is a one-to-one mapping between the two. The Management Server (MS) on Exadata Storage Server collects, computes, and manages metrics for the individual storage server. The Exadata Storage Server plug-in gathers metrics, typically from all the Database Machine storage servers, and presents them to the user in Grid Control as Enterprise Manager metrics.

In addition to cell-level metric thresholds, using the Exadata Storage Server plug-in you can set separate Grid Control thresholds for all the Exadata metrics supported by the plug-in. Prior to Exadata Storage Server software release 11.2.2.3.1, cell thresholds can only be set on a limited set of cell metrics so Grid Control thresholds offer the opportunity to automatically raise alerts based on a wider range of metrics.

When equivalent metrics can be defined in either Grid Control or at the cell level, the general recommendation is to define the threshold at the cell level so that the threshold definition and any associated alerts are maintained within the cell.

All Exadata Storage Server alerts are delivered by the storage server to Grid Control using Simple Network Management Protocol (SNMP). There are two types of server alerts that come from Exadata Storage Server:

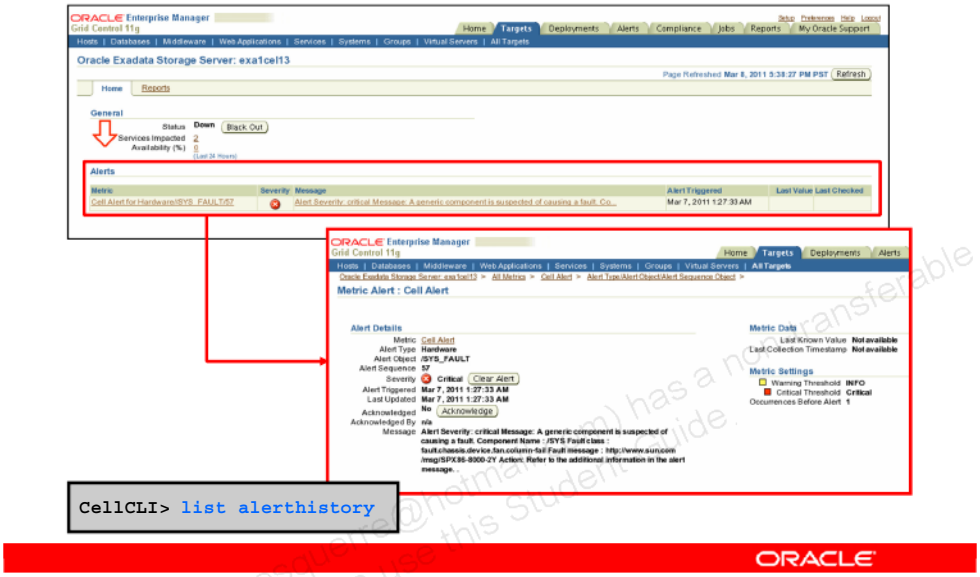
- For Integrated Lights Out Manager (ILOM)-monitored hardware components, ILOM reports a failure or threshold exceeded condition as an SNMP trap, which is received by MS. MS processes the trap, creates an alert for the storage server, and delivers the alert via SNMP to Grid Control.
- For MS-monitored hardware and software components, MS processes the failure or threshold exceeded condition, creates an alert, and delivers the alert via SNMP to Grid Control.

From an end-user perspective there is no difference between these two kinds of alerts.

An alert message may contain a corrective action which needs to be performed to resolve the alert. For example, the circled area in the screen shot on the slide indicates the action to take for the specified alert.

The fix for bug 8814019 should be installed to resolve incorrect text in the Action portion of Exadata Storage Server alerts.

Monitoring Hardware Failure and Sensor State

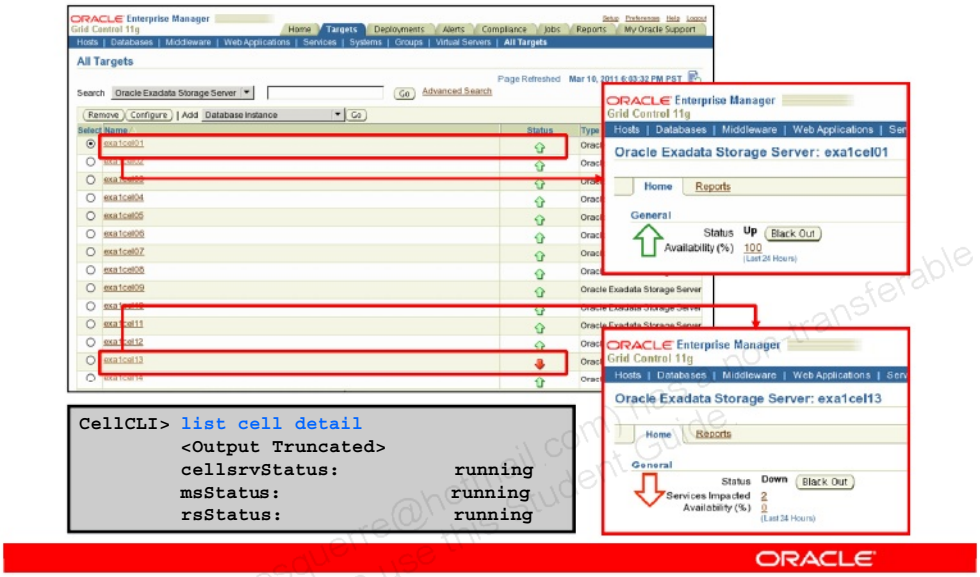


The hardware of an Exadata Storage Server is monitored collectively by ILOM and MS. ILOM monitors availability and sensor state using preset thresholds for Exadata Storage Server hardware components such as the system motherboard, processors, memory, power supplies, fans, and network interface controllers. MS monitors other hardware components directly, including the disk controller, hard disk drives, flash accelerator cards, and InfiniBand host channel adapter (HCA).

Together ILOM and MS provide full hardware monitoring and alerting. When an issue arises an alert is automatically generated within the cell, and is also propagated to Grid Control (if configured). The screen-shot in the slide shows an example of an alert notification for a hardware fault.

If Grid Control or some other SNMP manager is not configured, then administrators should configure cells to deliver alerts using email or they should periodically check the cell using the CellCLI `LIST ALERTHISTORY` command.

Monitoring Exadata Storage Server Availability



Monitoring of Exadata Storage Server availability is built into the Exadata Storage Server plug-in for Grid Control. The screen shots in the slide show examples of how Grid Control indicates that storage servers are available or unavailable.

For environments without Grid Control, storage server availability should be checked by confirming network connectivity. In addition, the status of the cell services should be checked using the CellCLI `LIST CELL DETAIL` command. The output for a healthy cell should contain the following:

```
cellsrvStatus:          running
msStatus:               running
rsStatus:               running
```

Checking for Undelivered Alerts

- Periodically check for undelivered alerts
  - CellCLI example checking one cell:

```
CellCLI> list alerthistory where notificationState != 1 and examinedBy = ''
         1_1     2011-04-25T12:09:22-07:00       warning         "The warning
     threshold for the following metric has been crossed. Metric Name        :
     CL_MEMUT  Metric Description : Percentage of total physical memory on the cell
     that is currently used  Object Name       : exalcel01  Current Value      :
     51.0 %  Threshold Value    : 50.0 %  "

CellCLI>
```

  - `dcli` example that checks all the cells listed in the `cell_group` file:

```
$ dcli -g cell_group cellcli -e "list alerthistory where notificationState != 1
  and examinedBy = \'\'"
```

- If undelivered alerts exist, check cell-to-agent network connectivity, agent availability and cell configuration

Relying on Grid Control for alert notifications depends on alerts being reliably propagated to Grid Control. While this is usually the case, administrators should periodically check for undelivered alerts.

The CellCLI `LIST ALERTHISTORY` command can be used to check for undelivered alerts. The first example in the slide shows how to check for alerts that don't have a success status (`1`) and which have not been marked as examined by an administrator. The second example shows how to use the same command in conjunction with `dcli` to check a group of cells at once.

Note that an alert might not have a success state for many reasons and that output generated by the monitoring command in the slide does not necessarily relate to a cell fault or misconfiguration. For example, an alert might not be propagated because the agent is unavailable. Also, if the cell is configured to deliver alerts using SNMP and email, it is possible for the alert to show an unsuccessful notification state when one delivery method succeeds but the other one fails. Finally, if communication is disrupted between a storage server and the agent where the Exadata Storage Server plug-in is deployed, alerts processed by MS may not be delivered to Grid Control. Remember that MS will retry up to 5 times to deliver the alert, so if the communication disruption is temporary you may observe an undelivered alert and then see the same alert delivered moments later.

Checking for Disk I/O Errors

- On each cell, create a warning threshold using the following CellCLI `CREATE THRESHOLD` command:

```
CellCLI> create threshold CD_IO_ERRS_MIN comparison='>', warning=0, -
> occurrences=1, observation=1
Threshold CD_IO_ERRS_MIN successfully created

CellCLI> list threshold CD_IO_ERRS_MIN detail
         name:                   CD_IO_ERRS_MIN
         comparison:             >
         observation:            1
         occurrences:            1
         warning:                0.0

CellCLI>
```


The slide shows a recommended storage server threshold setting that assists in checking for disk I/O errors. Disk I/O errors are typically handled by Exadata cell software automatically. The recommended threshold setting causes an alert to be generated by the cell when an I/O error is detected.

No action is required for a single warning alert. However many warnings for a single drive may by a precursor to a drive failure. Likewise, many warnings across all the drives in a single cell may indicate an impending fault with a controller or some other component.

If this warning is reported and a critical alert is also generated by MS for a disk component, follow the action specified in the critical alert.

Checking for Network Errors

- Monitor the storage network for dropped packets:
  - Using Grid Control
    - For the storage server target metrics:
      - Host MB Dropped Per Sec
      - Host RDMA MB Dropped Per Sec
    - Apply the following settings:
      - Warning threshold to Zero
      - Collection Schedule to Repeat Every 5 Minutes
      - Upload Interval to 3 Collections
  - Directly on the cell
    - Monitor the Host Interconnect Metrics, in particular:
      - `N_MB_DROP_SEC`
      - `N_MB_RDMA_DROP_SEC`

A functioning storage network is essential to the overall health and performance of Exadata Database Machine. A loss of functionality is often indicated by the amount of dropped network traffic. Normally, dropped network traffic is automatically dealt with inside Database Machine so occasional dropped packets are not a cause for concern. However many dropped packets for a single cell may be a precursor to a problem.

To raise the awareness of such situations it is recommended to set thresholds inside Grid Control for Host MB Dropped Per Sec and Host RDMA MB Dropped Per Sec as described on the slide. Note that the settings must be applied to each storage server target.

Periodic monitoring of the `N_MB_DROP_SEC` and `N_MB_RDMA_DROP_SEC` cell metrics is advised for environments without Grid Control. The use of thresholds and alerts to monitor these metrics at the cell level can be adopted in preference to setting Grid Control thresholds using Exadata Storage Server software release 11.2.2.3.1 and beyond.

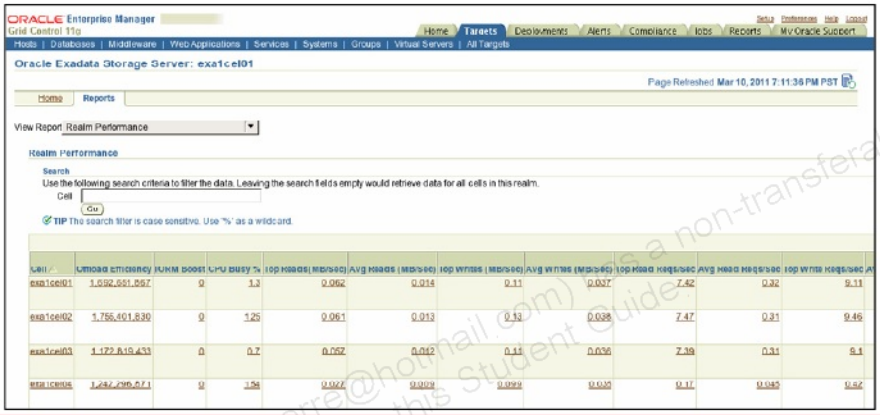
Monitoring Filesystem Free Space

- Filesystem free space on Exadata Storage Servers is monitored automatically by MS:
  - No direct administrator action is required
  - MS generates an alert if free space becomes low
  - MS automatically reclaims used space by purging old log files, trace files, crash dumps and other unnecessary files

Exadata Storage Server software automatically monitors filesystem utilization on each storage server. There is no need to set a metric threshold for filesystem utilization within the cell or in Grid Control. MS monitors filesystem free space, generates critical alerts when free space becomes low, and takes corrective action to free used space. For more detailed information see the section entitled Understanding Automated Cell Maintenance in the *Oracle Exadata Storage Server Software User's Guide 11g Release 2 (11.2)*.

Comparing Metrics Across Multiple Storage Servers

- Compare resource utilization across storage servers using cell metric observations and realm-wide reports in Grid Control
- Large imbalances may indicate a problem



Exactly how data is spread across a Database Machine depends upon many factors including the number of applications and databases that are consolidated on the environment, along with how storage is configured to meet various performance, availability, security and other requirements.

However, in general terms work performed by storage servers to satisfy database requests should be evenly spread across multiple storage servers. A significant imbalance in resource utilization or I/O performance for a single storage server compared to the others should be a cause for further investigation.

Various CPU, memory and I/O utilization metrics can be periodically compared across the storage servers to ensure there are no unwarranted imbalances. Grid Control users can utilize the various realm-wide reports to quickly and easily perform comparisons across a group of storage servers. The screen shot in the slide shows and example of the realm performance report which provides a very useful high-level comparison for key metrics across a realm of storage servers.

Monitoring Metrics Within a Storage Server

- There are no default metric thresholds for CPU, memory or I/O utilization
- 100% utilization may be normal in certain cases
- Metric observations are useful and relevant when compared with previously captured baselines
- Suggested methodology:
  1. Determine key metrics based on your applications and service requirements
  2. Record baseline metric observations based on normal and peak system usage
  3. Implement thresholds to generate alerts for observations outside expected ranges
     - Warning thresholds to indicate slightly abnormal observations
     - Critical thresholds to indicate situations that could affect service levels

Exadata Storage Server is designed to be a well balanced storage subsystem. As a result, CPU, memory, and I/O metric values on Exadata Storage Server can be higher than a typical database or general-purpose server, but still be in proper operating range.

For example, CPU may be 100% utilized while performing queries against data using Exadata Hybrid Columnar Compression (EHCC). In this case, 100% CPU utilization is not problematic and no alert should be generated.

Resource usage metric observations within a storage server are only useful and relevant if they are compared with values captured during a period of normal operation. The slide outlines a suggested methodology to determine important metrics and implement thresholds based on previously recorded observations.

Third Party Monitoring Tools

- Installing additional software, including any third party monitoring agent, is not supported on Exadata Storage Server
- Exadata Storage Server can be configured to send alerts to any SNMP subscriber

It is not permissible to install any additional software, including third party monitoring agents, on Exadata Storage Server.

It is permissible to configure Exadata Storage Server to send alerts using SNMP to third party management consoles.

Quiz

When using Grid Control to monitor Exadata Storage Servers, where should thresholds be defined?

a. In Grid Control only
b. Inside the Exadata cell only
c. Inside the Exadata cell if possible, otherwise in Grid Control
d. Either in Grid Control or inside the Exadata cell, it doesn't matter

Answer: c

When equivalent metrics can be defined in either Grid Control or at the cell level, the general recommendation is to define the threshold at the cell level so that the threshold definition and any associated alerts are maintained within the cell.

Quiz

Assuming that the Exadata Storage Server plug-in is properly configured, which of the following can occur?

a. A metric observation generates an alert in the cell but no alert is seen in Grid Control

b. A metric observation generates an alert in the cell and an alert is seen in Grid Control

c. A metric observation generates no cell alert but an alert is seen in Grid Control

d. A metric observation results in two alerts in Grid Control

ORACLE

Answer: a, b, c, d

A metric observation generates an alert in the cell but no alert is seen in Grid Control: This can occur when a cell-based threshold triggers a cell alert but the alert is not propagated to Grid Control due to a network communication problem.

A metric observation generates an alert in the cell and an alert is seen in Grid Control: This is normal and expected behaviour. It occurs when a cell-based threshold triggers an alert and the alert is propagated to Grid Control.

A metric observation generates no cell alert but an alert is seen in Grid Control: This occurs when a threshold set in Grid Control triggers an alert, and there is either no cell-level threshold for the metric or the cell-level threshold is set at a level which doesn't trigger an alert in the cell.

A metric observation results in two alerts in Grid Control: This occurs when a cell-level threshold and a Grid Control threshold are defined on the same metric, and the corresponding cell-level alert is propagated to Grid Control.

Summary

In this lesson, you should have learned how to:

- Describe Exadata Storage Server metrics, alerts and active requests
- Identify the recommended focus areas for Exadata Storage Server monitoring
- Describe how to monitor the recommended Exadata Storage Server focus areas

Additional Resources

- Lesson Demonstrations
  - Monitoring Exadata Storage Servers using Enterprise Manager Grid Control and the System Monitoring Plug-in for Exadata Storage Server
  - Managing Exadata Storage Server Alerts and Checking for Undelivered Alerts
  - Exadata Storage Server Monitoring and Management using Integrated Lights Out Manager (ILOM)

Practice 15 Overview: Monitoring Exadata Storage Server

In this practice you will monitor Exadata Storage Server using metrics, alerts and active requests.

16

Monitoring Exadata Database Machine Database Servers

ORACLE

Objectives

After completing this lesson, you should be able to describe the monitoring recommendations for Exadata Database Machine database servers.

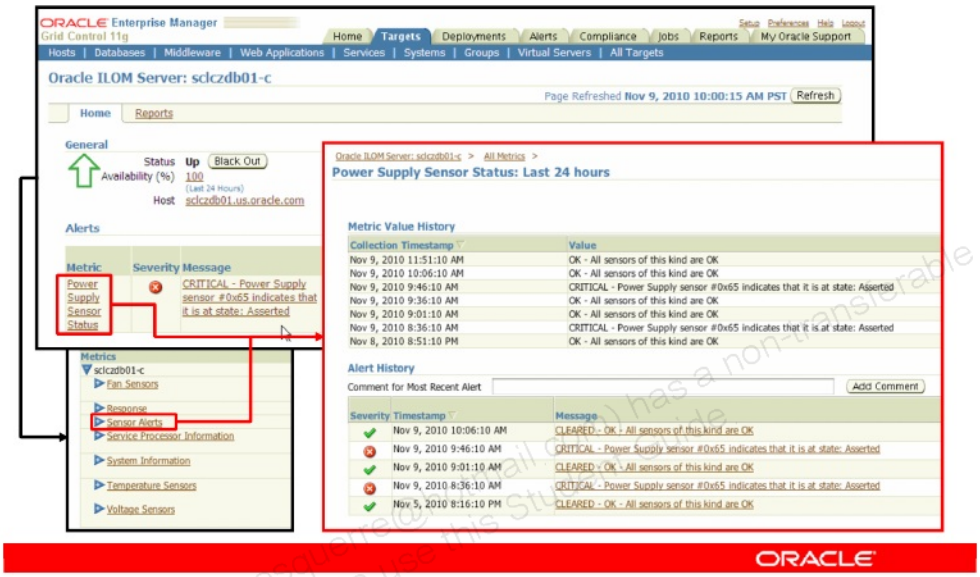
Monitoring Database Servers Overview

- Essentially, monitoring an Exadata Database Machine database server is the same as monitoring any clustered Oracle Database server
  - Current skills and practices are readily transferrable
- The lesson covers Database Machine specific differences and recommendations with a focus on using Grid Control
- Areas considered in this lesson include:
  - Monitoring hardware
  - Monitoring the Operating System
  - Monitoring Oracle Grid Infrastructure
  - Monitoring Oracle Database
  - Monitoring Oracle Management Agent

For the most part, monitoring Exadata Database Machine database servers is the same as monitoring any other clustered Oracle Database server. So the skills and practices already used by Oracle Database Administrators are readily transferrable into a Database Machine environment.

There are some specific differences and recommendations associated with Exadata Database Machine and Grid Control, and these are the focus of this lesson.

Monitoring Hardware



Integrated Lights Out Manager (ILOM) monitors availability and sensor state using preset thresholds for database server hardware components such as the system motherboard, processors, memory, power supplies, fans, and network interface controllers. The availability and sensor state can be monitored inside Grid Control using the ILOM plug-in. The following recommendations apply:

- There are no Exadata-specific thresholds to set for database server hardware monitoring. Failure conditions and threshold settings for the components monitored by ILOM are preset in ILOM and are sufficient for the necessary level of monitoring.
- To view current sensor readings, log in to the Grid Control Console and navigate to All Metrics from the plug-in home page. To view current component status, including those that have a Faulted status, expand the Sensor Alert section under All Metrics and review the metrics for each sensor. Components with a Faulted status will generate an alert. Any active alert will be visible on the home page of the ILOM target.
- Alerts generated by ILOM in a database server may be viewed using Grid Control in the same fashion as an alert generated by any other Grid Control target.
- To view the history of alerts generated by ILOM navigate to the target home page and expand the Sensor Alert section under All Metrics and review the observations for each sensor. A history of each sensor state is available for up to 31 days.

Monitoring the Operating System

- Operating system monitoring is built into Grid Control using the Host target type:
  - There are no Exadata-specific configuration requirements for general monitoring
  - Thresholds can be set or changed to accommodate site-specific requirements
  - Remember to monitor disk I/O in the storage servers and not on the database servers
  - Essential operating system alerts are generated by Grid Control based on default metric thresholds

The database server operating system is viewed in Grid Control as a Host target. The following recommendations apply:

- There are no Exadata-specific configuration requirements. The metrics and default thresholds provided by Grid Control are sufficient for general monitoring.
- Thresholds may be set or changed in the Metric and Policy Settings page in Grid Control to handle site-specific requirements.
- Remember that I/O against Oracle Database data files is directed to Exadata Storage Servers using the InfiniBand network. Therefore, metrics on database servers relating to disk I/O cannot be used to monitor Oracle data files. I/O relating to Oracle data files should be monitored on Exadata Storage Servers.
- Database server operating system alerts are generated automatically by Grid Control based on default metric thresholds for the Host target.

Monitoring Oracle Grid Infrastructure

- Grid Infrastructure monitoring is built into Grid Control using the ASM, Listener and Cluster target types:
  - There are no Exadata-specific configuration requirements for general monitoring
  - Thresholds can be set or changed to accommodate site-specific requirements
  - Essential alerts are generated by Grid Control based on default metric thresholds
- Monitoring of the ASM log file for cell connectivity issues is not provided by Grid Control
  - Check the ASM alert log file for messages such as:

```
connect: ossnet: connection failed to server <ipaddr>,
   result=5 (login: sosstcpreadtry failed)
```


Oracle Automatic Storage Management (ASM), listener, and clusterware are viewed in Grid Control as ASM, Listener, and Cluster targets, respectively. The following recommendations apply:

- There are no Exadata-specific configuration requirements. The metrics and default thresholds provided by Grid Control are sufficient for general monitoring.
- Thresholds may be set or changed in the Metric and Policy Settings page in Grid Control to handle site-specific requirements. For example, thresholds are typically configured to provide a warning when the amount of free space in a disk group (`USABLE_FREE_MB`) falls below the required amount of free space to maintain redundancy if a failure group is lost (`REQUIRED_MIRROR_FREE_MB`).
- Alerts are generated automatically by Grid Control based on default metric thresholds for the ASM, Listener, and Cluster targets.
- Monitoring of the ASM alert log file for cell connectivity issues is currently not integrated into Grid Control. Check the ASM alert log file for messages such as the example shown in the slide. This check should be performed every 5 minutes.

Monitoring Oracle Database

- Monitoring of Oracle databases is built into Grid Control using Cluster Database and Database Instance targets:
  - There are no Exadata-specific configuration requirements for general monitoring
  - Thresholds can be set or changed to accommodate site-specific requirements
  - Essential alerts are generated by Grid Control based on default metric thresholds
- Monitoring of the database instance log file for cell connectivity issues is not provided by Grid Control
  - Check the alert log file for messages such as:

```
connect: ossnet: connection failed to server <ipaddr>,
   result=5 (login: sosstcpreadtry failed)
```


Oracle databases are viewed in Grid Control as Cluster Database and Database Instance targets. The following recommendations apply:

- There are no Exadata-specific configuration requirements. The metrics and default thresholds provided by Grid Control are sufficient for general monitoring.
- Thresholds may be set or changed in the Metric and Policy Settings page in Grid Control to handle site-specific requirements.
- Alerts are generated automatically by Grid Control based on default metric thresholds for the Cluster Database and Database Instance targets.
- Monitoring of the database instance alert log file for cell connectivity issues is currently not integrated into Grid Control. Check the alert log file for messages such as the example shown in the slide. This check should be performed every 5 minutes.

Monitoring Oracle Management Agent

- Oracle Management Agent monitoring is built into Grid Control using the Agent target type
  - There are no Exadata-specific configuration requirements for general monitoring
  - Thresholds can be set or changed to accommodate site-specific requirements
  - Essential alerts are generated by Grid Control based on default metric thresholds
- Remember that agent availability is especially important for the agent hosting the monitoring plug-ins

Oracle Management Agents are viewed in Grid Control as agent targets. The following recommendations apply:

- There are no Exadata-specific configuration requirements. The metrics and default thresholds provided by Grid Control are sufficient for general monitoring.
- Thresholds may be set or changed in the Metric and Policy Settings page in Grid Control to handle site-specific requirements.
- Alerts are generated automatically by Grid Control based on default metric thresholds. By default, Grid Control generates an alert if an agent is down or loses contact with the management server.
- Always remember which agent is hosting the monitoring plug-ins because the ability to monitor many Database Machine components relies on the continued availability of the agent.

Quiz

Extensive retraining is required for existing Oracle Database Administrators to monitor and maintain Oracle databases on Exadata Database Machine:

a. TRUE

b. FALSE

Answer: b

Summary

In this lesson, you should have learned how to describe the monitoring recommendations for Exadata Database Machine database servers.

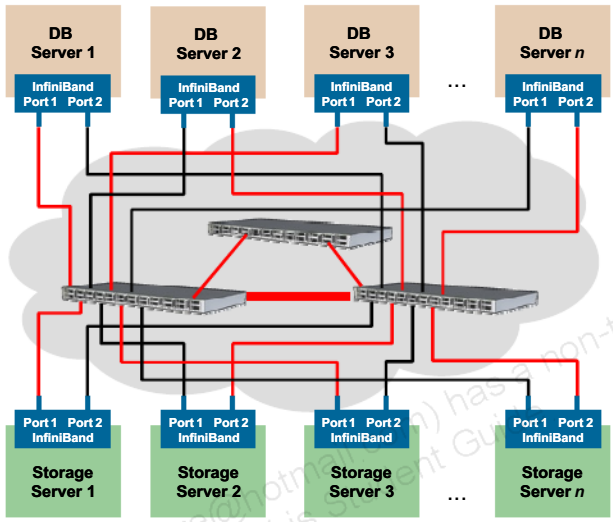
17

Monitoring the InfiniBand Network

Objectives

After completing this lesson, you should be able to describe and execute the commands to monitor the Exadata Database Machine InfiniBand network.

InfiniBand Network Monitoring Overview



The diagram in the slide illustrates the InfiniBand communications network contained inside Exadata Database Machine.

Each Database Machine contains at least two InfiniBand Switches which connect the database servers and storage servers. These are called leaf switches. A third switch, called the spine switch, connects to both leaf switches in Half Rack and Full Rack Database Machines. The spine switch facilitates connection of multiple racks to form a single larger Database Machine environment.

Each server contains at least one pair of InfiniBand ports which are bonded together using active/passive bonding. The active and passive connections are spread across both leaf switches for load balancing purposes. In addition, the leaf switches within a rack are connected to each other. The result is a Fat-Tree switched fabric network topology.

Monitoring of the InfiniBand network is divided into three main areas:

- InfiniBand switch monitoring
- InfiniBand port monitoring
- Monitoring of the InfiniBand fabric

The remainder of this lesson covers these three areas.

Manually Monitoring the InfiniBand Switches

- InfiniBand Switch monitoring checks for failed switch hardware and sensors that exceed preset thresholds
- Manual monitoring is required for switches running a switch software version prior to 1.1.3
  - Example of expected output:

```
# ssh root@dm01sw-ib2
root@dm01sw-ib2's password:
[root@dm01sw-ib2 ~]# showunhealthy
OK - No unhealthy sensors
[root@dm01sw-ib2 ~]# checkpower
PSU 0 present status: OK
PSU 1 present status: OK
```

  - Checks should be performed every 60 to 120 seconds
  - Use `env_test` to gather more information if required

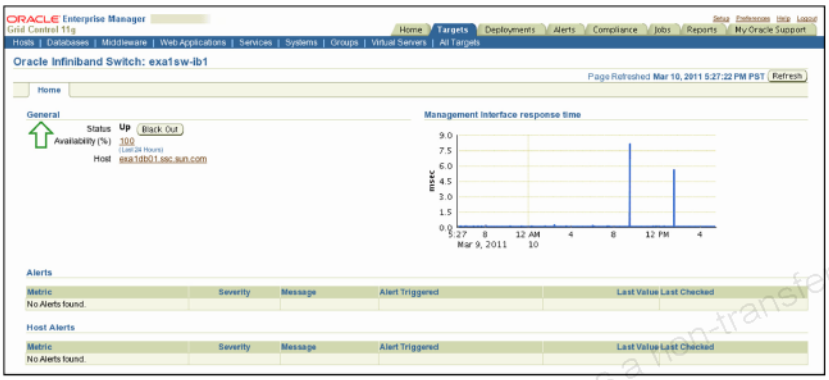
Monitoring of the InfiniBand switches provided with Oracle Exadata Database Machine involves checking for failed hardware components and sensors that exceed preset thresholds on the switch.

InfiniBand switches running switch software versions prior to version 1.1.3 do not have SNMP support. On these switches, monitoring can be performed by logging into the switch as the `root` user and running the `showunhealthy` and `checkpower` commands. An example is shown in the slide. It is recommended that the checks should be run every 60 to 120 seconds. In practical terms this means creating a script to perform the check and using an automated capability (such as the User Defined Metrics capability inside Grid Control) to run it repeatedly.

Note that `showunhealthy` will indicate no unhealthy sensors even if a power supply is offline, so they must be checked separately.

If `showunhealthy` or `checkpower` indicate any issues, run the `env_test` command to get more detailed status information for all switch sensors.

Monitoring the InfiniBand Switches with Grid Control



ORACLE

Commencing with version 1.1.3, the InfiniBand switch software supports SNMP. In conjunction with this, the InfiniBand Switch plug-in can be used to monitor switch status from Grid Control. A screen shot of the InfiniBand switch home page is shown in the slide.

Monitoring the InfiniBand switches using Grid Control is simply a matter of checking that the switches are up and dealing with any alerts that are generated. When alerts are generated, you can use the manual commands shown on the previous page to gather additional information.

Note that switches with switch software versions prior to version 1.1.3 can be upgraded to current software versions however such an upgrade may also require associated upgrades to other Database Machine components. Refer to My Oracle Support bulletin 888828.1 for detailed version compatibility information relating to Exadata Database Machine.

Monitoring the InfiniBand Switch Ports

- Check for switch port errors with the following command:

```
# ibqueryerrors.pl -s RcvSwRelayErrors,RcvRemotePhysErrors,XmtDiscards,
XmtConstraintErrors,RcvConstraintErrors,ExcBufOverrunErrors,VL15Dropped
```

  - Compare output with previous results
    - `SymbolErrors` or `RcvErrors` or `LinkIntegrityErrors` should not increase without `LinkDowned` increasing
  - Check can be executed from any database server or InfiniBand switch
  - Check should be performed every 60 to 120 seconds

InfiniBand port monitoring is required to check the health of InfiniBand network ports and interfaces. The recommended procedure is to manually execute the `ibqueryerrors.pl` command as shown in the slide.

A single invocation of the `ibqueryerrors.pl` command shown in the slide can be used to report switch port errors across all the InfiniBand switches. The output of the command should be compared with previously gathered output. When all is functioning correctly, the counts for `SymbolErrors` or `RcvErrors` or `LinkIntegrityErrors` should not increase without `LinkDowned` increasing. The check can be executed from any database server or InfiniBand switch. It is recommended that the check should be performed every 60 to 120 seconds.

Monitoring the InfiniBand Ports on Database Machine Servers

- InfiniBand port monitoring is automatically performed on Exadata Storage Servers
- Manually monitor database server ports with the following commands:

```
# ibstatus
# perfquery
# ifconfig
# ping <Remote InfiniBand Hostname>
# rds-ping <Remote InfiniBand Hostname>
```

  - Automatic monitoring with Grid Control can be performed using User Defined Metric (UDM) scripts which are available from My Oracle Support note 1110675.1

InfiniBand port monitoring on storage servers is performed by MS. No additional InfiniBand monitoring is required on storage servers. If an InfiniBand port is not functioning correctly, MS creates an alert, and delivers the alert Grid Control. Alert messages typically contain the corrective action to perform to resolve the alert.

Database servers have no built-in InfiniBand monitoring. The following list describes the recommended checks for each database server. Run the checks every 60 to 120 seconds.

- `ibstatus`: Check that every port shows the following output:

```
state:      4: ACTIVE
phys state  5: LinkUp
rate:       40 Gb/sec (4X QDR)
```

- `perfquery`: Compare output with previously gathered results. `SymbolErrors` or `RcvErrors` or `LinkIntegrityErrors` should not increase without `LinkDowned` increasing.
- `ifconfig`: Check that the InfiniBand interfaces (`bondib0`, `ib0` and `ib1`) are `UP`.
- `ping` and `rds-ping`: Check for connectivity to all database servers and storage servers over the InfiniBand network.

The checks in the above list can be automated in Grid Control using User Defined Metric (UDM) scripts which are available from My Oracle Support note 1110675.1.

Monitoring the InfiniBand Fabric: Subnet Manager Master Location

- The following checks are recommended once per day:
  - Check that the InfiniBand Subnet Manager (SM) master is located on one of the InfiniBand switches
    - Example:

```
# sminfo
sminfo: sm lid 1 sm guid 0x21283a8516a0a0, activity count 933330
priority 5 state 3 SMINFO_MASTER
# ibswitches
Switch : 0x0021283a8516a0a0 ports 36 "Sun DCS 36 QDR switch dm01sw-ib1"
enhanced port 0 lid 1 lmc 0
Switch : 0x0021283a8983a0a0 ports 36 "Sun DCS 36 QDR switch dm01sw-ib2"
enhanced port 0 lid 4 lmc 0
Switch : 0x0021283a89bda0a0 ports 36 "Sun DCS 36 QDR switch dm01sw-ib3"
enhanced port 0 lid 3 lmc 0
```

  - For networks containing spine switches, check that SM master is running on a spine switch

Daily, it is recommended to check that the Subnet Manager (SM) master is located on one of the InfiniBand switches. From one of the database servers or InfiniBand switches, execute the `sminfo` and `ibswitches` commands as shown in the slide. Determine the hostname for the SM master location by matching the `guid` reported by `sminfo` with the output from `ibswitches`. Alternatively, the `getmaster` command can be used, however note that `getmaster` is only available on the InfiniBand switches, not the database servers.

In a Half Rack, Full Rack or multi-rack Database Machine environment, SM master should run on a spine switch. A spine switch is a switch that only has other switches connected to it. The following command identifies spine switches.

```
ibnetdiscover -p | awk `
/^SW +[0-9]+ +[0-9]+ +0x[0-9\
a-e]+ +[0-9]+x .DR - [SW|CA].*/ {
  if (spine[$4]==””) spine[$4]=”yes”
  if ($8 == “CA”) spine[$4]=”no”
}
END {
  for (val in spine)
    if (spine[val]==”yes”)
      print val
}
`
```

Monitoring the InfiniBand Fabric: Network Topology and Link Status

- Use `verify-topology` to check the InfiniBand fabric
  - Example of expected output:

```
# /opt/oracle.SupportTools/ibdiagtools/verify-topology

        [ DB Machine Infiniband Cabling Topology Verification Tool ]
                [Version 11.2.1.3.b]

Looking at 1 rack(s).....
Spine switch check: Are any Exadata nodes connected .............[SUCCESS]
Spine switch check: Any inter spine switch connections...........[SUCCESS]
Spine switch check: Correct number of spine-leaf links...........[SUCCESS]
Leaf switch check: Inter-leaf link check.........................[SUCCESS]
Leaf switch check: Correct number of leaf-spine connections......[SUCCESS]
Check if all hosts have 2 CAs to different switches..............[SUCCESS]
Leaf switch check: cardinality and even distribution.............[SUCCESS]
```

- Use `iblinkinfo.pl -Rl` to monitor link status
  - Compare output with previously gathered output

In addition to checking the SM master, it is recommended to monitor the InfiniBand network topology and the state of all the InfiniBand network links.

Once per day, it is recommended to execute the `verify-topology` command from one of the Database Machine database servers. `verify-topology` conducts a series of topology-level tests to verify that the expected storage network links are cabled correctly. An example of the expected output for `verify-topology` is shown in the slide.

To check the state of all the InfiniBand network links, run the following command as `root` on one of the Database Machine database servers:

```
# iblinkinfo.pl -Rl
```

Compare the resulting output with previously gathered output from a time when the InfiniBand network was in a known good state. It is recommended to run this check frequently (every 1 to 5 minutes).

Quiz

Which of the following is true?

a. By default, InfiniBand port monitoring is automatically performed by MS on Exadata Storage Servers
b. By default, InfiniBand port monitoring on database servers is performed by Grid Control using User Defined Metrics

Answer: a

User Defined Metrics can be used to monitor the InfiniBand ports on database servers, however they are not provided inside Grid Control by default. The User Defined Metrics must be defined for each database server.

Summary

In this lesson, you should have learned how to describe and execute the commands to monitor the Exadata Database Machine InfiniBand network.

Additional Resources

- Lesson Demonstration
  - Monitoring the Database Machine InfiniBand network
- My Oracle Support Notes
  - Oracle Database Machine Monitoring Best Practices
  - Database Machine and Exadata Storage Server 11*g* Release 2 (11.2) Supported Versions

18

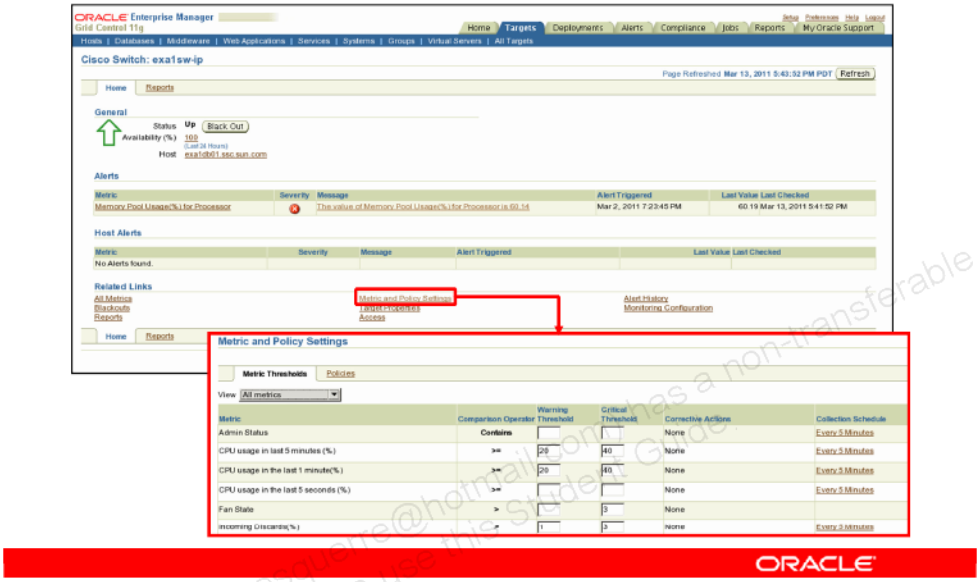
Monitoring Other Exadata Database Machine Components

Objectives

After completing this lesson, you should be able to monitor the following additional Exadata Database Machine components:

- Cisco Catalyst Ethernet Switch
- Sun Power Distribution Units
- Avocent MergePoint Unity KVM Switch

Monitoring the Cisco Catalyst Ethernet Switch



The Cisco Catalyst Ethernet switch is not in the critical data path. It does not participate in connecting application clients to the database or in connecting database servers to the storage servers. However, monitoring and administrative traffic does depend on the availability of the Cisco switch.

The primary goal of monitoring the Cisco switch is to identify hardware component failure and environmental conditions that can lead to switch malfunction. The switch monitors availability and sensor thresholds for its hardware components.

A Grid Control plug-in facilitates easy monitoring of the Cisco switch. An example of the Cisco switch home page is shown in the slide. The metrics and default thresholds provided by the plug-in are sufficient for the level of monitoring necessary to ensure switch availability. Thresholds may be changed or set in the Metric and Policy Settings page to handle site-specific requirements. Alerts within the switch are reported using an SNMP trap which can be received and displayed by Grid Control.

Manual monitoring of the Cisco switch can be achieved using a command-line interface that is built into the Cisco switch. For further information, consult the documentation for Cisco Catalyst 4500 Series switches available at www.cisco.com.

Monitoring the Sun Power Distribution Units

- Each rack contains two PDUs
- Monitor PDUs to:
  - Ensure continuous power supply
  - Measure power consumption
- Monitoring options:
  - Physical inspection of the PDUs
  - Remote monitoring using the Grid Control plug-in or other SNMP manager
    - Appropriate threshold settings inside each PDU are required to facilitate remote monitoring

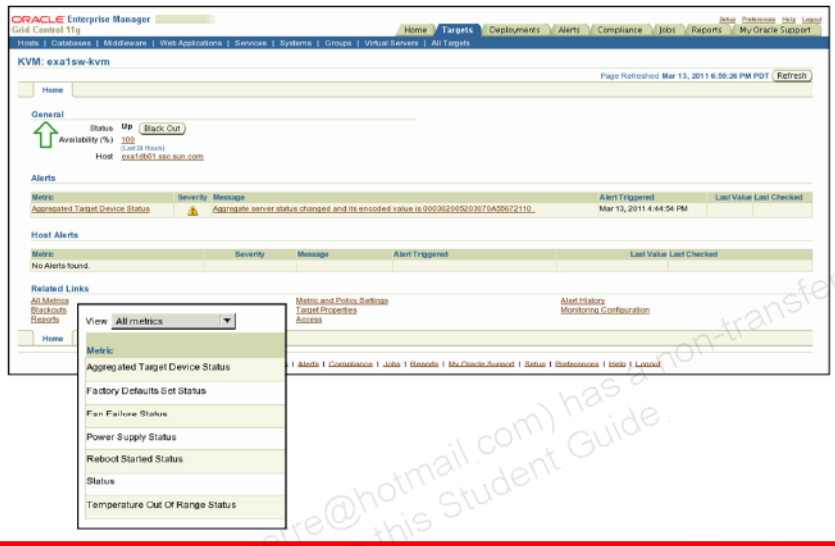
Every Exadata Database Machine rack contains two Power Distribution Units (PDUs). The PDUs provide a redundant power supply capability which helps to maintain high availability of the system. The primary goal of monitoring the PDUs is to ensure that a continual power supply is maintained. PDU monitoring can also measure power usage for data center management purposes.

On each PDU, the PDU metering unit enables you to monitor the current being used by equipment connected in the Oracle Exadata Database Machine rack. You can monitor the power usage in person by viewing the LCD screen on the PDU or remotely from a system on the network. A picture of the PDU power consumption monitor is shown in the slide.

To enable remote monitoring, a Grid Control plug-in is supplied. The plug-in shows the status of the PDU along with any alerts generated by the PDU.

Note that appropriate thresholds must be configured inside each PDU to automatically generate alerts indicating abnormal power consumption. The specific threshold values depend on several factors including the model of the Database Machine, the size of the Database Machine and the power supply voltage and type. The *Oracle Exadata Database Machine Owner's Guide* contains the recommended threshold values for each situation.

Monitoring the Avocent MergePoint Unity KVM Switch



The KVM switch provides remote console access to the Exadata Storage Servers and database servers in Database Machine. Its primary purpose is to provide keyboard, video, and mouse control to facilitate various administration tasks. The KVM switch is monitored mainly to ensure it is available.

The KVM switch can be monitored using the Exadata Avocent MergePoint Unity Switch plug-in. An example of a KVM switch home page is shown in the slide. The metrics and default thresholds provided by the plug-in are sufficient for the level of monitoring necessary to ensure switch availability. Alerts within the switch are reported using an SNMP trap which can be received and displayed by Grid Control.

Note that no KVM switch exists on Exadata Database Machine X2-8. On this model, server console functionality can be accessed using ILOM.

Quiz

It is most important to monitor which of the following components in order to maintain the overall availability of Database Machine?

a. Cisco Ethernet Switch

b. Sun Power Distribution Units

c. Avocent MergePoint Unity KVM Switch

Answer: b

Summary

In this lesson, you should have learned how to monitor the following additional Exadata Database Machine components:

- Cisco Catalyst Ethernet Switch
- Sun Power Distribution Units
- Avocent MergePoint Unity KVM Switch

Additional Resources

- Lesson Demonstration
  - Monitoring the Cisco Catalyst Ethernet switch and the Avocent MergePoint Unity KVM using Grid Control

19

Other Useful Monitoring Tools

ORACLE

Objectives

After completing this lesson, you should be able to describe the following useful tools:

- Exachk
- DiagTools
- ADRCI
- Imageinfo and Imagehistory
- OSWatcher

Exachk Overview

- Exachk is a utility that:
  - Collects data regarding Database Machine component versions and best practices
  - Allows administrators to check their environment against supported version levels and best practices
  - Is pre-installed on new Exadata Database Machines
  - Is available from My Oracle Support note 1070954.1
  - Should be executed periodically as a regular part of Database Machine monitoring and maintenace
  - Does not alter any Database Machine configuration settings
  - Is lightweight and has minimal impact on the system
- Exachk is NOT:
  - A continuous monitoring tool
  - A replacement for Enterprise Manager

Exachk collects data regarding key software, hardware, and firmware versions and configuration best practices specific to Oracle Database Machine.

The output assists Database Machine administrators to periodically review and cross reference the current data for key Database Machine components against supported version levels and recommended best practices.

Exachk is pre-installed on new Exadata Database Machines. The latest updated version of exachk is available for download from My Oracle Support note 1070954.1. Exachk can be executed as desired and should be executed regularly as part of the maintenance program for an Oracle Database Machine.

Exachk is not a continuous monitoring utility and should not be considered as an replacement for other monitoring or alerting tools such as Enterprise Manager or ILOM.

Exachk is non-intrusive and does not change anything on the Database Machine apart from writing output files and some small temporary files which are deleted after they are used.

Exachk will also offer to configure SSH user equivalence if it is not configured.
Exachk is lightweight so the impact to the target machine is minimal. On a Full Rack Database Machine X2-2 system exachk normally takes a little more than one hour to run.

Running Exachk

As the Oracle Database software owner (`oracle` OS user):

1. Download exachk.zip onto a database server
2. Unzip exachk.zip

```
$ unzip exachk.zip
```

   - Leave the `exachk` script and driver files (`collections.dat` and `rules.dat`) together in the same directory

3. Ensure the `exachk` script is executable

```
$ chmod +x exachk
```

4. Run `exachk`
   - Follow the prompts, read and understand all the messages
   - Supply the requested passwords, otherwise some checks are skipped
5. Review the reports
   - Review the summary report to identify areas for further investigation
   - Review the detailed report for recommendations and further information

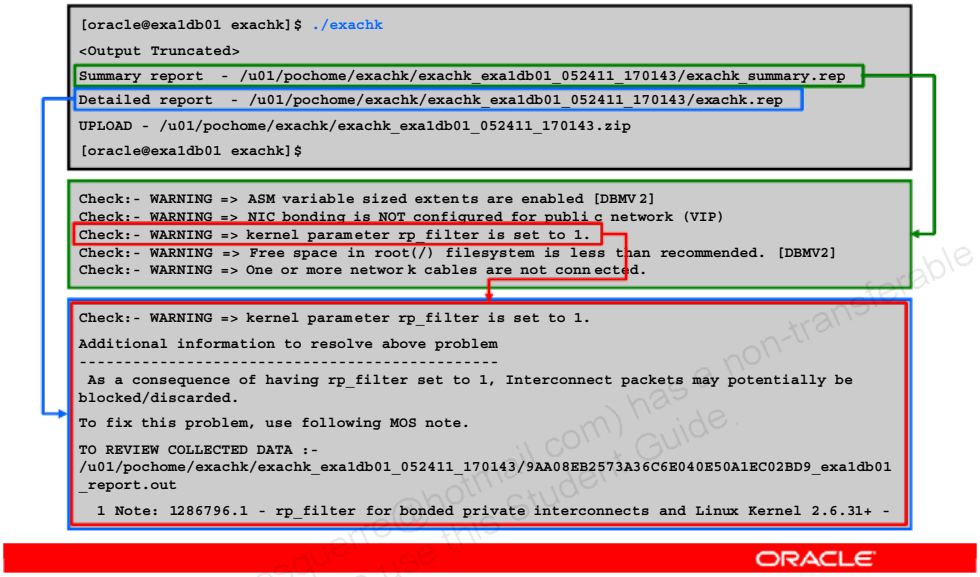
The slide outlines the recommended process for installing and running exachk.

The exachk executable does not require any mandatory command line options because the tool automatically scans the Database Machine environment to gather most of the required information. Additional information is gathered using prompts at the beginning of the exachk session. After the administrator input is gathered, exachk can be left to run unattended.

The following usage considerations should also be noted:

- It is recommended that exachk should be staged and operated from a local filesystem on a single database server in order to deliver the best performance.
- To maximize the number of checks that are performed, execute exachk when the Grid Infrastructure services and at least one database are up and running.
- While exachk is a minimal impact tool, it is a best practice to execute it during times of least load on the system.
- To avoid possible problems associated with running the tool from terminal sessions on a network attached computer, consider running the tool using VNC so that if there is a network interruption the tool will continue to execute to completion.
- If the execution of the tool fails for some reason, then it can be re-run from the beginning; exachk does not resume from the point of failure.

Exachk Output



Exachk output files are written to a time stamped directory under the location where exachk is installed. The main output files are:

- `exachk_summary.rep` which contains a summary of information, warnings and failure notices for the exachk checks.
- `exachk.rep` which contains more detailed information for all of the entries in the summary report. The entries in this report typically describe the issues in detail and provide recommendations for remediation. A link to a note in My Oracle Support may also be included to provide further information, especially if the recommendations are frequently changing or they are not straight-forward.

Example output from both of the main exachk reports is shown in the slide.

Numerous additional log files are written which contain raw output from the exachk checks. These files may be referenced in the `exachk.rep` report as shown in the example output on the slide.

A zip file containing all the exachk output files is also written to the directory containing the exachk executable.

DiagTools Overview

- `Diagtools.zip` contains two scripts which collect trace, log and configuration information:
  - `DbmCheck.sh` collects general configuration information about all of the Exadata Storage Servers
  - `diagget.sh` collects Oracle software log and trace files plus OS information from all of the servers
- The scripts are designed to provide a consolidated package of information for use by Oracle Support
- `Diagtools.zip` is available from My Oracle Support note 735323.1
- See demonstration Using DiagTools

`Diagtools.zip` is an archive that contains two scripts which collect trace files, log files and other configuration information from Exadata Database Machine:

- `DbmCheck.sh` connects to all the storage cells and collects general configuration information such as information about cell disks, grid disks and flash cache.
- `diagget.sh` connects to all the Database Machine servers and collects log and trace files related to Oracle clusterware, ASM and database. It also collects various operating system details including information generated by OSWatcher.

The scripts are primarily designed to assist administrators to provide a consolidated package of information about a Database Machine environment which can be used by Oracle Support technicians to assist in resolving support requests.

`DiagTools.zip` is available from My Oracle Support note 735323.1. The scripts rely on the group files (files listing the hostnames of all the Database Machine servers) generated during initial deployment of Database Machine. The scripts should be installed at `/opt/oracle.SupportTools/onecommand` on the first database server. The diagnostic information collected by the scripts is written to date stamped directories under `/opt/oracle.SupportTools/onecommand/diagfiles`

An example of how to use the DiagTools scripts is provided in the demonstration entitled *Using DiagTools*.

Using ADRCI on Exadata Storage Servers

- Exadata Storage Server software uses the Automatic Diagnostic Repository (ADR) structure to manage diagnostic data
- Using ADRCI on an Exadata Storage Server is essentially the same as using it on a database server.
- The default ADR home location is:

```
/opt/oracle/cell<cell_version>/log/diag/asm/cell/<hostname>
```

- See demonstration Using ADRCI on an Exadata Storage Cell

The Automatic Diagnostic Repository (ADR) is a file-based repository for diagnostic data such as traces, dumps, alert logs, health monitor reports, and more. Exadata Storage Server software uses an ADR structure that is essentially the same as the ADR used by Oracle Database software. Like the Oracle Database ADR, the cell ADR can be managed using ADRCI; the ADR command interpreter utility.

Using ADRCI on an Exadata Storage Server is essentially the same as using it on a database server. The main difference is the location of the ADR. On Exadata Storage Servers, `/opt/oracle/cell<cell_version>/log/diag/asm/cell/<hostname>` is the default ADR home location, where `<cell_version>` is a string denoting the cell software version and `<hostname>` is the name of the cell host. Note that each Exadata cell has a symbolic link at `/opt/oracle/cell` that points to the current software version directory (`/opt/oracle/cell<cell_version>`).

An example of how to use ADRCI to collect trace files for a particular incident on an Exadata Storage Server is provided in the demonstration entitled *Using ADRCI on an Exadata Storage Cell*.

Imageinfo Overview

```
[root@exa1cel01 ~]# imageinfo

Kernel version: 2.6.18-194.3.1.0.3.el5 #1 SMP Tue Aug 31 22:41:13 EDT 2010 x86_64
Cell version: OSS_11.2.0.3.0_LINUX.X64_101206.2
Cell rpm version: cell-11.2.2.2.0_LINUX.X64_101206.2-1

Active image version: 11.2.2.2.0.101206.2
Active image activated: 2011-02-24 12:22:44 -0800
Active image status: success
Active system partition on device: /dev/md5
Active software partition on device: /dev/md7

In partition rollback: Impossible

Cell boot usb partition: /dev/sdac1
Cell boot usb version: 11.2.2.2.0.101206.2

Inactive image version: 11.2.2.2.1.110131
Inactive image activated: 2011-02-23 18:30:44 -0800
Inactive image status: success
Inactive system partition on device: /dev/md6
Inactive software partition on device: /dev/md8

Boot area has rollback archive for the version: undefined
Rollback to the inactive partitions: Impossible
```


Software is initially installed on Exadata Database Machine storage servers and database servers using an Exadata-specific imaging process. Exadata Storage Server patch bundles are also provided as software images.

Two utilities are provided to monitor the Database Machine server images:

- `imageinfo` displays information relating to the images currently installed on a server.
- `imagehistory` displays historical information about all the images installed on the server.

The slide shows an example of the output displayed by `imageinfo`. In this example, the command shows information associated with a storage server.

There is significantly less output associated with a database server because database servers have no concept of active and inactive images, and they do not contain a cell boot USB device.

The behavior and output of `imageinfo` can be modified using a series of optional parameters. Execute `imageinfo -h` to obtain a list of the available options.

Imagehistory Overview

```
[root@exalcel01 ~]# imagehistory
Version                               : 11.2.2.2.0.101206.2
Image activation date                 : 2011-01-12 14:56:44 -0800
Imaging mode                          : fresh
Imaging status                        : success

Version                               : 11.2.2.2.1.110131

Image activation date                 : 2011-02-23 18:30:44 -0800
Imaging mode                          : out of partition upgrade
Upgrade logs                          : /var/log/cellos/patch/rollback_20
1102242002_11.2.2.2.0.101206.2_11.2.2.2.1.110131.tar.gz
Imaging status                        : success

Version                               : 11.2.2.2.0.101206.2
Image activation date                 : 2011-02-24 12:22:44 -0800
Imaging mode                          : out of partition rollback
Imaging status                        : success
```


`imagehistory` displays historical information about all the images installed on the server.

The slide shows an example of the output displayed by `imagehistory`. In this example, the command shows information associated with a storage server. From the output it can be seen that the storage server was initially imaged with storage server software version 11.2.2.2.0. Then it was patched to version 11.2.2.2.1. And finally the patch was rolled back leaving the server with the initial software version.

Typically the output associated with a database server would only relate to the initially installed image because database servers are not patched using imaging techniques.

`imagehistory` has only 3 optional parameters:

- `-all` shows all the available image attributes.
- `-help` shows basic command using instructions.
- `-version <version>` outputs information relating to the specified version only

OSWatcher Overview

- OSWatcher collects and archives operating system and network metrics to aid in diagnosing performance issues
- OSWatcher gathers OS data on a regular basis, invoking utilities such as `vmstat`, `netstat` and `iostat`
- OSWatcher runs on all Exadata Database Machine servers with the following default settings:
  - Metric snapshots are recorded every 15 seconds
  - 168 hours (7 days) of archived metric observations are maintained
  - Installation location: `/opt/oracle.oswatcher/osw`
  - Archive location: `/opt/oracle.oswatcher/osw/archive`
- The OSWatcher User Guide is available from My Oracle Support note 301137.1.

OSWatcher is a collection of UNIX shell scripts designed to collect and archive operating system and network metrics to aid in diagnosing performance issues. OSWatcher operates as a set of background processes on the server and gathers OS data on a regular basis, invoking utilities such as `vmstat`, `netstat` and `iostat`.

By default, OSWatcher is configured to run on all Exadata Database Machine storage and database servers with the following settings:

- Metric snapshots are recorded every 15 seconds.
- 168 hours (7 days) of archived metric observations are maintained.

OSWatcher is installed under `/opt/oracle.oswatcher/osw`. Metric observations are maintained in the `archive` subdirectory.

The OSWatcher User Guide is available from My Oracle Support note 301137.1.

Quiz

Exachk is an alternative to using Grid Control:

a. TRUE
b. FALSE

Answer: b

Quiz

Which utility should you use to provide Oracle Support with a consolidated package of information about a Database Machine environment?

a. Exachk
b. DiagTools
c. ADRCI
d. Imageinfo
e. Imagehistory
f. OSWatcher

Answer: b

Summary

In this lesson, you should have learned how to describe the following useful tools:

- Exachk
- DiagTools
- ADRCI
- Imageinfo and Imagehistory
- OSWatcher

Additional Resources

- Lesson Demonstrations
  - Using DiagTools
  - Using ADRCI on an Exadata Storage Cell
- My Oracle Support Notes
  - Oracle Exadata Database Machine exachk or HealthCheck
  - Exadata Storage Server Diagnostic Collection Guide
  - OS Watcher User Guide

20

Backup and Recovery

Objectives

After completing this lesson, you should be able to:

- Describe how RMAN backups are optimized using Exadata Storage Server
- Describe the recommended approaches for disk-based and tape-based backups of databases on Database Machine.
- Describe the recommended best practices for backup and recovery on Database Machine.

Backup and Recovery Overview

- Backup and recovery of databases on Database Machine
  - Use RMAN
  - Typical strategies:
    - Disk-based backups
    - Tape-based backups
    - Hybrid strategy
- Backup and recovery of Database Machine software
  - Database server software
  - Exadata Storage Server software

This lesson focuses on the best practice recommendations for backup and recovery in conjunction with Database Machine. The slide lists the topics that are considered throughout the lesson.

Using RMAN with Database Machine

- Using RMAN on Database Machine is essentially the same as using RMAN elsewhere
  - Same concepts
  - Same commands
- Incremental backup performance is improved
  - Block filtering is offloaded to Exadata Storage Server
  - Fewer blocks need to be processed by RMAN
  - Offload processing is automatic and transparent
- Exadata Hybrid Columnar Compression can assist to further improve backup performance
  - Reduced data size results in smaller, quicker backups

Oracle recommends the use of RMAN for database backup and recovery in conjunction with Database Machine. In essence, using RMAN with Database Machine is the same as using RMAN without Database Machine.

To optimize the performance of incremental backups, the database can offload block filtering to Exadata Storage Server. This optimization is only possible when taking backups using RMAN. The offload processing is done transparently without user intervention. During offload processing, Exadata Storage Server filters out the blocks that are not required for the incremental backup in progress. Therefore, only the blocks that are required for the backup are sent to the database. This can significantly improve backup performance, especially if the proportion of changed blocks is low.

Exadata Hybrid Columnar Compression can assist to further improve backup performance. When a database uses Exadata Hybrid Columnar Compression, then the number of blocks included in the backup is reduced compared to the same data in a database that uses a less space-efficient form of compression, or no compression at all. The reduced data size associated with Exadata Hybrid Columnar Compression results in smaller and ultimately quicker backups.

General Recommendations for RMAN

Use RMAN to backup and recover databases on Database Machine

- Use RMAN incremental backups and block change tracking
- Use an external RMAN recovery catalog repository
- Set `DB_RECOVERY_FILE_DEST_SIZE` to bound space used in the Fast Recovery Area

The following recommendations apply when using RMAN in conjunction with Database Machine:

- **Use RMAN incremental backups and block change tracking:** For fast incremental backups, enable block change tracking. Block change tracking allows RMAN to avoid scanning blocks that have not changed, when creating incremental backups. Also, when performing incremental backups of databases on Database Machine, additional block inspection is offloaded from the database servers. Block change tracking provides the greatest benefit for databases where fewer than 20% of the blocks are changed between incremental backups. You may still benefit by using block change tracking with change rates greater than 20%, but testing is recommended to ensure that backup times are reduced.
- **Use an external RMAN recovery catalog repository:** The RMAN recovery catalog should be hosted on a server outside Database Machine. In practice, it is common to have a server that hosts the RMAN catalog along with other management repositories such as the Oracle Enterprise Manager repository and the Oracle Secure Backup catalog.

- **Set `DB_RECOVERY_FILE_DEST_SIZE` to bound space used in the Fast Recovery Area:** The database writes archived redo log files and any additional recovery files to the Fast Recovery Area. These include any disk backup files such as level 0 image copies and level 1 backup sets as well as Flashback log files (if enabled). It is important that you set the value of this parameter to less than the total free space in the disk group, taking into account at least one disk failure, and preferably one Exadata cell failure. Additionally, if multiple databases are sharing the Fast Recovery Area, ensure that the sum of the space allocated to the different databases is less than the free space in the disk group.

Disk Based Backup Strategy

For disk-based database backups, Oracle recommends:

- Use a Fast Recovery Area
- Perform an initial level 0 (full) backup
- Perform periodic incremental level 1 backups
- Update your level 0 backup by applying the second to last level 1 backup

For disk-based database backups, Oracle recommends incremental backups to a Fast Recovery Area (FRA). The recommended strategy is outlined in the slide. Using this approach, you can achieve a good compromise between the time and effort required during backup and recovery operations. The approach also efficiently manages the amount of storage required for backups.

Remember that the initial configuration procedure for Database Machine creates a disk group for the FRA, and that the initial size of the disk group is determined by the backup method chosen in the configuration worksheet.

Disk Based Backup Recommendations

- Fast Recovery Area (FRA) configuration:
  - Default: FRA disk group striped across all available Exadata Storage Servers along with data disk groups
    - High availability with the best throughput
  - Alternative: FRA disk group and data disk groups on separate Exadata Storage Servers
    - Separation of data and backups at the cost of some throughput
- Additional RMAN recommendations:
  - Instances and channels:
    - Initially, run RMAN using one database instance and two RMAN channels
    - Add another instance with two channels if required
    - Add an additional two channels per instance if required
  - Configure an Oracle Service to use as the RMAN target

By default, the installation and configuration procedure for Database Machine yields a disk group for the Fast Recovery Area (FRA) which is spread across all of the disks in all of the Exadata Storage Servers. In this configuration, the FRA shares each disk with disk groups containing database data files. By default, all the disk groups are created with at least `NORMAL` ASM redundancy (2-way mirroring). The result is that Database Machine can tolerate the loss of an entire Exadata Storage Server and still maintain full data integrity.

Where customers desire a separation between database files and the FRA, the recommended alternative is to configure the disk group containing the FRA on a dedicated set of Exadata Storage Servers and to configure the required data disk groups on separate storage servers. This configuration means that the I/O for a single workload can no longer benefit from being striped across all the storage servers, however it also provides additional protection if multiple simultaneous failures affect either the database or FRA. Multiple simultaneous failures to both the database and FRA can still result in data loss.

Another strategy is to purchase additional high capacity Exadata Storage Servers specifically to store the FRA. This allows your databases to leverage the full capacity of the Database Machine storage servers. To implement this strategy, the FRA storage servers need to be hosted in a separate rack and they need to be connected to the Database Machine storage network using the spare ports on the Database Machine InfiniBand switches.

In addition to the general recommendations for RMAN in conjunction with Database Machine, Oracle recommends the following configuration for disk based backups:

- Testing shows that optimal backup rates are achieved with 2 database instances and 2 to 4 RMAN channels per instance. For the 2 database instances designated as backup servers, less than 10% CPU and less than 40% IO bandwidth are utilized. During backup operations, I/O intensive parallel queries should avoid these designated backup servers. Start by running RMAN on one database instance using two RMAN channels. If greater backup throughput is required, use a second instance with two more channels. Finally, if required add two more channels per instance.
- Configure an Oracle Service to run against specific database instances in the cluster. The service is used by RMAN to automatically spread the backup load evenly among the target instances offering the service.
- For incremental backups, use the RMAN `BACKUP` command option `FILESPERSET 1`. This backup option will allow faster single file restore operations.
- Set the initialization parameter `_file_size_increase_increment=2143289344` to optimize the space used when incremental (level 1) backups are written to the FRA.
- For systems running Oracle Database release 11.2.0.1, set the initialization parameter `_backup_ksfq_bufsz=4194304` to optimize read performance. Do not set the initialization parameter `_backup_ksfq_bufsz` on systems running Oracle Database release 11.2.0.2 or later.
- For systems running Oracle Database release 11.2.0.1, set the initialization parameter `_backup_ksfq_bufcnt` as follows:
  - On a Quarter Rack configuration set `_backup_ksfq_bufcnt=32`.
  - On a Half Rack, Full Rack, or multiple rack configuration set `_backup_ksfq_bufcnt=64`.

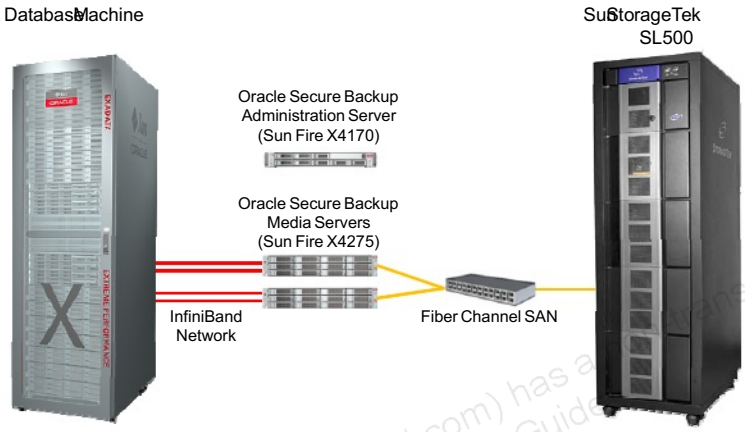
Tape Based Backup Strategy

For tape based database backups, Oracle recommends:

- Use media management software that is integrated with RMAN, like Oracle Secure Backup
- Perform periodic level 0 (full) database backups
- Perform more frequent cumulative level 1 backups and also backup the Oracle Secure Backup catalog

For tape based database backups, Oracle recommends the use of Oracle Secure Backup or other media management software that is integrated with RMAN. The recommended backup strategy is to perform periodic (weekly) level 0 (full) database backups. In addition, more frequent (daily) cumulative level 1 backups should be taken along with a backup of the Oracle Secure Backup catalog if it is the media management software which is used.

Tape Based Backup Architecture



The slide illustrates the recommended reference architecture for a tape based backup configuration using Oracle Secure Backup. The key recommendations associated with the configuration follow:

- Connect the media servers to Database Machine using the high-performance InfiniBand network. The InfiniBand connection to each media server should be bonded for high availability. Alternatively, you can use a Gigabit Ethernet network between the media servers and Database Machine, however if you use this configuration be aware that the network will likely be the constraining factor for backup and recovery performance.
- Consider the bandwidth capability of the InfiniBand connection from the Database Machine to the media server, compared to the SAN links between the media server and the tape library. Testing shows that data rates of approximately 3 GB per second can be sustained over an InfiniBand link. Hence you need to provide the appropriate number of SAN links and configure them to deliver the required throughput.
- Typically, backup performance is limited by tape drive throughput. Backup performance scales when you add more tape drives and RMAN channels. Allocate a sufficient number of tape drives so the media servers can achieve their maximum backup and restore rates. Add tape drives until the bandwidth of the media servers is saturated.
- Start with at least two media servers. Add media servers if you have enough tape drives to keep them busy without saturating Database Machine resources.

Tape Based Backup Recommendations

- Media server to Database Machine network configuration recommendation:
  - Use InfiniBand for the best backup rates:
    - Configure bonding of the media server InfiniBand interfaces
    - Update OpenFabrics Enterprise Distribution on the media server
    - Configure IP over InfiniBand connected mode for best performance
    - Set the message transfer unit (MTU) size to 65520 for the InfiniBand interface
    - Configure the media management software to use the InfiniBand network
- Media server SAN configuration recommendation:
  - Configure persistent bindings for tape devices

Media servers can be directly connected to the Database Machine InfiniBand fabric by adding an InfiniBand Quad Data Rate (QDR) host channel adapter (HCA) to the media server. Use InfiniBand for the best backup rates, especially for large databases that require fast backup rates and low CPU overhead.

Follow these recommendations:

- For high availability, connect the HCA to two different Database Machine InfiniBand leaf switches to eliminate the switch as a single point of failure. This provides transparent failover if connectivity is lost to one of the ports. Configure bonding of the media server InfiniBand (IB) interfaces. Use Active-Passive bonding.
- You must use an OpenFabrics Enterprise Distribution (OFED) version on the media server that is compatible with the version found in Database Machine. For details, refer to My Oracle Support bulletin 888828.1.
- Configure the InfiniBand interface to use IP over InfiniBand connected mode for best performance. On Linux, edit the /etc/ofed/openib.conf file so that it contains the entry `SET_IPOIB_CM=yes`. Reboot the server to enable the setting. To verify the setting, check the contents of `/sys/class/net/ib0/mode` and `/sys/class/net/ib1/mode`. Both files should contain the entry `connected`.

- Set the message transfer unit (MTU) size to 65520 for the InfiniBand interface. On Linux, edit the `/etc/sysconfig/network-scripts/ifcfg-ib*` and `/etc/sysconfig/network-scripts/ifcfg-bond0` files so that they contain the entry `MTU=65520`. Verify the MTU setting by examining the output of the `ifconfig` command for the InfiniBand interfaces.
- To direct the backup and restore traffic over the InfiniBand network, configure the media management software to favor InfiniBand. Note that each media management software type has its own method of enabling this configuration. For example, Oracle Secure Backup has the concept of a preferred network interface, which can be set on the media server for a specific list of clients. Other media management software may require this configuration to be defined when the software is installed. See the media management software for information about how to direct traffic over a particular network.

For SAN attached tape devices, configure persistent bindings so the device address does not change. If the device address changes, the media servers cannot access the device unless you update the device configuration within the media server software. Therefore, it is very important that your environment maintains consistent device addresses.

Persistent bindings are a part of the SAN infrastructure setup. Typically, persistent bindings are configured through the HBA or the operating system. The configuration steps will vary by platform and vendor. See My Oracle Support note 971386.1 for an example of creating persistent bindings for device attachments.

Connecting the Media Server Using Ethernet

- Ethernet can be used if throughput is sufficient:
  - GigE: Expect up to 120 MB/sec from each interface
  - 10*g*igE: Expect up to 1 GB/sec from each interface
- Recommendations:
  - Use a dedicated backup network:
    - Configure dedicated network interfaces on each Database Machine database server
  - Use bonded network interfaces:
    - Configure LACP for maximum throughput:
      - Availability still maintained if one link is lost
      - Configuration required on media server, network switch and database servers
    - Use Active-Passive bonding for high availability otherwise:
      - Configure database servers and media server
      - No specific switch configuration required

Alternatively, you can use a Gigabit Ethernet (GigE) or 10 Gigabit Ethernet (10 *g*igE) network between Database Machine and the media servers if throughput is sufficient.

Using a dedicated backup network eliminates any impact on the client access network. Often, a dedicated backup network is already in place. The maximum throughput with the GigE network is approximately 120 MB/sec for each interface. For 10 *g*igE, the maximum throughput is approximately 1 GB/sec for each interface. Hence a Full Rack Exadata Database Machine X2-2 can achieve throughput up to 960 MB/sec using GigE, or 8 GB/sec using 10 GigE, with a single dedicated network interface on each database server.

Higher throughput can be achieved using Link Aggregation Control Protocol (LACP). LACP enables a bonded network interface to use both network channels simultaneously. To use LACP you must configure a bonded network interface on each database server and set the `BONDING_OPTS` parameter to `mode=4` in the associated bonding configuration file (`ifcfg-bond1` for example). You must also configure LACP on the associated network switch and on the media server network interfaces. If LACP is not used, use Active-Passive bonded network interfaces to provide high availability in case of network interface failure.

Be aware that the ability to use a dedicated, bonded network interface for backup and recovery depends on the rest of the Database Machine configuration. For example, there may not be sufficient network interfaces available if bonding is also used for the client access network. In this case, customers must balance the competing requirements.

Tape Based Backup Recommendations

- Run RMAN across all the available database instances
  - Create a Database Service that runs across the cluster

```
$ srvctl add service -d <dbname> -s <service name>
  -r <instance1>, ... ,<instancen>
```

  - Use the service name and SCAN address to connect RMAN

```
$ rman target sys/<passwd>@<scan_address>/<service_name>
  catalog ...
```

- Allocate one RMAN channel per tape drive
- Configure IORM and DBRM to control resource allocation between backups and application workloads

Further recommendations:

- For your RAC databases, configure RMAN to run across all the database instances. It is recommended that you create a Database Service to run across the RAC cluster. Then when running RMAN, use the service name and SCAN address in the connect string for the RMAN target as shown in the slide.
- Allocate one RMAN channel per tape drive. A single RMAN channel in Database Machine can stream dataat a rate of approximately 750 MB/sec fromDatabase Machine to the media server. Typical tape drive backup rates are between 100 MB/sec and 240 MB/sec, depending on the drive type and compression options. Note that tape drive compression becomes less effective when backing up tables that are compressed at the database level.
- If Database Machine resources must be prioritized between application workloads and backups, then configure I/O Resource Manager (IORM) and Database Resource Manager (DBRM). This is more likely to be required in cases where time-consuming backups to large databases must run at the same time as production application workloads.

Hybrid Backup Strategy

- A hybrid backup strategy combines the disk-based and tape-based backup approaches:
  - Level 0 (full) database backups are stored on tape
    - Relatively cheap backup media that can be stored off site
  - Level 1 incremental backups are stored on disk in the FRA
    - Readily available with high performance access
- Follow recommendations for both approaches

ORACLE

A hybrid backup strategy combines the disk-based and tape-based backup approaches. For example, a reasonable hybrid approach would result in level 0 (full) database backups being stored on tape, while the level 1 incremental backups would be stored on disk in the FRA. The hybrid approach also combines the benefits of each storage type as indicated on the slide. If you choose to implement a hybrid backup approach you should follow the recommendations for disk-based backups and tape-based backups.

Restore and Recovery Recommendations

- Restore into existing data files if possible
  - Restore performance is better
  - Restore using all database instances
- If no existing data files are present, restore using up to 2 database instances
- Recommended number of RMAN channels:
  - For disk-based restoration, use 2 to 8 RMAN channels per database instance
  - For tape-based restoration, set the number of RMAN channels based on the total number of tape drives

Higher restore rates are achieved when avoiding the overhead of initial data file allocation. If existing files are present prior to database restoration, do not delete the data files and perform database restoration into the existing files to take advantage of this optimization.

With pre-existing files, create a restore service across all database instances. If no existing data files are present prior to the restore operation, create a restore service with only two database instances. Use 2 to 8 RMAN channels per database instance for disk-based restoration. For tape-based restoration, the number of channels should be set to the number of tape drives.

Backup and Recovery of Database Machine Software

- Database Server software
  - Perform file system level backup and recovery
    - Use your chosen file system backup management software or Oracle Secure Backup can be used
    - Copies of the Oracle Cluster Registry are automatically maintained on the Database Server file system and should be included in Database Server file system backups
- Exadata Storage Server software
  - File system level backups are not recommended
    - System areas are mirrored
      - Use CellCLI commands to recover if one system disk fails
    - Use the Exadata Software Rescue Procedure if both system disks fail simultaneously
      - The rescue procedure uses a built-in USB flash drive

It is recommended that you use file system level backup and recovery techniques for the database servers. Use your chosen file system backup management software and infrastructure or use Oracle Secure Backup if desired. Note that a backup of the Oracle Cluster Registry, which also contains Voting Disk information, is automatically maintained on the file system of the first database server and should be included in your database server file system backups. The default location for the Oracle Cluster Registry backup is `<Grid_Home>/cdata/<Cluster_Name>`, where:

- `<Grid_Home>` is the location of the Grid Infrastructure software as specified during the initial configuration of Database Machine. The default `<Grid_Home>` location is `/u01/app/11.2.0/grid`.
- `<Cluster_Name>` is the name of your cluster. This is the same as the DB Machine Prefix specified during the initial configuration of Database Machine.

File system level backups are not recommended for Exadata cell software. Exadata Storage Server maintains mirrored system areas on separate physical disks. In the rare event that both system disks fail simultaneously, you must use the rescue functionality provided on the CELLBOOT USB flash drive that is built into every Exadata Storage Server.

Quiz

To facilitate disk-based backups you can install additional high capacity Exadata Storage Servers into the available space in a Half Rack Database Machine based on high performance disks:

a. True
b. False

Answer: b

Adding Exadata Storage Servers into the same rack as a Half Rack Database Machine is only supported using the Half Rack to Full Rack upgrade package, and the resulting Full Rack Database Machine can only utilize one disk type; high performance or high capacity. To add high capacity cells to a Half Rack Database Machine using high performance cells would require the addition of an Exadata Storage Expansion Rack or to install the high capacity cells into a separate rack and connect them to the Database Machine storage network using the spare ports on the Database Machine InfiniBand switches.

Quiz

How many RMAN channels should you use for tape-based backups?

a. 2 per database instance
b. 4 per database instance
c. 1 per tape drive

Answer: c

Summary

In this lesson, you should have learned how to:

- Describe how RMAN backups are optimized using Exadata Storage Server
- Describe the recommended approaches for disk-based and tape-based backups of databases on Database Machine.
- Describe the recommended best practices for backup and recovery on Database Machine.

Additional Resources

- Lesson Demonstrations
  - Backup Optimization Using RMAN and Exadata Storage Server
  - Recovery Optimization Using RMAN and Exadata Storage Server
- My Oracle Support Notes
  - Database Machine and Exadata Storage Server 11*g* Release 2 (11.2) Supported Versions
  - OSB - Create persistent bindings for device attachments on OEL

Practice 20 Overview: Using RMAN Optimizations for Database Machine

In these practices, you will examine the backup and recovery optimizations that are enabled when Oracle Recovery Manager (RMAN) is used in conjunction with Exadata storage.

21

Exadata Database Machine Maintenance Tasks

ORACLE

Objectives

After completing this lesson, you should be able to perform the following Database Machine maintenance tasks:

- Powering Database Machine on and off
- Safely shutting down a single Exadata Storage Server
- Replacing a damaged physical disk on a cell
- Replacing a damaged flash card on a cell
- Moving all disks from one cell to another
- Using the Exadata cell software rescue procedure

Database Machine Maintenance Overview

- Maintaining Database Machine is similar to maintaining any other clustered Oracle Database environment
- Database Machine-specific tasks outlined in this lesson:
  - Powering Database Machine on and off
  - Safely shutting down a single Exadata Storage Server
  - Replacing a damaged physical disk on a cell
  - Replacing a damaged flash card on a cell
  - Moving all disks from one cell to another
  - Using the Exadata cell software rescue procedure
- Additional references:
  - *Oracle Exadata Database Machine Owner's Guide*
  - My Oracle Support

In many respects, maintaining Exadata Database Machine is similar to maintaining any other clustered Oracle Database environment. The procedures for maintaining Oracle clusterware, ASM and RAC are essentially the same on Database Machine as they are on other platforms, the main difference being the references to the Exadata cell objects.

This lesson focuses on a series of Database Machine-specific maintenance tasks that administrators are most likely to encounter. Additional less-common maintenance tasks are also documented in the *Oracle Exadata Database Machine Owner's Guide* Administrators should also consult My Oracle Support for notes on other maintenance issues.

Note: Patching guidelines for Exadata Database Machine are considered in a separate lesson later in the course.

Powering Database Machine Off and On

- Power Off sequence:
  1. Database servers

```
# <Grid_Home>/bin/crsctl stop cluster
# shutdown -h -y now
```

     – Ensure all database servers are shut down before proceeding
  2. Exadata Storage Servers

```
# shutdown -h -y now
```

     – Ensure all storage servers are shut down before proceeding
  3. Rack, including network switches
- Power On sequence:
  1. Rack, including network switches
     – Apply power for a few minutes before proceeding
  2. Exadata Storage Servers
     – Check that all cells are running before proceeding
  3. Database servers

The slide lists the recommended sequence for powering Database Machine off and on.

When performing either sequence it is important to ensure that each step is completely finished before moving on to the next step. Failure to execute steps in the proper order may result in Database Machine not functioning correctly.

When powering on the Exadata Storage Servers and the database servers, power can be applied by pressing the power button on the front of each server, or by logging into the ILOM interface for each server and issuing the `reset /SYS` command.

To power the rack on or off, use the switches located on the Power Distribution Units (PDUs) which are located at the rear of the rack.

Safely Shutting Down a Single Exadata Storage Server

- Safe shutdown sequence:
  1. Check to make sure that shutting down the storage server will not take any ASM disk group offline

```
CellCLI > LIST GRIDDISK WHERE asmdeactivationoutcome != 'Yes'
```

  2. Make all the grid disks inactive

```
CellCLI> ALTER GRIDDISK ALL INACTIVE
```

  3. Verify that all the grid disks are inactive

```
CellCLI> LIST GRIDDISK WHERE STATUS != 'inactive'
```

  4. Shut down the storage server
- Startup sequence:
  1. Start the storage server
     - Cell services start automatically
  2. Make all the grid disks active

```
CellCLI> ALTER GRIDDISK ALL ACTIVE
```

  3. Verify that all the grid disks are active

```
CellCLI> LIST GRIDDISK ATTRIBUTES name, asmmodestatus
```

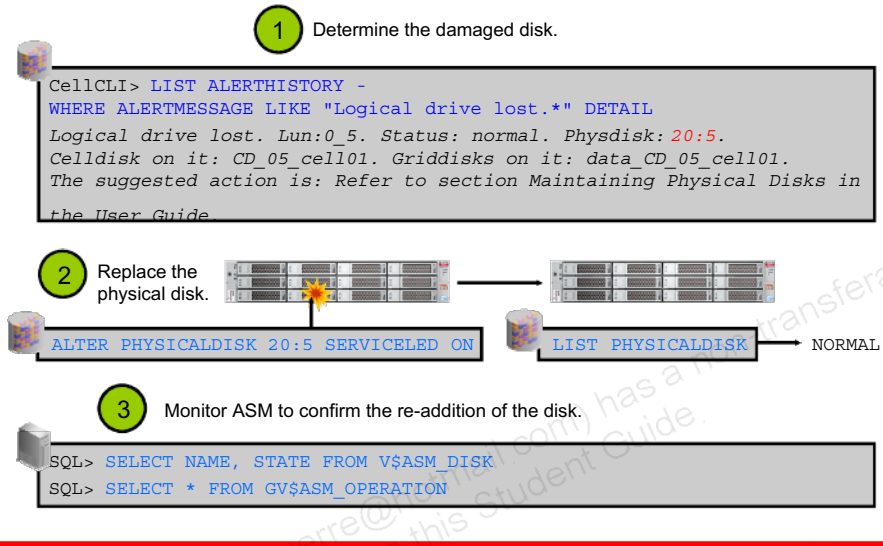
ORACLE

In some maintenance scenarios a single Exadata Storage Server must be shut down in isolation. For example, a hardware component, such as a flash memory card or a disk controller, may be indicating an intermittent fault so the storage server must be shut down to replace the component while the rest of the system continues to support business transactions. In these cases the desired result is that the Database Machine environment continues to support processing activities without a substantial impact on the system users.

To safely and gracefully shut down a single Exadata Storage Server use the commands shown in the slide. When checking to make sure that shutting down the storage server will not take any ASM disk group offline you should expect to see no output from the first `LIST GRIDDISK` command. If any output is returned, then it is not safe to take Exadata Storage Server offline because proper Oracle ASM disk group redundancy will not be maintained. Taking Exadata Storage Server offline when one or more grid disks are in this state will cause ASM to dismount the affected disk group, causing the databases to shut down abruptly. In this case you would need to analyze the situation and bring other grid disks online in order to proceed safely.

After the Exadata Storage Server is restarted and the grid disks are reactivated, you should check that all the grid disks return to an `asmmodestatus` of either `ONLINE` or `UNUSED`.

Replacing a Damaged Physical Disk



Replacing a physical disk due to problem or failure is probably the most likely hardware maintenance operation that an Exadata Storage Server might ever require. Assuming you are using ASM redundancy, the procedure to replace a problem disk is quite simple.

The first step requires that you identify the problem disk. This could occur in a number of ways:

- Hardware monitoring using ILOM may report a problem disk.
- If a disk fails, an Exadata alert is generated. The alert includes specific instructions for replacing the disk. If you have configured the system for alert notifications, the alert will be sent to the designated email address or SNMP target. The `LIST ALERTHISTORY` command could also be used to identify a failed disk.
- The `LIST PHYSICALDISK` command may identify a disk reporting an abnormal status. Even if the cell is still functioning, the problem may be a precursor to a disk failure.
- The `CALIBRATE` command may identify a disk delivering abnormally low throughput or IOPS. Even if the cell is still functioning, a single bad physical disk can degrade the performance of other good disks so you may decide to replace the identified disk. Note that running `CALIBRATE` at the same time as the cell is active will impact performance.

You can use the `ALTER PHYSICALDISK` command to light a service LED which assists in correctly translating the disk name to the corresponding physical disk location.

When a disk fails, the Oracle ASM disks associated with the grid disks on the physical disk are automatically dropped with the `FORCE` option, and an Oracle ASM rebalance follows to restore the data redundancy. This process is known as proactive disk quarantine.

If you wish to replace a disk that is delivering poor performance, but which has not been isolated by proactive disk quarantine, you must manually drop the associated grid disks using the `ALTER DISKGROUP ... DROP DISK` command.

After you have identified the problem disk, you can replace it. When you remove the disk, you will get an alert. When you replace a physical disk, the disk must be acknowledged by the RAID controller before it can be used. This does not take a long time, and you can use the `LIST PHYSICALDISK` command to monitor the status until it returns to `NORMAL`.

The grid disks and cell disks that existed on the previous disk in the slot will be automatically re-created on the new disk. If these grid disks were part of an Oracle ASM disk group with `NORMAL` or `HIGH` redundancy, they will be added back to the disk group and the data will be rebalanced based on disk group redundancy and the `asm_power_limit` parameter.

Re-creating the ASM disk and rebalancing the data may take some time to complete. You can monitor the progress of these operations within ASM. You can monitor the status of the disk as reported by `V$ASM_DISK.STATE` until it returns to `NORMAL`. You can also monitor the rebalance progress using `GV$ASM_OPERATION`.

Review the following considerations when replacing a failed disk:

- The disk could be dropped by ASM, and the rebalance operation may have been successfully run. Check the Oracle ASM alert logs to confirm this. After the failed disk is replaced, a second rebalance is required.
- The disk could be dropped, and the rebalance operation is currently running. Check the `GV$ASM_OPERATION` view to determine if the rebalance operation is still running. In this case the rebalance operation following the disk replacement will be queued.
- The disk could be dropped by ASM, and the rebalance operation failed. Check `GV$ASM_OPERATION.ERROR` to determine why the rebalance operation failed. Monitor the rebalance operation following the disk replacement to ensure it runs.
- Rebalance operations from multiple disk groups can be done on different Oracle ASM instances in the same cluster if the physical disk being replaced contains grid disks from multiple disk groups. Multiple rebalance operations cannot be run in parallel on just one Oracle ASM instance. The operations will be queued for the instance.

Replacing a Damaged Flash Card

1 Determine the damaged flash card.

```
CellCLI> LIST PHYSICALDISK DETAIL
         name:         [9:0:2:0]
         diskType:     FlashDisk
         ...
         slotNumber:   "PCI Slot: 1; FDOM: 2"
         status:       critical
```

2 Power down the cell.

3 Replace the flash card.

4 Power up the cell.

5 If the card contained a flash-based grid disk, monitor ASM to confirm the re-addition of the disk.

```
SQL> SELECT NAME, STATE FROM V$ASM_DISK
SQL> SELECT * FROM GV$ASM_OPERATION
```

ORACLE

Each Exadata Storage Server is equipped with 4 PCI flash memory cards. Each card has 4 flash modules (FDOMs) for a total of 16 flash modules on each cell.

Identifying a damaged flash module is similar to identifying a damaged physical disk. Hardware monitoring using ILOM or a drop in performance indicated by the `CALIBRATE` command may indicate a problem. If a failed FDOM is detected, an alert is generated. The alert message includes notice of any flash-based grid disks that were on the flash module.

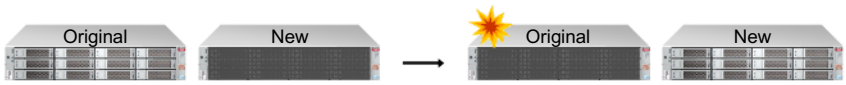
As shown in the slide, a damaged flash module can also be reported using the `LIST PHYSICALDISK DETAIL` command. The `slotNumber` attribute shows the PCI slot and the FDOM number. In this example, the `status` attribute indicates a critical fault.

If there were no grid disks on the flash module, the flash module was probably being used for Exadata Smart Flash Cache. In this case, the bad flash module results in a decreased amount of flash memory on the cell. The performance of the cell is affected proportional to the size of flash memory lost, but the database and applications are not at risk of failure.

Although technically the PCI slots in a Exadata Storage Server are hot-replaceable, it is recommended to power down the cell while replacing a damaged flash card.

After replacing the card and powering up the cell, no additional steps are required to re-create any flash-based grid disks, however you should monitor ASM to confirm the re-addition of any flash-based grid disks and to verify that all rebalance operations complete successfully.

Moving All Disks from One Cell to Another



1. Make the grid disks inactive:

```
CellCLI> ALTER GRIDDISK ALL INACTIVE
```

2. Back up the operating system configuration files that may change when the new cell is booted.
3. Move the disks from the srcinal cell to the new cell.
   - Ensure the system disks occupy the first two slots.
4. Boot the new cell.
5. Restart Exadata cell services:

```
CellCLI> ALTER CELL RESTART SERVICES ALL
```

6. Import the cell disks:

```
CellCLI> IMPORT CELLDISK ALL
```


You may need to move all drives from one storageserver to another server. Thismay be necessary when there is a chassis-level component failure, or when troubleshooting a hardware problem. To move the drives, perform the following steps:

1. If possible, use the `ALTER GRIDDISK ALL INACTIVE` command to make the grid disks inactive.
2. If possible, back up `/etc/hosts`, `/etc/modprobe.conf`, and the files in `/etc/sysconfig/network` and `/etc/sysconfig/network-scripts`. This is mainly a precautionary step and is also useful in case you wish to move the disks back to the srcinal chassis.
3. Shut down the srcinal server and move the disks to the new server.
   Caution: Ensure the first two disks, which are the system disks, are in the same first two slots. Failure to do so causes the Exadata Storage Server to function improperly.
4. Start the cell. The cell operating system will be automatically reconfigured to suit the new server hardware.
5. Restart the cell services using `ALTER CELL RESTART SERVICES ALL`.
6. Import the cell disks using `IMPORT CELLDISK ALL`.

If you are using ASM redundancy and the procedure is completed before the amount of time specified in the `DISK_REPAIR_TIME` ASM initialization parameter, then the ASM disks will be automatically brought back online and updated with any changes made during the cell outage.

Using the Exadata Cell Software Rescue Procedure

- Every Exadata Storage Server is equipped with a CELLBOOT USB flash drive to facilitate cell rescue
  - Required if both system disks fail simultaneously
  - Use with extreme caution
- To perform cell rescue:
  1. Connect to the Exadata Storage Server using the console
  2. Boot the cell, and as soon as you see the "Oracle Exadata" splash screen, press any key on the keyboard
  3. In the displayed list of boot options, select the last option, `CELL_USB_BOOT_CELLBOOT_usb_in_rescue_mode`, and press Enter
  4. Select the rescue option, and proceed with the rescue
  5. At the end of the rescue process, ensure that the cell boots from the system disks
  6. Reconfigure the cell

Exadata Storage Server maintains mirrored system areas on separate physical disks. If one system area becomes corrupt or unavailable, the mirrored copy is used to recover.

In the rare event that both system disks fail simultaneously, you must use the rescue functionality provided on the CELLBOOT USB flash drive that is built into every Exadata Storage Server.

It is important to note the following when using the rescue procedure:

- Use extreme caution when using this procedure, and pay attention to the prompts. The rescue procedure can potentially rewrite some or all of the disks in the cell. If this happens, then you can irrevocably lose the contents of those disks. Ideally, you should use the rescue procedure only with assistance from Oracle Support Services.
- The rescue procedure does not destroy the contents of the data disks or the contents of the data partitions on the system disks unless you explicitly choose to do so during the rescue procedure.
- The rescue procedure restores the Exadata Storage Server software to the same release. This includes any patches that existed on the cell as of the last successful boot.

- The following is not be restored using the rescue procedure:
  - The crash kernel support rpms `kernel-debuginfo-common`, and `kernel-debuginfo`. You will need to reinstall them. These cannot be restored due to space limitations on the CELLBOOT USB flash drive.
  - Some cell configuration details, such as alert configurations, SMTP information, and administrator e-mail address. Note that the cell network configuration is restored, along with SSH identities for the cell, and the `root`, `celladmin` and `cellmonitor` users.
  - ILOM configurations. Typically, ILOM configurations remain undamaged even in case of Exadata software failures.
- The rescue procedure does not examine or reconstruct data disks or data partitions on the system disks. If you have data corruption on the grid disks, then do not use the rescue procedure. Instead use the database backup and recovery procedures.

The following rescue options are available for the rescue procedure:

- Partial reconstruction recovery: During partial reconstruction recovery, the rescue process re-creates partitions on the system disks and checks the disks for the existence of a file system. If a file system is discovered, then the process attempts to boot. If the cell boots successfully, then you use the CellCLI commands, such as `LIST CELL DETAIL`, to verify the cell is usable. You must also recover any data disks, as appropriate. If the boot fails, then you must use the full srcinal build recovery option.
- Full srcinal build recovery: This option rewrites the system area of the system disks to restore the Exadata software. It also allows you to erase any data on the data disks, and data partitions on the system disks.
- Re-creation of the CELLBOOT USB flash drive: This option is used to make a copy of the CELLBOOT USB flash drive.

To perform a rescue using the CELLBOOT USB flash drive:

1. Connect to Exadata using the console.
2. Boot the cell, and as soon as you see the "Oracle Exadata" splash screen, press any key on the keyboard. The splash screen remains visible for only 5 seconds.
3. In the displayed list of boot options, scroll down to the last option, `CELL_USB_BOOT_CELLBOOT_usb_in_rescue_mode`, and press Enter.
4. Select the rescue option, and proceed with the rescue.
5. When prompted at the end of the rescue process proceed as follows:
   a. Choose to enter a shell. Do not choose the reboot option at this point.
   b. Run the `reboot` command from the shell.
   c. During reboot, but before you see the "Oracle Exadata" splash screen, press F8 to access the boot device selection menu.
   d. Select the RAID controller as the boot device.
6. After a successful rescue, you must reconfigure the cell to return it to the pre-failure configuration, and reinstall the `kernel-debuginfo` and `kernel-debuginfo-common` rpms to use crash kernel support. If you chose to preserve the data when prompted by the rescue procedure, then import the cell disks. If you chose not to preserve the data, then you should create new cell disks, and grid disks.

Quiz

When shutting down an Exadata Database Machine, the Exadata Storage Servers must be shut down first:

a. True
b. False

Answer: b

Quiz

An Exadata Storage Server should be shut down to replace failed hardware components other than hard disk drives:

a. True
b. False

Answer: a

While the flash memory cards inside Exadata Storage Server are hot-swappable, Oracle recommends that cells should be shut down to replace hardware components inside the chassis.

Quiz

If an Exadata Storage Server disk fails, which of the following are true?

a. The associated ASM grid disks are automatically dropped and an ASM rebalance occurs in order to quickly restore redundancy
b. The disk may be replaced without shutting down the storage server
c. The storage server must be shut down to replace the disk
d. Multiple ASM instances can participate in the rebalance operation of a single disk group

Answer: a, b

Summary

In this lesson, you should have learned how to perform the following Database Machine maintenance tasks:

- Powering Database Machine on and off
- Safely shutting down a single Exadata Storage Server
- Replacing a damaged physical disk on a cell
- Replacing a damaged flash card on a cell
- Moving all disks from one cell to another
- Using the Exadata cell software rescue procedure

22

Patching Exadata Database Machine

Objectives

After completing this lesson, you should be able to:

- Describe how software in maintained on different Database Machine components
- Locate recommended patches for Database Machine
- Describe the recommended patching process for Database Machine
- Describe the characteristics of an effective test system

Patching and Updating Overview

- There are three categories of software that must be maintained in Database Machine:
  - Software and firmware on the Exadata Storage Servers
  - Software and firmware on the database servers
  - Software and firmware for other components, like the InfiniBand switches for example
- Compatibility between these different pieces of software needs to be maintained
- Patches and updates are rolling in nature wherever possible
- Key information is maintained in My Oracle Support bulletin 888828.1

There are three broad categories of software that must be maintained in Database Machine. That is, software and firmware in the Exadata Storage Servers, database servers and other components like the InfiniBand switches. Compatibility between these different pieces of software is vital for Database Machine to function.

Wherever possible, Database Machine patches and updates are rolling in nature, meaning that they are applied in a manner that facilitates ongoing system availability by rolling the patch across the environment one server at a time rather affecting all the servers at once.

Important information relating to Database Machine patches and updates is maintained in My Oracle Support bulletin 888828.1. This bulletin is constantly updated with new information. Oracle recommends that all Database Machine customers should sign up for an automated alert when this bulletin changes.

The remainder of this lesson outlines additional recommendations regarding Database Machine patching and updating.

Maintaining Exadata Storage Server Software

- Exadata Storage Server patches:
  - Are complete software images
  - Maintain consistency across all cell components
  - Are released independently from Oracle Database patches
  - Have dependencies on other software and firmware versions
- Most Exadata Storage Server patches can be applied while databases remain up and running
- Firmware levels in Exadata Storage Server are maintained automatically
- No additional software, RPMs or otherwise, should be installed on Exadata Storage Server

Patches for Exadata Storage Server are provided as complete software images which contain updates to the Linux operating system, cell server software, InfiniBand software and component firmware. This ensures that all the software and firmware components on an Exadata Storage Server remain consistent with each other.

Exadata Storage Server patches are supplied independent of Oracle Database Server patches. However, an Exadata Storage Server patch may require a specific Oracle Database Server patch level, or database server firmware or operating system version. Details are provided in the patch documentation and in My Oracle Support bulletin 888828.1.

Exadata Storage Server patches can generally be applied in a rolling manner, while the databases continue to run. A rolling patch is performed sequentially on each storage server. The rolling patch process relies on ASM redundancy to maintain disk group availability while each storage server is updated and rebooted.

Patches may also include instructions for parallel installation on multiple Exadata Storage Servers. This approach requires down time but may be preferred if a scheduled maintenance window exists for the system.

Exadata Storage Server automatically maintains firmware within the server. Firmware levels are periodically checked while Exadata Storage Server is running and the correct firmware is automatically applied to components when the server reboots. Disk firmware is also automatically updated when a disk is replaced.

Oracle recommends that no additional software, whether RPMs or otherwise, should be installed on Exadata Storage Server.

Maintaining Database Server Software

- You can patch and update the database server software as you would for an Oracle Database server outside of Database Machine
  - Oracle Database patches applied by OPatch
    - Double-check compatibility of patches with Database Machine
    - Oracle supplies Bundle Patches for Database Machine
      - Periodic bundle of database patches recommended for Database Machine
  - Operating system and firmware updated using regular channels
    - Must maintain consistency with InfiniBand (OFED) software
- Check Exadata Storage Server patches for database server firmware and operating system updates

In essence, you can patch and update the database server software using the same techniques and approaches as you would for an Oracle Database server outside of Database Machine.

OPatch is used to apply and manage Oracle Database patches. It is recommended that you raise a Service Request with Oracle Support to verify the compatibility of patches with Database Machine. In addition to regular database patches and updates, Oracle supplies periodic Bundle Patches for Database Machine. These bundle a series of database patches that are recommended for use in conjunction with Database Machine. Oracle recommends that you should apply Bundle Patches as part of your patching regime.

Operating system and firmware patches should be managed using standard patching practices. My Oracle Support bulletin 888828.1 lists current constrains and requirements for operating system and firmware patches. The main consideration is to ensure that any patches or updates maintain consistency with the OFED software version which underpins the InfiniBand network.

Some Exadata Storage Server patches also include database server firmware and operating system updates. This may occur, for example, when the firmware for the InfiniBand HCA is updated on Exadata Storage Server and the update must also be applied to the database servers.

Assisted Patching Using OPlan

- OPlan is a utility that provides step-by-step patching instructions specific to the target environment
  - Can create instructions for Apply and Rollback
- OPlan works in conjunction with Exadata Database Machine recommended Bundle Patches
  - Starting with 11.2.0.2 Bundle Patch 2
- Using OPlan:
  - As the Oracle software owner:
    - Download the Bundle Patch to a local directory
    - Set the `$ORACLE_HOME` environment variable and execute:

```
$ORACLE_HOME/oplan/oplan generateApplySteps <bundle patch location>
```

    - Locate the customized patch installation instructions at:

```
$ORACLE_HOME/cfgtoollogs/oplan/<TimeStamp>/InstallInstructions.txt
```


OPlan is a utility that facilitates the patch installation process by providing step-by-step patching instructions specific to your environment. OPlan works in conjunction with Database Machine recommended Bundle Patches.

Without OPlan, administrators must read the patch documentation and translate the generic patch installation instructions into specific commands suitable for the target environment. For example, references to the Oracle home directory or the path to the patch directory must be translated into the appropriate directory paths. Also, administrators may be required to choose between a number of options and evaluate the configuration of the environment to generate the right set of commands.

OPlan assists administrators by automatically analyzing the environment and collecting the required configuration information. Using this information, OPlan generates a set of instructions and commands which is customized specifically for the target configuration. OPlan can create the instruction required to apply a patch and also the instruction required to perform a patch rollback.

The OPlan utility can be downloaded from My Oracle Support using patch number 11846294. OPlan supports all Exadata Database Machine recommended Bundle Patches commencing with release 11.2.0.2 Bundle Patch 2. See My Oracle Support bulletin 1306814.1 for further information on OPlan.

Maintaining Other Software

- Other components in Database Machine that have software or firmware:
  - InfiniBand switches
  - Power Distribution Units (PDUs)
  - KVM switch
  - Cisco Ethernet switch
- Always refer to My Oracle Support bulletin 888828.1 for Exadata-specific requirements

Other Database Machine components that have software or firmware include:

- InfiniBand switches
- Power Distribution Units (PDUs)
- Keyboard, Video and Mouse (KVM) switch
- Cisco Ethernet switch

The software and firmware on all components must always be maintained in accordance with the guidelines published in My Oracle Support bulletin 888828.1. In particular this is critical for the InfiniBand switches. For other components the requirements may be less specific.

Recommended Patching Process

1. Review the patch documentation (README file)
   - Read and understand it all before proceeding
2. Validate the patch installation on a test system
   - Run exachk before and after patch application
   - Automate the patch application steps where possible
   - Test the failback procedure
3. Validate the patch functionality on a test system
   - Verify that the patch provides the desired functionality
   - Evaluate system performance
4. Apply the patch in production
   - Run exachk before and after patch application
   - Make sure all cells are healthy before applying the patch
   - Evaluate system performance

No two patches are the same. This adage is true whether or not you are using Database Machine. Hence you should always follow a methodical patching process. Following is an outline of the recommended patching process for Exadata Database Machine:

1. Review the patch README file for known issues, patch installation and deinstallation instructions, and special notes.
2. Validate the patch installation on a proper test system:
   - Run exachk to verify the environment before and after the patch application.
   - Where possible, automate the steps to reduce human error.
   - Test the fallback procedures in case the patch must be rolled back.
3. Validate patch functionality on the test system.
   - If applying the patch to address a specific issue, verify that the patch provides the desired functionality.
   - Verify that there are no performance, availability, or operational regressions. Test using a workload that is representative of the production workload. Compare metrics observed during the test with baseline metrics observed in production. Real Application Testing may be used to replay a production workload on the test system. Automatic Workload Repository (AWR) and SQL Performance Analyzer may be used to assess performance on the test system.

4. Apply the patch on the production system.
   - Run exachk to verify the environment before and after the patch application.
   - Make sure that all the Exadata Storage Servers are functioning correctly. This is especially important because an unnoticed storage server fault that doesn't seem to impact Database Machine operations may result in diskgroups being offlined and databases terminating abruptly when another storage server taken offline for patching.
   - Compare metrics captured before and after patch application to evaluate system performance. SQL Performance Analyzer may be used to assess performance improvement or regression resulting from the patch.

See the lesson entitled *Other Useful Monitoring Tools* and My Oracle Support note 1070954.1 for details regarding exachk.

Test System Recommendations

- An effective test system:
  - Is an exact replica of the production Database Machine
  - Is not used for any other purpose during testing
  - Contains a full copy of the production data set with identical statistics
  - Is able to mimic production transaction volumes and concurrency
  - Is able to compare workload and performance metrics on the test system with metrics collected from the production system
  - Includes the operational and availability test suites used in production

The following list describes the recommendations for an effective Database Machine test environment:

- A test system should be an exact replica of the primary Exadata Database Machine including any standby systems for high availability and disaster recovery.
- When the test system is shared for other purposes, such as development, exclusive time should be allocated to perform patch validation.
- The test system should contain a full copy of the production data set with identical statistics.
- A workload framework, such as Real Application Testing, should be used to mimic production transaction volumes and concurrency.
- AWR data should be collected from the production database to compare against AWR data collected on the test system.
- Any operational and availability test suites used in production should also be ported to the test environment.

Quiz

Which of the following statements is a broad overview of the recommended approach for patching and updating Database Machine?

a. All updates for Database Machine are specific to Database Machine, and these are the only updates that should be applied to Database Machine

b. Use Unbreakable Linux Network (ULN) to update Linux on database servers and Exadata servers, and apply specific database and Exadata updates for the remaining software

c. Use Exadata-specific update bundles for the Exadata servers, and use normal database updating practices for the database servers

Answer: c

Summary

In this lesson, you should have learned how to:

- Describe how software in maintained on different Database Machine components
- Locate recommended patches for Database Machine
- Describe the recommended patching process for Database Machine
- Describe the characteristics of an effective test system

Additional Resources

- Lesson Demonstrations
  - Exadata Storage Server Rolling Patch Application
  - Exadata Storage Server Patch Rollback
- My Oracle Support Notes
  - Database Machine and Exadata Storage Server 11g Release 2 (11.2) Supported Versions
  - Oracle Software Patching with OPLAN
  - Oracle Exadata Database Machine exachk or HealthCheck

23

Exadata Database Machine Automated Support Ecosystem

ORACLE

Objectives

After completing this lesson, you should be able to:

- Describe the Auto Service Request (ASR) function and how it relates to Exadata Database Machine
- Describe the implementation requirements for ASR
- Describe the ASR configuration process
- Describe Oracle Configuration Manager (OCM) and how it relates to Exadata Database Machine

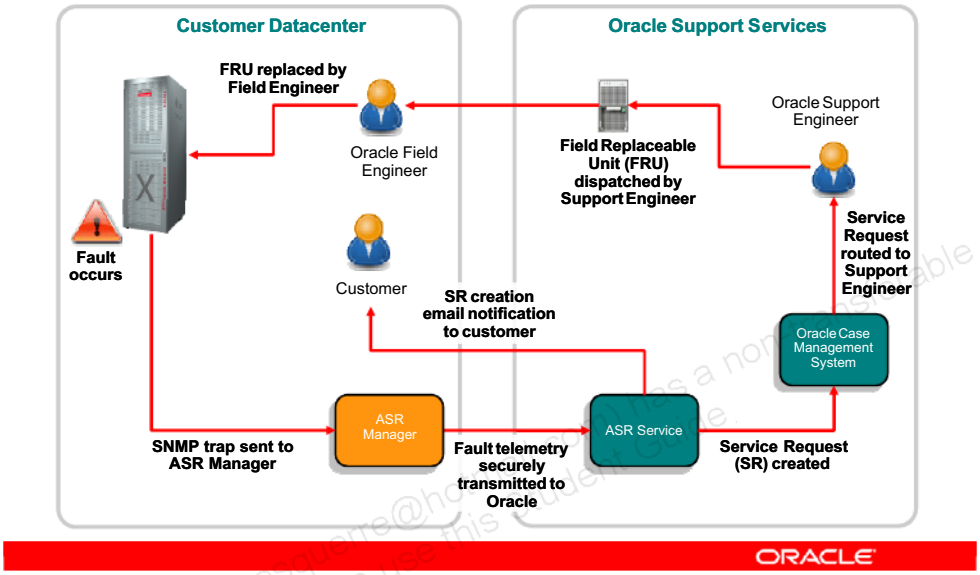
Auto Service Request Overview

- Automatically and securely opens service requests with Oracle for common server hardware faults
  - Minimal data is collected and transmitted
  - No IP addresses are communicated
  - One-way (customer to Oracle) SSL encrypted communications are used
- Enables fast and accurate resolution of hardware faults
  - Improved availability, less downtime
- Can be integrated with existing monitoring tools
  - ASR manager can send SR notifications via SNMP traps to existing monitoring tools
- Included with hardware warranty and Oracle Premier Support for Systems

Auto Service Request (ASR) is a secure, scalable, customer-installable software solution available as a feature of your Oracle or Sun hardware warranty, and Oracle Premier Support for Systems or valid Sun support plan. The ASR software helps to resolve problems faster by using auto-case generation for Oracle's Sun server and storage systems when specific hardware faults occur.

With the release of the Exadata software version 11.2.1.3.1, ASR functionality is extended to Exadata Database Machine. ASR automatically opens service requests (SRs) with Oracle Support when specific hardware faults occur either in the Exadata Storage Servers or the database servers. Note that ASR is currently applicable only for hardware faults detected on the following server components: CPUs, disk controllers, disks, flash cards, flash modules, InfiniBand cards, memory modules, system boards, power supplied, and fans.

The ASR Process



The diagram in the slide illustrates the ASR process. It shows how a Service Requests (SR) is automatically opened by the ASR Manager after it receives an SNMP trap that is triggered by a server hardware fault. Within Oracle Support Services the automatic service request is acknowledged via email and serviced according to normal procedures.

Customers should note that there are occasions when a SR may not be automatically filed. This can happen due to loss of connectivity to the ASR manager, for example. Oracle recommends that customers continue to monitor their systems for faults and engage Oracle Support if they notice a fault but do not receive notice that a service request has been automatically filed.

ASR Requirements

- A server capable of running ASR Manager:
  - An existing ASR Manager (version 2.7 or greater) can be used to monitor Database Machine
  - A Database Machine database server can be used as the ASR Manager (not recommended)
- Connectivity between the ASR manager and the Database Machine management network
- HTTPS connectivity (either directly or via a proxy) from the ASR manager to Oracle Support
- My Oracle Support account with current contact information for all Database Machine assets
- Sun Online Account (SOA)
- Exadata software 11.2.1.3.1 or later on the database servers and Exadata Storage Servers

The slide lists the requirements for using ASR in conjunction with Database Machine.

See the Hardware and Network Configuration Recommendations for ASR at http://www.oracle.com/technetwork/server-storage/asr/overview/hardware-recommendations-330108.html for more details regarding the server requirements for the ASR Manager.

When checking for the correct Exadata software on the Database Machine database servers, check for the existence of the following file:

```
/opt/oracle.cellos/compmon/exadata_mon_hw_asr.pl
```

If the file is missing, then update the database servers using a current Bundle Patch.

By default, configuration of ASR is performed by Oracle engineers as part of the initial installation and configuration of Exadata Database Machine. Customers can opt-out of configuring ASR however this is not recommended by Oracle. The next section in this lesson outlines the ASR configuration process. This process can be used by existing Exadata customers to configure ASR after Exadata Database Machine is already installed, or it can be used simply as a quick reference so that customers can better understand the ASR configuration process.

Configuring ASR Manager

1. Install ASR Manager components:
   - Oracle Service Tools Bundle (STB)
   - Service Tags
   - SAMS Package
   - ASR Package
2. Register ASR Manager:
   - As `root` on the ASM Manager server, run:

```
# asr register
```

     - Follow the prompts and provide requested information
   - To check registration status, run:

```
# asr show_reg_status
```

   - To test the network connectivity, run:

```
# asr test_connection
```


Configuring a new ASR Manager server consists of two main tasks:

1. Installing the ASR Manager software components
2. Registering the ASR Manager

The procedure for installing the ASR Manager software components varies depending on the ASR Manager version and the platform selected to run the ASR Manager. Consult the ASR documentation at http://www.oracle.com/technetwork/server-storage/asr/documentation/index.html for a detailed installation guide. A quick installation guide entitled *Oracle Exadata Quick Installation Guide for ASR* can also be found at the same location.

Configuring Exadata Database Machine for ASR

1. Configure SNMP trap destinations:
   - As `root` on each database server, run:

```
# /opt/oracle.cellos/compmon/exadata_mon_hw_asr.pl -set_snmp_subscribers \
>  "(type=asr,host=<ASR Manager>,port=162,community=public,)"
```

   - As `root` or `celladmin` on each Exadata Storage Server, run:

```
CellCLI> alter cell snmpsubscriber = -
>  ((host='<ASR Manager>', port=162, community=public, type=ASR))
```

2. Verify ASR SNMP subscribers:
   - As `root` on each database server, run:

```
# /opt/oracle.cellos/compmon/exadata_mon_hw_asr.pl -get_snmp_subscribers \
>  -type asr
```

   - As `root` or `celladmin` on each Exadata Storage Server, run:

```
CellCLI> list cell attributes snmpsubscriber
```


Each Exadata Database Machine database server and Exadata Storage Server must be configured to deliver SNMP traps to the ASR Manager server. Use the commands shown in the slide to configure and verify the required SNMP trap destinations.

In the commands, any reference to `<ASR Manager>` should be replaced with the ASR Manager server hostname or IP address.

Remember that the SNMP subscriber configuration is required for every Database Machine database server and Exadata Storage Server. If you prefer, you can use the `dcli` utility to perform the configuration simultaneously across multiple servers. For example, the following command could be used to perform the SNMP subscriber configuration on all the storage servers listed in the `cell_group` file:

```
# dcli -g cell_group -l celladmin -n "cellcli -e \
> alter cell snmpsubscriber=\(\(host=<ASR Manager>, \
> port=162, community=public, type=ASR\)\)"
```

Activating ASR Assets

- As `root` on the ASR Manager server:
  - Activate the ASR Manager host:

```
# asr activate_asset -i <ASR Manager IP>
```

  - Activate each Database Machine database server ILOM and each Exadata Storage Server ILOM:

```
# asr activate_asset -i <Asset ILOM IP>
```

  - Activate each Database Machine database server and each Exadata Storage Server:

```
# asr activate_exadata -i <Asset IP> \
>  -h <Asset hostname> -l <Asset ILOM IP>
```


The slide shows the commands required to activate the ASR assets. All commands are performed as `root` on the ASR Manager server. Activating the ASR Manager host is a one-time operation, while the other commands must be executed once for each Database Machine database server and Exadata Storage Server.

In the commands, `<ASR Manager IP>` should be replaced with the ASR Manager server IP address, `<Asset ILOM IP>` should be replaced with the database server or storage server ILOM interface IP address, `<Asset IP>` should be replaced with the database server or storage server management network IP address, and `<Asset hostname>` should be replaced with the database server or storage server hostname.

Approve and Verify ASR Assets

- Approve Database Machine assets in My Oracle Support:
  1. Click the *More...* tab and then *Settings*
  2. Click *Pending ASR Activations* in the *Settings* left-hand section
  3. Select a Database Machine server from the list
  4. Review the *ASR Activation – Asset* window
  5. Click *Approve* to complete activation
  6. Repeat steps 3, 4 and 5 for until all the Database Machine assets are activated
- Verify the ASR configuration:
  - As `root` on each database server, run:

```
# /opt/oracle.cellos/compmon/exadata_mon_hw_asr.pl -validate_snmp_subscriber \
>  -type asr
```

  - As `root` or `celladmin` on each Exadata Storage Server, run:

```
CellCLI> alter cell validate snmp type=ASR
```

ORACLE

After the Database Machine server assets are activated in the ASR Manager, the servers are listed inside My Oracle Support as Pending ASR Activations. To complete the activation process the servers must be approved in My Oracle Support using the following process:

1. On the My Oracle Support home page, click the *More...* tab and then *Settings.*
2. Click *Pending ASR Activations* in the *Settings* panel on the left side of the window .
3. Select a Database Machine server from the list.
4. Review the *ASR Activation – Asset* window. Update information relating to your Database Machine environment as required.
5. Click *Approve* to complete activation.
6. Repeat steps 3, 4 and 5 for until all the Dat abase Machine assets are activated.

After the assets are approved, the ASR configuration is completed and the system is enabled. To test the configuration and verify that it functions correctly, execute the commands shown on the slide. Each time a validation command is executed, the end-to-end ASR process is tested and a message is sent to the Sun Online Account (SOA) email account indicating that the ASR service inside Oracle Support received the test message. No Service Request is opened for validation requests.

Oracle Configuration Manager Overview

Oracle Configuration Manager collects configuration information for propagation to Oracle Support

- Benefits include:
  - More efficient problem diagnosis
  - Easier service request reporting
- Configuration information can be uploaded to Oracle either automatically or manually
- Configuration information remains confidential
- Oracle recommends the use of OCM in conjunction with Exadata Database Machine

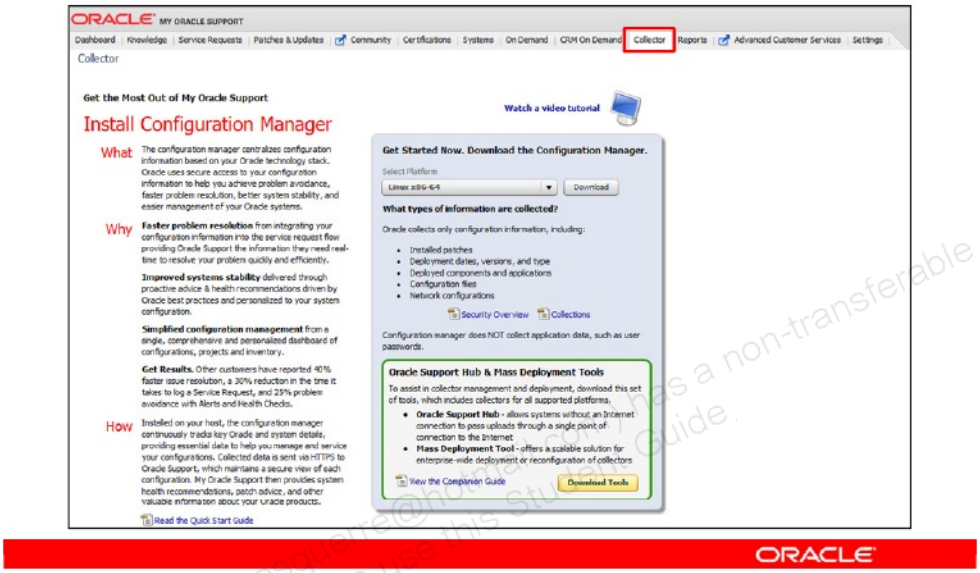
Oracle Configuration Manager (OCM) automatically collects configuration information from your environment at regular intervals. This configuration information can be uploaded to My Oracle Support. This helps Oracle to maintain up-to-date information about your environment, diagnose support issues more efficiently, and offer consistently better support outcomes. OCM is recommended for use in conjunction with Exadata Database Machine.

Oracle recommends that OCM is set up so that the configuration information can be automatically collected and uploaded to the Oracle Customer Configuration Repository and My Oracle Support. Alternatively, information can be manually uploaded by an administrator.

No business or personal information is collected and uploaded, except for contact information which is used in the event of transmission problems. All information is kept strictly confidential.

Note that OCM is not a function of Grid Control and is separate from the Grid Control Configuration Management Pack.

Configuring Oracle Configuration Manager



Like ASR, configuration of OCM is performed by Oracle engineers as part of the initial installation and configuration of Exadata Database Machine. Customers can opt-out of configuring OCM however this is not recommended by Oracle.

For customers who wish to configure OCM against an existing Database Machine, the easiest way is to use the resources under the *Collector* tab in My Oracle Support. The slide show the interface presented to uses which enables them to get started with OCM.

Note that the OCM collector is installed on the Database Machine database servers. No OCM components are installed on the Exadata Storage Servers. Therefore select *Linux x86-64* or *Oracle Solaris on x86-64 (64-bit)* from the *Select Platform* drop-down list depending on how your Database Machine is configured.

Quiz

Oracle recommends that the Auto Service Request Manager is configured on one of the Exadata Database Machine servers:

a. True
b. False

Answer: b

The ASR Manager can be configured on an Exadata database server, but this is not recommended.

Quiz

ASR will create an service request with Oracle Support when which of the following occur?

a. A disk fault in a storage server
b. An ORA-600 error in a database server
c. A power supply in a server fails
d. A hardware fault in a server fan
e. All of the above

Answer: a, c, d

Quiz

The Grid Control Configuration Management Pack is required to use Oracle Configuration Manager:

a. True
b. False

Answer: b

Oracle Configuration Manager is not a function of Grid Control and is separate from the Grid Control Configuration Management Pack.

Summary

In this lesson, you should have learned how to:

- Describe the Auto Service Request (ASR) function and how it relates to Exadata Database Machine
- Describe the implementation requirements for ASR
- Describe the ASR configuration process
- Describe Oracle Configuration Manager (OCM) and how it relates to Exadata Database Machine

Additional Resources

- Documentation and recommendations
  - Oracle Auto Service Request User Documentation and Product Qualification Information
  - Hardware and Network Configuration Recommendations for ASR

24

Quality of Service Management

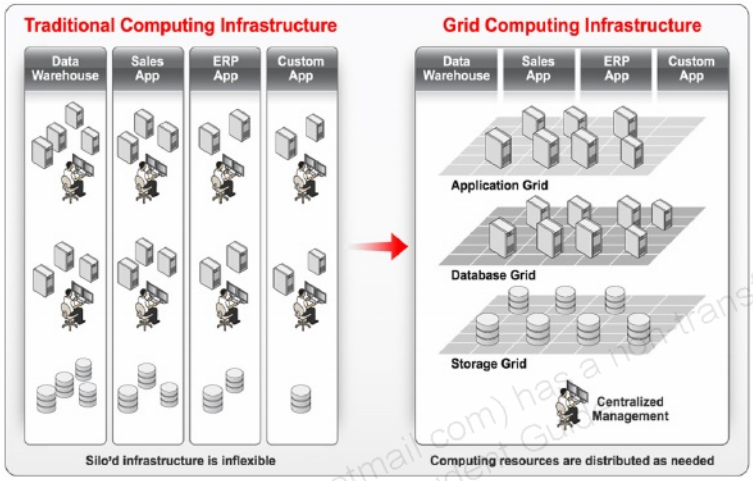
ORACLE

Lesson Objectives

After completing this lesson, you should be able to:

- Describe the purpose of Oracle Database Quality of Service (QoS) Management
- Describe the benefits of using Oracle Database QoS Management
- Describe the components of Oracle Database QoS Management
- Describe the operation of Oracle Database QoS Management

QoS Management Background

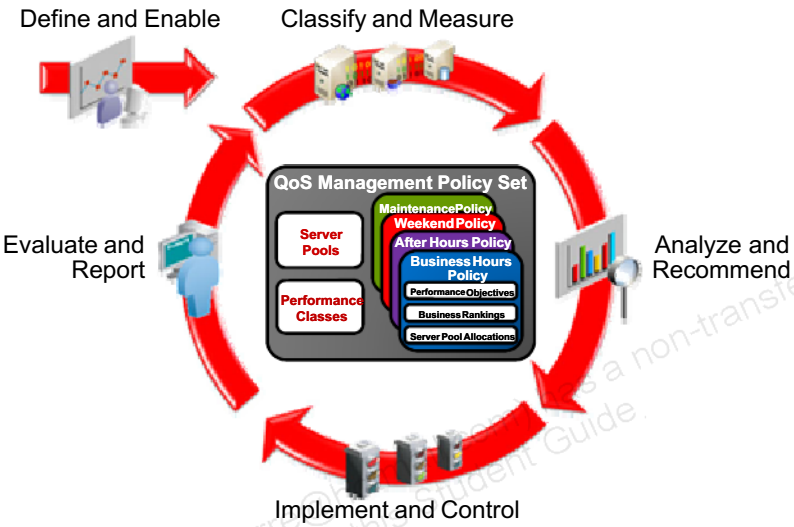


In previous Oracle Database releases, you could use services for workload management and isolation. For example, a group of servers might be dedicated to data warehouse work, while another is dedicated to your sales application, a third group is used for ERP processing, and a fourth group to a custom application. Using services, the database administrator can allocate resources to specific workloads by manually changing the number of servers on which a database service is allowed to run. The workloads are isolated from each other, so that demand spikes, failures and other problems in one workload do not affect the other workloads. The problem with this type of deployment is that each workload needs to be separately provisioned for peak demand because resources are not shared.

You could also define services which shared resources by overlapping server allocations. However even with this capability, you had to manually mange the server allocations and each service was mapped to a fixed group of servers.

Starting with Oracle Database 11*g*, you can use server pools to logically partition a cluster and provide workload isolation. Server pools provide a more dynamic and business focused way of allocating resources because resource allocations are not dependant on which servers are up. Rather, the server pool allocations dynamically adjust when servers enter and leave the cluster to best meet the priorities defined in the server pool policy definitions.

QoS Management Overview



Many companies are consolidating and standardizing their data center computer systems. In parallel with this, the migration of applications to the Internet has introduced the problem of managing demand surges that cannot be fully anticipated . In this type of environment, it is necessary to pool resources and have management tools that can detect and resolve bottlenecks in real time. Policy-managed server pools provide a foundation for dynamic workload management however they can only adjust resource allocations in response to server availability changes.

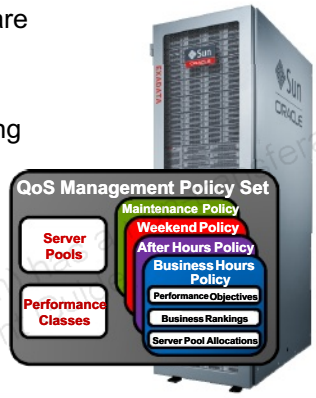
QoS Management is an automated, policy-based workload management (WLM) system that monitors and adjusts the environmentto meet businesslevel performance objectives. Based on resource availability *and* workload demands, QoS Management identifies resource bottlenecks and provides recommendations for how to relieve them. It can make recommendations for the system administrator to move a server from one server pool to another, or to adjust access to CPU resources using the Database Resource Manager, in order to satisfy the current performance objectives.

Using QoS Management enables the administrator to ensure the following:

- When sufficient resources are available to meet the demand, business level performance objectives for each workload are met, even if the workloads change.
- When sufficient resources are not available to meet all demands, QoS Management attempts to satisfy more critical business objectives at the expense of less critical ones.

QoS Management and Exadata Database Machine

- In its initial form, QoS Management is a feature of the Oracle Database product family:
  - Introduced in Oracle Database 11*g* release 2
  - Associated with Oracle RAC software
  - Released exclusively on Exadata Database Machine
  - Focused on environments supporting multiple OLTP workloads
  - Not Exadata-specific technology
  - The first step along the road towards a broader solution



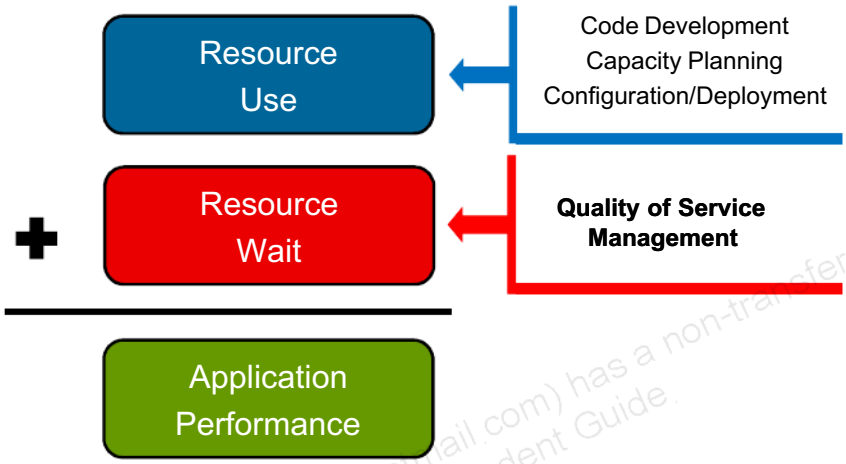
The initial incarnation of QoS Management is as a feature of the Oracle Database product family in association with Oracle Real Application Clusters (RAC) software. It was first introduced in Oracle Database 11 *g* release 2.

The initial set of features and benefits associated with QoS Management are exclusively available to Exadata Database Machine customers and are best suited to customers using Database Machine predominantly as a consolidation platform for OLTP applications.

QoS Management software can operate on non-Exadata environments where Oracle Database 11*g* release 2 is available. Commencing with version 11.2.0.3, a limited subset of QoS Management functionality has been released which enables non-Exadata users to monitor performance classes, but not generate and implement changes in response to the currently observed workload.

In its current form, QoS Management provides a powerful database-focused capability which represents the first step along the road towards a broader workload management solution.

QoS Management Focus



QoS Management monitors the performance of each work request on a target system. By accurately measuring the two components of performance, resource use and wait, bottlenecks can be quickly detected and resources reallocated to relieve them, thus preserving or restoring service levels. Changing or improving the execution time generally requires application source code changes. QoS Management therefore only observes and manages wait times.

QoS Management bases its decisions on observations of how long work requests spend waiting for resources. Examples of resources that work requests might wait for include hardware resources, such as CPU cycles, disk I/O queues, and global cache blocks.

Other waits can occur within the database, such as latches, locks, pins, and so on. While these database waits are accounted for by QoS Management, they are not broken down by type or managed. Minimizing unmanaged waits requires changes that QoS Management cannot perform, such as application code changes and database schema optimizations for example. QoS Management is still beneficial in these cases because the measurement and notification of unmanaged waits can be used as a tool to measure the effect of application optimization activities.

QoS Management Benefits

- Determines where additional resources are needed
- Determines if additional hardware can be added to maintain acceptable performance
- Reduces number of critical performance outages
- Reduces time to resolve performance objective violations
- Improves system stability as the workload changes
- Helps to ensure that SLAs are met
- Facilitates effective sharing of hardware resources

Some of the benefits of QoS Management include:

- By categorizing and measuring database work, QoS Management can help administrators determine where additional resources are needed.
- QoS Management is Oracle RAC-aware, and it uses this fundamental understanding to determine if additional hardware can be added to maintain acceptable performance.
- QoS Management helps reduce the number of critcal performance outages. By reallocating runtime resources to the busiest business-critical applications, those applications are less likely to suffer from a performance outage.
- QoS Management reduces the time needed to resolve performance objective violations. Rather than requiring administrators to understand and respond to changes in performance, much of the work can be automated. Administrators are provided with a simple interface to review and implement the recommended changes.
- Performance stressescan often leadto system instability. By moving resourcesto where they are most needed, QoS Management reduces the chance that systems will suffer from performance stress and related instability

- QoS Management allows the administrator to define performance objectives that help to ensure Service Level Agreements (SLAs) are being met. Once the objectives are defined, QoS Management tracks performance and recommends changes if the SLAs are not being met.
- As resource needs change, QoS Management can reallocate hardware resources to ensure applications make more effective usedof those resources. Resources can be removed from applications that no longer require them, and added to an application that is suffering from performance stress.

QoS Management Functional Overview

QoS Management works with Oracle RAC and Oracle Clusterware to:

- Manage database server CPU resources by evaluating CPU wait times to identify workloads that are not meeting performance objectives
  - QoS Management can recommend:
    - Adjustments to the size of server pools
    - Alterations to consumer group mappings
    - Adjustments to the CPU resources allocated to different database instances within a server pool
- Manage memory pressure due to number of sessions or runaway workloads
  - QoS Management restricts new sessions from being established on servers that are suffering from memory stress

QoS Management works with Oracle RAC, Oracle Clusterware and Cluster Health Monitor (CHM) to manage database resources to meet service levels, and manage memory pressure for managed servers.

Typically, database services are used to group related work requests and for measuring and managing database work. For example, a user initiated query against the database might use a different service from a report generation application. To manage the resources used by a service, some services may be deployed on several Oracle RAC instances concurrently, while others may be deployed on only one instance.

In an Oracle RAC database, QoS Management monitors the nodes on which user-defined database services are offered. Services are created in a specific server pool and the service runs on all servers in the server pool. If a singleton service is required because the application cannot effectively scale across multiple RAC servers, the service can be hosted in a server pool with a maximum size of one.

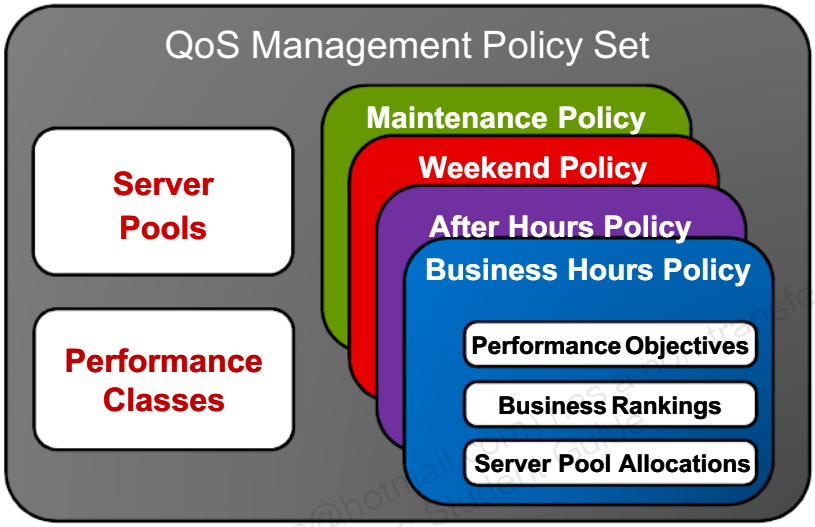
QoS Management periodically evaluates database server CPU wait times to identify workloads that are not meeting performance objectives. If needed, QoS Management provides recommendations for adjusting the size of the server pools or alterations to Database Resource Manager (DBRM) consumer group mappings. Starting with Oracle Database release 11.2.0.3, QoS Management also supports moving CPUs between databases within the same server pool.

The DBRM is an example of a resource allocation mechanism; it can allocate CPU shares among a collection of resource consumer groups based on a resource plan specified by an administrator. A resource plan allocates the percentage of opportunities to run on the CPU. QoS Management does not adjust DBRM plans; it activates a shared multi-level resource plan and then, when implementing a recommendation, it moves workloads to specific resource consumer group to meet performance objectives for all the different workloads.

Enterprise database servers can run out of available memory due to too many open sessions or runaway workloads. Running out of memory can result in failed transactions or, in extreme cases, a reboot of the server and loss of a valuable resource for your applications. QoS Management eases memory pressure by temporarily shutting down the services for database instances on a server suffering from memory stress. This causes new sessions to be directed to lighter loaded servers. Rerouting new sessions protects the existing workloads and the availability of the memory-stressed server.

When QoS Management is enabled and managing an Oracle Clusterware server pool, it receives a metrics stream from Cluster Health Monitor that provides real-time information about memory resources for a server, including the amount of available memory, the amount of memory currently in use, and the amount of memory swapped to disk for each server. If QoS Management determines that a node is under memory stress, the Oracle Clusterware managed database services are stopped on that node preventing new connections from being created. After the memory stress is relieved, the services are restarted automatically and the listener can send new connections to the server. The memory pressure can be relieved in several ways, for example, by existing sessions closing or by user intervention.

QoS Management Policy Sets



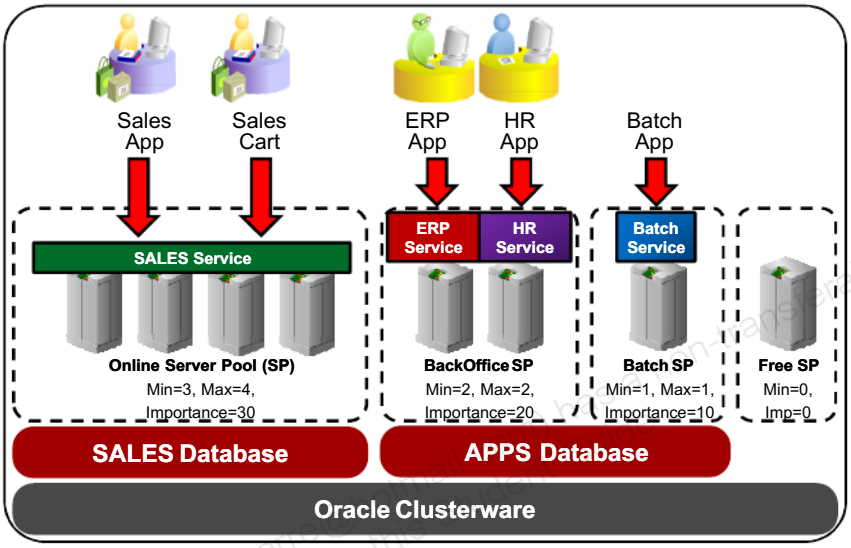
A central concept in QoS Management is the policy set. A policy set allows you to specify your resources, performance classes (workloads), and a collection of performance policies that specify the performance objective for each performance class and sets constraints for resource availability. QoS Management uses a system-wide policy set that defines performance objectives based upon the classes of work and the availability of resources. Specific performance policies can be enabled based upon a calendar schedule, maintenance windows, events, and so on. Only one performance policy can be in effect at any time.

To maintain the current performance objectives, QoS Management makes resource reallocation recommendations and predicts their effect. The recommendations can be easily implemented with a single button click.

A policy set consists of the following:

- The server pools that are being managed by QoS Management.
- Performance classes, which are work requests with similar performance objectives.
- Performance policies, which describe how resources should be allocated to the performance classes by using performance objectives and server pool directive overrides. Within a performance policy, performance objectives are ranked based on business importance which enables QoS Management to focus on specific objectives when the policy is active.

Server Pools



A server pool is a logical division of a cluster. Server pools facilitate workload isolation within a cluster while maintaining agility and allowing users to derive other benefits associated with consolidation. Administrators can define server pools which are typically associated with different applications and workloads. An example is illustrated on the slide. QoS Management can assist in by managing the size of each server pool and also managing the allocation of resources within a server pool.

When Oracle Grid Infrastructure is first installed, a default server pool, called the Free pool, is created. All servers are initially placed in this server pool. Specific server pools can then be created for each of the workloads that needs to be managed. When a new server pool is created, the servers assigned to that server pool are automatically moved out of the Free pool and placed into the newly created server pool.

After a server pool is created, a database can be configured to run on the server pool, and cluster-managed services can be established for applications to connect to the database.

For an Oracle RAC database to take advantage of the flexibility of server pools, the database must be created using the policy-managed deployment option, which places the database in one or more server pools.

A key attribute of policy-based management is the allocation of resources to server pools based on cardinality and importance.

When the cluster starts or when servers are added, all the server pools are filled to their minimum levels in order of importance. After the minimums are met, server pools continue to be filled to their maximums in order of importance. If there are any left-over servers, they are allocated to the Free pool.

If servers leave the cluster for any reason, a server reallocation may take place. If there are servers in the Free pool and another server pool falls below its maximum value, a free server is allocated to affected server pool. If there are no free servers, then server reallocation takes place only if a server pool falls below its minimum level. If that occurs, a server will be sourced from one of the following locations in the following order:

1. The server pool with the lowest importance that has more than its minimum number of servers.
2. The server pool with the lowest importance that has at least one server and has lower importance than the affected server pool.

Using these mechanisms, server pools can maintain an optimal level of resources based on the current number of servers that are available.

Consider the example shown on the slide. If one of the servers in the Online server pool failed, the server currently residing in the Free server pool would automatically move to the Online server pool.

Now, if one of the servers from the BackOffice server pool failed, there would be no servers to allocate from the Free server pool. In this case, the server currently servicing the Batch server pool would be dynamically reallocated to the BackOffice server pool because the failure would cause the BackOffice server pool to fall below its minimum and it has a higher importance than Batch.

If one node is later returned to the cluster, it will be allocated to the Batch pool in order to satisfy the minimum for that server pool.

Any additional nodes added to the cluster after this point will be added to the Free pool since all the other pools are filled to their maximum level.

Performance Classes

- A performance class is a group of work requests whose service level needs to be managed
- Work requests are defined by performance classifiers containing the database service name and optional session parameters
- Initial set of performance classifiers are automatically discovered and created from cluster-managed services
- Performance objectives are defined on performance classes

Performance classes are used to categorize workloads with similar performance requirements. A set of classification rules are evaluated against work requests when they arrive at the edge of the system. These rules allow value matching against attributes of the work request; when there is a match between the type of work request and the criteria for inclusion in a performance class, the work request is classified into that performance class.

This classification of work requests applies the user-defined name, or tag, that identifies the performance class (PC) to which the work request belongs. All work requests that are grouped into a particular PC have the same performance objectives. In effect, the tag connects the work request to the performance objective that applies to it. Tags are carried along with each work request so that every component of the system can take measurements and provide data to QoS Management for evaluation against the applicable performance objectives.

QoS Management supports user-defined combinations of connection parameters called classifiers to map performance classes to the actual workloads running in the database.

These connection parameters fall into two general classes and can be combined to create fine-grained Boolean expressions:

- **Configuration Parameters:** The supported configuration parameters are SERVICE_NAME and USERNAME. Each classifier in a performance class must include one or more cluster-managed database services. Additional granularity can be achieved by identifying the Oracle Database user that is making the connection from either a client or the middle tier. The advantage of using these classifiers is that they do not require application code changes to define performance classes.
- **Application Parameters:** The supported application parameters are MODULE, ACTION, and PROGRAM. These are optional parameters set by the application as follows:
  - **OCI:** Use OCI_ATTR_MODULE and OCI_ATTR_ACTION
  - **ODP.NET:** Specify the ModuleName and ActionName properties on the OracleConnection object
  - **JDBC:** Set MODULE and ACTION in SYS_CONTEXT

The PROGRAM parameter is set or derived differently for each database driver and platform. Please consult the appropriate Oracle Database developer's guide for further details and examples.

To manage the workload for an application, the application code directs database connections to a particular service. The service name is specified in a classifier, so all work requests that use that service are tagged as belonging to the performance class created for that application. If you want to provide more precise control over the workload generated by various part of the application, you can create additional performance classes and use classifiers that include MODULE, ACTION, or PROGRAM in addition to the SERVICE_NAME or USERNAME.

The performance classes used in an environment can change over time. A common scenario is to replace a single performance objective with multiple, more-specific performance objectives, dividing the work requests into additional performance classes. For example, application developers can suggest performance classes for QoS Management to use. In particular, an application developer can define a collection of database classifiers using the MODULE and ACTION parameters, and then put them in separate performance classes so each type of work request is managed separately.

Classification and Tagging

- Each session is classified:
  - The classification is determined by evaluating session parameters against the performance class classifiers
  - Occurs only when a session is established or when session parameters change
    - This minimizes the overhead associated with classification
- Each work request is tagged:
  - The tag is based on the current session classification
  - The tag connects the work request with a performance class
    - It enables measurements associated with the work request to be recorded against the appropriate performance class

To enable QoS Management, work requests must be classified and tagged.

When a database session is established, the session parameters are evaluated against the performance class classifiers to determine a classification. Work associated with the session is then tagged based on the session classification until the session ends or the session parameters change. If the session parameters change, the classification is re-evaluated. Thus the overhead associated with classification is very small since the classification is only evaluated when a session is established or when session parameters change.

Tags are permanently assigned to each work request so that all the measurements associated with the work request can be recorded against the appropriate performance class. In effect, the tag connects the work request to a performance class and its associated performance objective.

Performance Policies

- Performance policies are named sets of performance objectives and server pool overrides to meet business objectives
  - Performance objectives can be ranked according to their importance
- Only one policy is active at any time

To manage various performance objectives, a QoS Management administrator defines one or more performance policies. For example, the administrator might define a performance policy for normal business hours, another for weekday non-business hours, one for weekend operations, and another to be used during processing for the quarter-end financial closing. Note that at any time, only one performance policy is in effect.

A performance policy has a collection of performance objectives in effect; one or more for each application that is being managed on the system. Some performance objectives are always more critical to the business than others, while other performance objectives might be more critical at certain times, and less critical at other times. The ability to define multiple performance policies inside the policy set provides QoS Management with the flexibility required to implement different priority schemes when they are required.

Performance Class Ranks

- Performance class ranks assign a relative level of business criticality to each performance class within a performance policy:
  - Highest
  - High
  - Medium
  - Low
  - Lowest

Within a performance policy, you can also rank each performance class. This rank assigns a relative level of business criticality to each performance objective. When there are not enough resources available to meet all the performance objectives for all performance classes, the performance objectives for the more critical performance classes must be met at the expense of the less critical ones. The available rank settings are Highest, High, Medium, Low, or Lowest. Note that if more than one class is assigned a particular rank (for example, Medium), classes are then ordered within that ranking alphabetically.

Performance Objectives

- Performance objectives can be derived from your SLAs. They specify:
  - A business requirement, and
  - The performance class to which it applies
- Average response time per database call is currently the only performance objective type
  - Response time is the total time from the time the database receives the request to when the response leaves the server
  - Response time does not include network traffic time

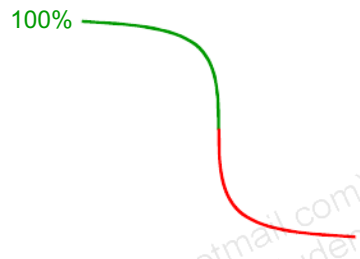
You create a performance objective for each performance class to specify the desired performance level for that performance class. A performance objective specifies both a business requirement, and the work to which it applies (the performance class). For example, a performance objective might say that database work requests that use the SALES service should have an average response time of less than 60 milliseconds.

Each performance policy includes a performance objective for each and every performance class, unless the performance class is marked measure-only. In this release, QoS supports only one type of performance objective, average response time.

Response time is based upon database client calls from the point that the database server receives the request over the network until the request leaves the server. Response time does not include the time it takes to send the information over the network to or from the client. The response time for all database client calls in a performance class is averaged and presented as the average response time.

Performance Satisfaction Metrics

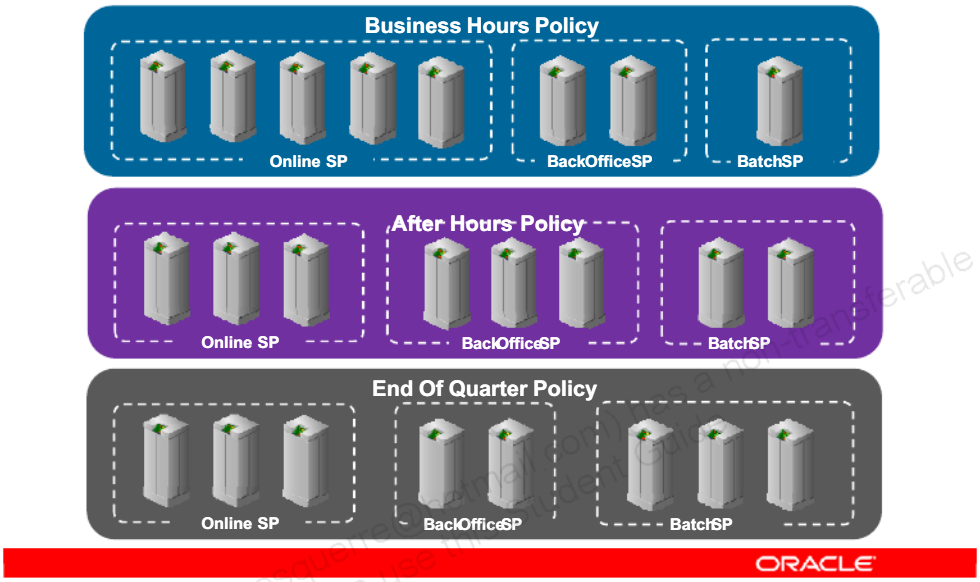
- Different performance objectives can be compared using the Performance Satisfaction Metric (PSM)
- The PSM quickly shows how the system is coping with the objective



Different performance objectives are used to measure the performance of different workloads. QoS Management currently supports only OLTP workloads and uses only the average response time performance objective. When configuring QoS Management, you can have very different performance objectives for each performance class. For example, one performance objective may specify that a Checkout call should complete within 1 millisecond, while another performance objective may specify that a Browse call should complete within 1 second. As more performance objectives are added to a system, it can be difficult to compare them quickly.

Because of this, it is useful to have a common and consistent numeric measure indicating how the current workload for a performance class is measuring up against its current performance objective. This numeric measure is called the Performance Satisfaction Metric. The Performance Satisfaction Metric is thus a normalized numeric value (between +100% and -100%) that indicates how well a particular performance objective is being met, and which allows QoS Management to compare the performance of the system for widely differing performance objectives.

Server Pool Directive Overrides



A performance policy can also include a set of server pool directive overrides. A server pool directive override sets the minimum server count, maximum server count and importance attributes for a server pool when the performance policy is in effect. Server pool directive overrides serve as constraints on the recommendations proposed by QoS Management because the server pool directive overrides are honored while the performance policy is active. For example, QoS Management will never recommend moving a server out of a server pool if doing so will leave the server pool below its minimum server count value.

Server pool directive overrides can be used to define the normal state of server pools at different points in time. The slide illustrates an example. Under normal conditions, these server pool settings would be expected to handle the prevailing workload. If there is a sudden increase in the workload requests for a performance class, then the associated server pool might require additional resources beyond what is specified in the performance policy.

Overview of Metrics

- QoS Management uses a standardized set of metrics
- There are two metric types:
  - Performance metrics give an overview of where time is spent in the system
  - Resource metrics measure the time that work requests use a resource or wait for a resource
- Metrics are used to identify bottlenecked resources and to determine the best corrective action:
  - For a performance class, the bottlenecked resource is the resource that contributes the largest average wait time

QoS Management uses a standardized set of metrics, which are collected by all the servers in the system. There are two types of metrics: performance metrics and resource metrics. These metrics enable direct observation of the use and wait time incurred by work requests in each performance class, for each resource requested, as it traverses the servers, networks and storage devices that form the system.

Performance metrics are collected at the entry point to each server in the system. They give an overview of where time is spent in the system and enables comparisons of wait times across the system. Data is collected periodically and forwarded to a central point for analysis, decision-making, and historical storage.

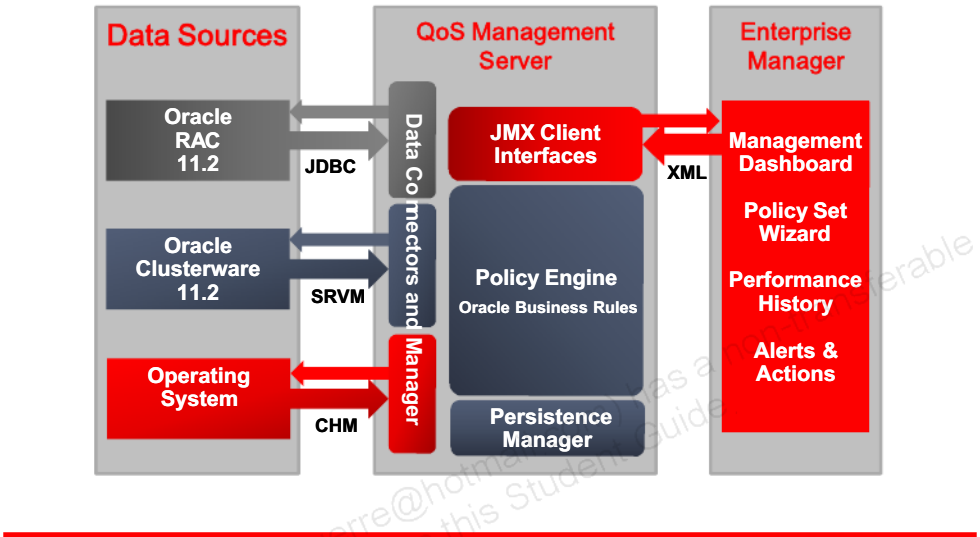
A key performance metric is response time, or the difference between the time a request comes in and the time a response is sent out. The response time for all database calls in a Performance Class is averaged and presented as the Average Response Time. Another important performance metric is the arrival rate of work requests. This provides a measure of the demand associated with each Performance Class.

Resource metrics exist for the following resources; CPU, Storage I/O, Global Cache, and Other (database waits). Two resource metrics are provided for each resource:

- Resource usage time - measures how much time is spent using the resource.
- Resource wait time - measures the time spent waiting to get the resource.

QoS Management metrics provide the information needed to systematically identify performance class bottlenecks in the system. When a performance class is violating its performance objective, the bottleneck for that performance class is the resource that contributes the largest average wait time for each work request in that performance class.

QoS Management Architecture



QoS Management retrieves metrics data from each database instance running in managed server pools and correlates the data by performance class every 5 seconds. The data includes many metrics; for example, call arrival rate and CPU, I/O and Global Cache use, and wait times. The data is combined with the current topology of the cluster and the health of the servers in the Policy Engine to determine the overall performance profile of the system with regard to the current performance objectives established by the active performance policy.

The performance evaluation occurs once a minute and results in a recommendation if there is a performance class not meeting its objective. The recommendation specifies what resource is bottlenecked. Specific corrective actions are included, if possible, along with the projected impact on all performance classes in the system. The slide shows the collection of data from various data sources by the data connectors component of QoS Management:

- Oracle RAC 11.2 communicates with the data connector using JDBC.
- Oracle Clusterware 11.2 communicates with the data connector using the SRVM component of Oracle Clusterware.
- The server operating system communicates with the data connector using Cluster Health Monitor (CHM).

Enterprise Manager displays the information in a variety of ways, including the Management Dashboard, Policy Set Wizard, Performance History, and Alerts and Actions screens.

QoS Management Recommendations

- If performance objectives are not being met, QoS Management makes a recommendation
- Each recommendation focuses on improving the highest ranked performance class exceeding its performance objective
- Recommendations may include:
  - Changing consumer group mappings
    - Reprioritize work within existing resource boundaries
  - Moving servers between server pools
    - Reprioritize resources between server pools to meet workload demands
  - Moving CPUs between databases within a server pool
    - Reprioritize CPU resources within existing server pool boundaries

If your business experiences periodic demand surges, then to retain performance levels for your applications you can acquire additional hardware to be available when needed, and sit idle when not needed. Rather than have extra servers sit idle for most of the time, you might decide to use those servers to run other application workloads. However, if the servers are busy running other applications when a demand surge hits, your main business applications are not able to perform as expected. QoS Management helps to manage such situations.

When you implement a performance policy, QoS Management continuously monitors the system and manages it using an iterative process. When one or more performance objectives are not being met, each iteration seeks to improve the performance of a single performance objective; the highest ranked performance objective that is currently not being met. When all performance objectives are being met, QoS Management makes no further recommendations.

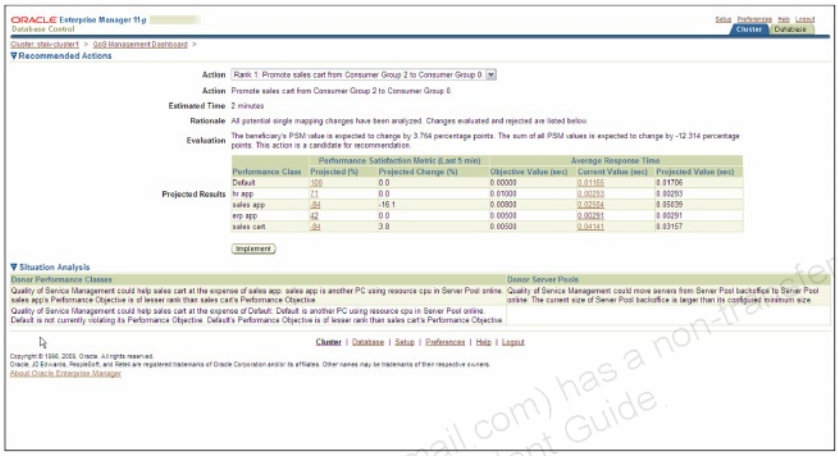
The recommendations take the form of moving servers between server pools, changing consumer group mappings, or moving CPUs between databases within a server pool.

Changing consumer group mappings may involve promoting a specific workload so that it gets a greater share of resources, or it may involve demoting a competing workload as a way of making additional resources available to the target performance class. In both cases, workloads are reprioritized within existing resource boundaries.

Moving servers between server pools is another approach used by QoS Management. This approach alters the distribution of servers to meet workload demands.

Commencing with Oracle Database release 11.2.0.3, QoS Management can also move CPU resources between databases within the same server pool. This alters the distribution of CPU resources between database instances using instance caging and provides additional control for environments where multiple databases are consolidated within the same Exadata Database Machine environment.

Implementing Recommendations



When QoS Management is working to improve the performance of a particular performance class, it recommends to add more of the bottleneck resource (such as CPU time) for that performance class, or to make the bottleneck resource available more quickly to work requests in the performance class.

Implementing a recommendation makes the resource less available to other performance classes. The negative impact on the performance classes from which the resource is taken may be significantly smaller than the positive impact on the service that is getting better access, resulting in a net win for the system as a whole. Alternatively, the performance class being penalized may be less business critical than the one being helped.

When generating recommendations, QoS Management evaluates the impact to the system performance as a whole. If the improvement for one performance class is rather small, but the negative impact on another performance class is large, then QoS Management might report that the performance gain is too small, and not recommended. If there is more than one way to resolve the bottleneck, then QoS Management advises the best overall recommendation factoring in variables such as the calculated impact on all the performance classes along with the predicted disruption and settling time associated with the action. Using Oracle Enterprise Manager, you can view the current recommendation and the alternative recommendations.

Performance data is sent to Oracle Enterprise Manager for display in the QoS Management Dashboard and Performance History pages. Alerts are generated to drive notifications that one or more performance objectives are not being met or that a problem has developed that prevents one or more server pools from being managed. As a result of these notifications the administrator can implement the recommendation.

In this release, QoS Management does not implement the recommendations automatically. It suggests a way of improving performance, which must then be implemented by the administrator clicking on the Implement button. After implementing a recommendation, the system is allowed to settle before any new recommendations are made. This is to ensure stable data is used for further evaluations and also to prevent recommendations that result in oscillating actions.

Quiz

Oracle Database Quality of Service Management helps to meet performance objectives by reducing resource usage:

a. True
b. False

Answer: b

Oracle Database of Service Management helps to meet performance objectives by managing and reducing resource wait times, not resource usage.

Quiz

Oracle Database Quality of Service Management recommendations can include:

a. Moving servers between server pools
b. Adding spindles to improve I/O performance
c. Changing consumer group mappings
d. Recommending partitioning strategies to ease global cache bottlenecks

Answer: a, c

Oracle Database Quality of Service Management can identify I/O performance issues or global cache bottlenecks, but cannot address them in this release.

Summary

In this lesson, you should have learned how to:

- Describe the purpose of Oracle Database Quality of Service (QoS) Management
- Describe the benefits of using Oracle Database QoS Management
- Describe the components of Oracle Database QoS Management
- Describe the operation of Oracle Database QoS Management

Additional Resources

- Lesson Demonstrations
  - Configuring Quality of Service Management
  - Using Quality of Service Management

Managing Exadata Database Machine with Enterprise Manager Cloud Control 12c

ORACLE

Objectives

After completing this lesson, you should be able to outline the key Exadata Database Machine management capabilities included in Enterprise Manager Cloud Control 12c.

Lesson Overview

- This lesson provides an outline of some key Exadata-specific enhancements in Enterprise Manager Cloud Control 12c.
- Areas covered in this lesson include:
  - Configuring Exadata Database Machine as an Enterprise Manager target
  - Visualizing Exadata Database Machine in Enterprise Manager
  - Exadata Storage Server management and performance monitoring
  - Exadata Database Machine InfiniBand network monitoring and administration

This lesson provides an outline of some of the Exadata-specific monitoring and administration enhancements included in Enterprise Manager Cloud Control 12c. The following areas are covered:

- Configuring Exadata Database Machine as an Enterprise Manager target.
- Visualizing Exadata Database Machine in Enterprise Manager.
- Exadata Storage Server management and performance monitoring.
- Exadata Database Machine InfiniBand network monitoring and administration.

The intention of this lesson is to provide an outline of some of the key areas which are enhanced in the new release of Enterprise Manager.

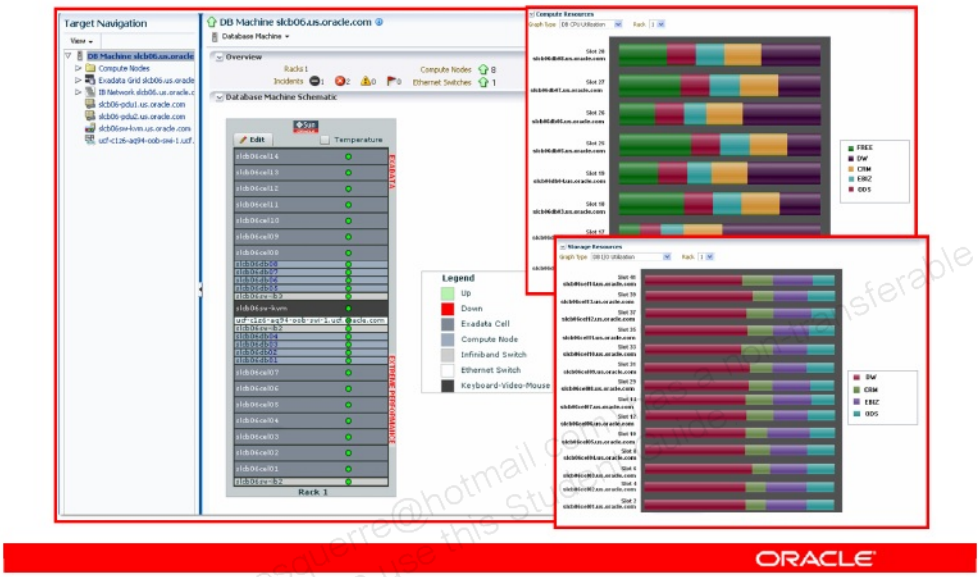
Configuring Exadata Database Machine as an Enterprise Manager Target

- New target type for Exadata Database Machine
- New wizard-driven guided discovery process:

The new Exadata Database Machine target type in Enterprise Manager Cloud Control 12c provides the basis for managing Database Machine as an integrated whole rather than as a series of related components. This fundamental enhancement to the Enterprise Manager architecture underpins many of the Database Machine specific monitoring and management improvements in this release.

Compared to earlier releases, the process of configuring Enterprise to manage Exadata Database Machine is greatly simplified. The new wizard-driven guided discovery process makes configuration significantly easier and faster. There are no separate plug-ins to install and the discovery of Database Machine components is done automatically by the Enterprise Manager agent.

Visualizing Exadata Database Machine in Enterprise Manager



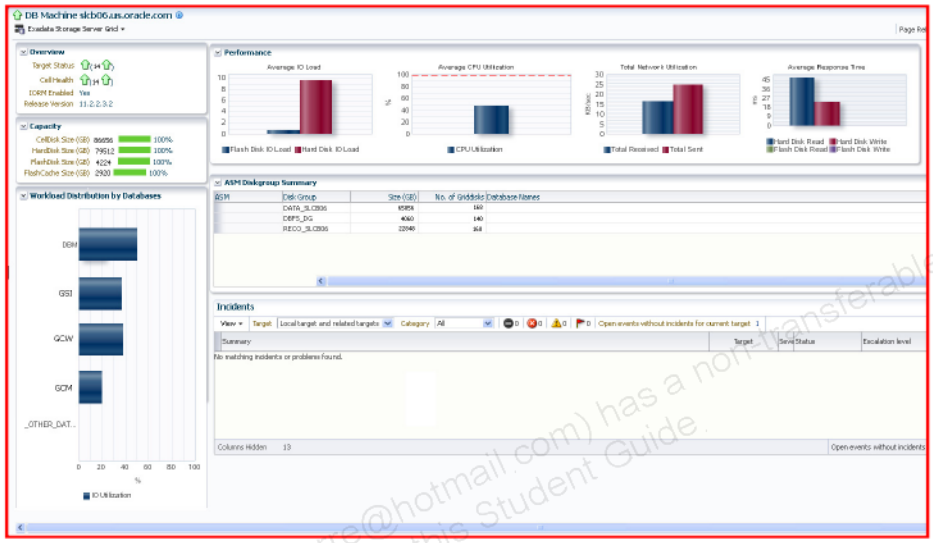
Enterprise Manager Cloud Control 12c introduces a new user interface for visualizing Exadata Database Machine in its components. It provides an integrated view of the Database Machine hardware and software components. The slide shows a selection of screen images.

The left-most image shows a graphical representation of a Full Rack Exadata Database Machine. Using this view administrators can easily relate severs and switches back to their physical location in the rack. This is particularly useful when a hardware component raises an alert and the administrator can easily identify the component with a high degree of certainty.

The images on the right side of the screen show charts relating to different cell and database server metrics. These server-organized screen layouts are specific to Database Machine, making it easy for administrators to relate the information on the screen back to specific Database Machine components.

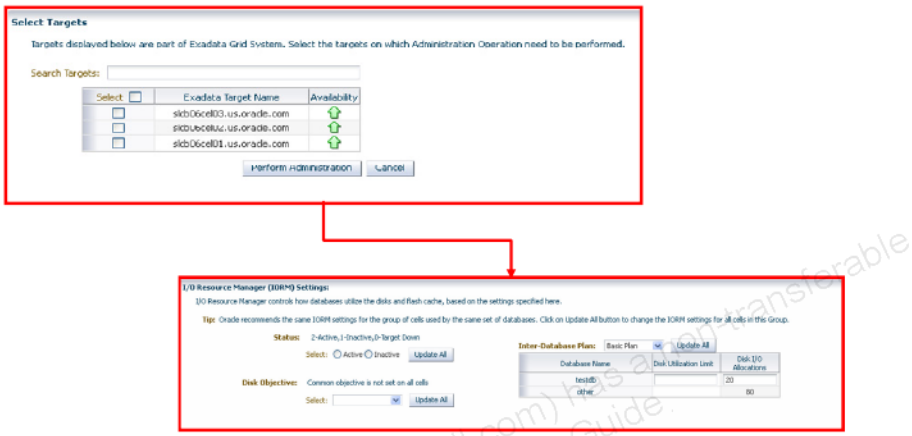
In addition to the hardware-organized Database Machine views, the are numerous system and software oriented views which allow administrators to monitor performance and availability along with resource usage by databases, services and clusters. Software alerts for databases, clusters and ASM are also presented. A topology-organized view of Database Machine clusters and database is also provided while the built-in configuration management repository tracks the version information associated with all the Database Machine components.

Exadata Storage Server Monitoring



Enterprise Manager Cloud Control 12c includes enhanced group-based monitoring and administration capabilities. By default, all the cells in a Database Machine are grouped into a cell group. As shown in the screen image, consolidated monitoring capabilities are provided for the cell group. Group-level administration operations are also provided so that actions can be applied consistently across all the cells in the group.

Exadata Storage Server Administration

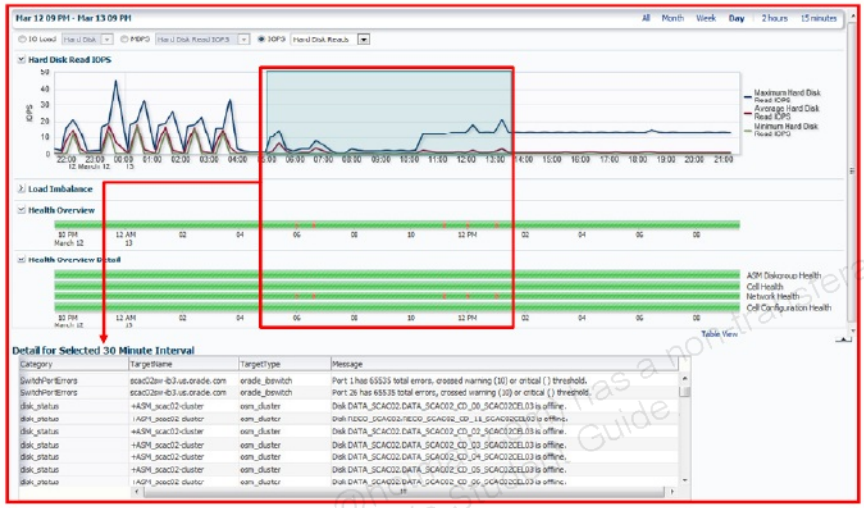


Enterprise Manager Cloud Control 12c includes the ability to perform cell configuration and administration tasks. The slide shows an example of the interface that can be used to enable, disable and configure I/O Resource Manager (IORM). Enterprise Manager Cloud Control 12c also includes a range of other cell administration functions such as the ability to start and stop cell services, verify cell connectivity and configure SSH access.

Exadata Storage Server Performance Monitoring

The screen image on the slide show an example of the Exadata Storage Server performance monitoring capabilities in Enterprise Manager Cloud Control 12c. The information presented by Enterprise Manager provides a composite view of the performance indicators for a single cell or cell group. When users point at different chart elements, additional information is displayed.

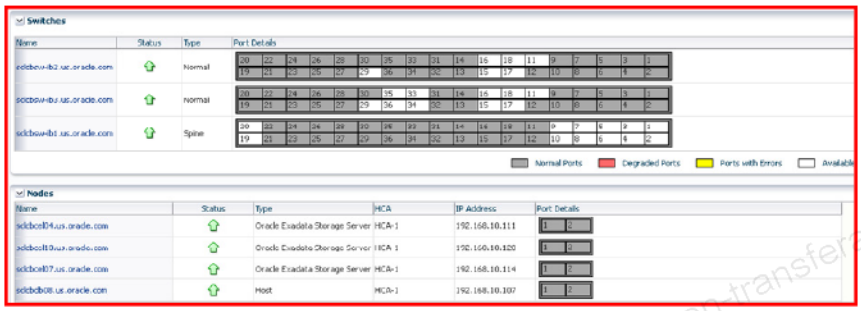
Exadata Storage Server Health Monitoring



The screen image on the slide show an example of the Exadata Storage Server health monitoring capabilities in Enterprise Manager Cloud Control 12c. The information presented by Enterprise Manager provides a composite view of the health indicators for a cell group.

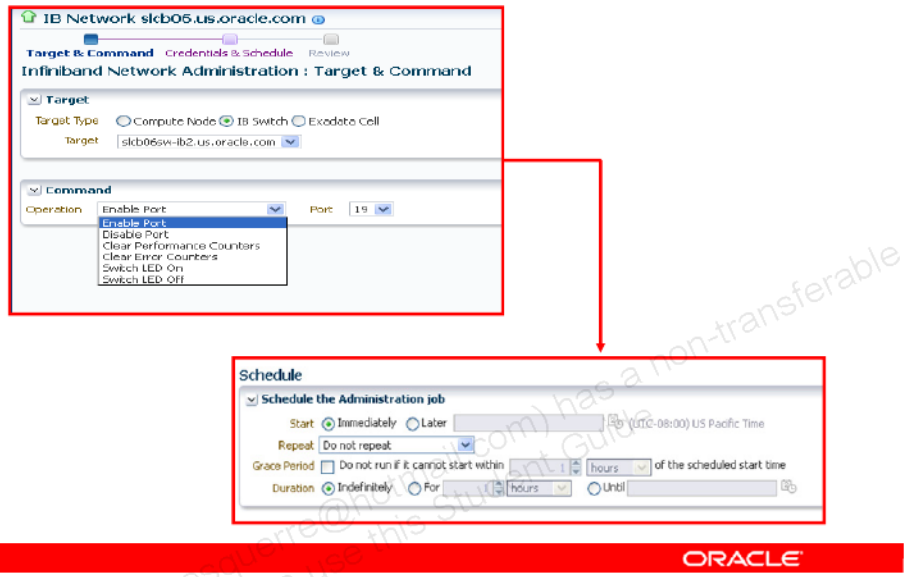
A slider enables the targeted examination of a specific time period, More detailed information for the selected time period can also be examined. The detailed information can relate to issues such as response degradation, I/O load imbalances, ASM related problems, cell software or hardware failures, cell configuration issues and network related failures.

Exadata Database Machine
InfiniBand Network Monitoring



In Enterprise Manager Cloud Control 12c, the Exadata Database Machine InfiniBand network and switches are native targets which are automatically discovered during the guided discovery process. Real time and historical performance metrics relating to the InfiniBand network are available for administrators to analyze. Also, Enterprise Manager will highlight alert conditions which are generated within the InfiniBand switches or which relate to thresholds maintained inside Enterprise Manager. Configuration metrics are used to detect and notify administrators about network changes and situations that violate best-practice recommendations.

Exadata Database Machine
InfiniBand Network Administration



New Exadata Database Machine InfiniBand network administration function are also available in Enterprise Manager Cloud Control 12c. The screen images in the slide shows an example of some of the InfiniBand network administration screens. The available administration actions include:

- Enable an InfiniBand port.
- Disable an InfiniBand port.
- Clear performance counters.
- Clear error counters.
- Switch on an indicator LED.
- Switch off an indicator LED.

Summary

In this lesson you should have learned how to outline the key Exadata Database Machine management capabilities included in Enterprise Manager Cloud Control 12c.