Hide Navigation

Reference
Home · Master Index · Master Glossary · Master Book List

Contents
- Oracle Exadata Storage Server Software User's Guide
- Preface
- **Introducing Oracle Exadata Storage Server Software**
- Configuring Oracle Exadata Storage Server Software
- Administering Oracle ASM Disk Groups on Oracle Exadata Storage Servers
- Configuring Security for Oracle Exadata Storage Server Software
- Maintaining Oracle Exadata

Oracle® Exadata Storage Server Software User's Guide
12c Release 1 (12.1)
E50471-14

Home Book List Contents Index Contact Us

Previous Next

PDF · Mobi · ePub

1 Introducing Oracle Exadata Storage Server Software

This chapter introduces Oracle Exadata Storage Server Software. This chapter contains the following topics:

- Overview of Oracle Exadata Storage Server Software
- Key Features of Oracle Exadata Storage Server Software
- Oracle Exadata Storage Server Software Components

Overview of Oracle Exadata Storage Server Software

Oracle Exadata Storage Server is a highly optimized storage server that runs Oracle Exadata Storage Server Software to store and access Oracle Database data. With traditional storage, data is transferred to the database server for processing. In contrast, Oracle Exadata Storage Server Software provides database-aware storage services, such as the ability to offload SQL and other database processing from the database server, while remaining transparent to the SQL processing and database applications. Oracle Exadata Storage Servers process data at the storage level, and pass only what is needed to the database server. Oracle Exadata Storage Server can also be used in addition to traditional storage arrays and products.

Oracle Exadata Storage Server Software offloads some SQL processing from the database server to the Oracle Exadata Storage Servers. Oracle Exadata Storage Server Software enables function shipping between the database instance and the underlying storage, in addition to traditional data shipping. Function shipping greatly reduces the amount of data processing that must be done by the database server. Eliminating data transfers and database server workload can greatly benefit query processing operations that often become bandwidth constrained. Eliminating data transfers can also provide a significant benefit to online transaction processing (OLTP) systems that include large batch and report processing operations.

The hardware components of Oracle Exadata Storage Server are carefully chosen to match the needs of high performance processing. The cell software is optimized to maximize the advantage of the hardware components. Each cell delivers outstanding processing bandwidth for data stored on disk, often several times better than traditional solutions.

Oracle Exadata Storage Servers use state-of-the-art InfiniBand interconnections between servers and storage. Each InfiniBand link provides 40 gigabits per second of bandwidth, many times higher than traditional storage or server networks. Additionally, Oracle interconnection protocol uses direct data placement, also referred to as direct memory access (DMA), to ensure low CPU overhead by directly moving data from the wire to database buffers with no extra copies. The InfiniBand network has the flexibility of a LAN network with the efficiency of a storage area network (SAN). With an InfiniBand network, Oracle eliminates network bottlenecks that could reduce performance. This InfiniBand network also provides a high performance cluster interconnection for Oracle Real Application Clusters (Oracle RAC) servers.

The Oracle Exadata Storage Server architecture scales to any level of performance. To achieve higher performance or greater storage capacity, you add more storage cells to the configuration. As more cells are added, capacity and performance increase linearly. Data is mirrored across cells to ensure that the failure of a cell does not cause loss of data or availability. The scale-out architecture achieves near infinite scalability, while lowering costs by allowing storage to be purchased incrementally on demand.

Note:
Oracle Exadata Storage Server Software must be used with Oracle Exadata Storage Server hardware, and only supports databases on the database servers of Oracle Exadata Database Machines. Information is available on My Oracle Support at

`http://support.oracle.com/`

and on the Products page of Oracle Technology Network at

`http://www.oracle.com/technetwork/index.html`

Key Features of Oracle Exadata Storage Server Software

The key features of Oracle Exadata Storage Server Software include the following:

- Reliability, Modularity, and Cost-Effectiveness
- Compatibility with Oracle Database
- Smart Flash Technology
- Centralized Storage
- I/O Resource Management
- Offloading of Data Search and Retrieval Processing
- Offloading of Incremental Backup Processing

- Protection Against Data Corruption
- Fast File Creation
- Storage Index

Reliability, Modularity, and Cost-Effectiveness

Oracle Exadata Storage Server Software enables cost-effective modular storage hardware to be used in a scale-out architecture while providing a high level of availability and reliability. All single points of failure are eliminated in the Oracle Exadata Storage Server architecture by data mirroring, fault isolation technology, and protection against disk and other storage hardware failure.

In the Oracle Exadata Storage Server architecture, one or more storage cells can support one or more databases. The placement of data is transparent to database users and applications. Storage cells use **Oracle Automatic Storage Management (Oracle ASM)** to distribute data evenly across the cells. Because Oracle Exadata Storage Servers support dynamic disk insertion and removal, the online dynamic data redistribution feature of Oracle ASM ensures that data is appropriately balanced across the newly added, or remaining, disks without interrupting database processing. Oracle Exadata Storage Servers also provides data protection from disk and cell failures.

Compatibility with Oracle Database

Oracle Exadata Storage Server Software 12c Release 1 (12.1) requires Oracle Database and Oracle ASM 12c Release 1 (12.1). All Oracle Database features are fully supported with Oracle Exadata Storage Server Software. Oracle Exadata Storage Server Software works equally well with single-instance or Oracle RAC deployments of Oracle Database. Oracle Data Guard, Oracle Recovery Manager (RMAN), Oracle Streams, and other database administration tools are managed the same with Oracle Exadata Storage Server as with traditional storage. This enables database administrators to use the same tools with which they are familiar.

Smart Flash Technology

Oracle has implemented smart flash cache directly in Oracle Exadata Storage Server. Oracle Exadata Smart Flash Cache holds frequently-accessed data in very fast flash storage while most data is kept in very cost-effective disk storage. This happens automatically without the user having to take any action. Oracle Exadata Smart Flash Cache is smart because it knows when to avoid trying to cache data that will never be reused or will not fit in the cache. Oracle Database and Oracle Exadata Storage Server Software allow the user to provide directives at the database table, index and segment level to ensure that specific data is retained in flash. Tables can be moved in and out of flash with a simple command, without the need to move the table to different tablespaces, files or LUNs as is done with traditional storage using flash disks.

Oracle Exadata Smart Flash technology is also used to reduce the latency of log write I/O operations by eliminating performance bottlenecks that might occur due to database logging. The time to commit user transactions is very sensitive to the latency of log write operations. In addition, many performance-critical database algorithms, such as space management and index splits, are very sensitive to log write latency.

Although the disk controller has a large battery-backed DRAM cache that can accept writes very quickly, some write operations to disk can still be slow during periods of high I/O. Even with relatively few redo log write operations that are slow, these write operations can cause performance issues. It is these situations that Oracle Exadata Smart Flash Log is designed to alleviate.

The goal of the Oracle Exadata Smart Flash Log is to perform redo write operations simultaneously to both flash memory and disk, and complete the write operation when the first of the two completes. This gives Oracle Exadata the best of both worlds by avoiding problems due to latency spikes on either type of media. Smart Flash Logging is most beneficial during busy periods when the disk controller cache occasionally becomes filled with blocks that have not been written to disk and therefore degrades to real disk performance versus disk cache performance. It is important to note that Smart Flash Logging improves latency of log write operations, but it does not improve total disk throughput. If an application is bottlenecked on disk throughput, then Smart Flash Logging can provide little benefit because log response time is not the limiting factor to performance.

It is also crucial to note that the purpose of Smart Flash Logging is not to use flash to consistently beat disk controller performance. It is used as an auxiliary destination that provides low latency when disks occasionally become slow, thus avoiding a negative impact on database performance.

Oracle Exadata Smart Flash Log improves user transaction response time, and increases overall database throughput for I/O intensive workloads by accelerating performance critical database algorithms.

See Also:
"ALTER CELL" for information about write back and write through flash cache

Centralized Storage

You can use Oracle Exadata Storage Server to consolidate your storage requirements into a central pool that can be used by multiple databases. Oracle Exadata Storage Server Software with Oracle ASM evenly distributes the data and I/O load for every database across available disks in the storage pool. Every database can use all of the available disks

to achieve superior I/O rates. Oracle Exadata Storage Servers can provide higher efficiency and performance at a lower cost while also lowering your storage administration overhead.

I/O Resource Management

I/O Resource Management (IORM) and the Database Resource Manager process enable multiple databases and pluggable databases to share the same storage while ensuring that I/O resources are allocated across the various databases. Oracle Exadata Storage Server Software works with IORM and Database Resource Manager to ensure that customer-defined policies are met, even when multiple databases share the grid. As a result, one database cannot monopolize the I/O bandwidth and degrade the performance of the other databases.

IORM enables storage cells to service disk I/O resources among multiple applications and users across all databases in accordance with sharing and prioritization levels established by the administrator. This improves the coexistence of online transaction processing (OLTP) and reporting workloads, because latency-sensitive OLTP applications can be given a larger share of disk I/O bandwidth than throughput-sensitive batch applications. Database Resource Manager enables the administrator to control processor utilization on the database host on a per-application basis. Combining IORM and Database Resource Manager enables the administrator to establish more accurate policies.

IORM for a database or pluggable database is implemented and managed from the Database Resource Manager. Database Resource Manager in the database instance communicates with the IORM software in the storage cell to manage user-defined service-level targets. Database resource plans are administered from the database, while interdatabase plans are administered on the storage cell.

See Also:
Chapter 6, "Managing I/O Resources" for additional information about IORM

Offloading of Data Search and Retrieval Processing

One of the most powerful features of Oracle Exadata Storage Server Software is that it offloads the data search and retrieval processing to the storage cell. Oracle Exadata Storage Server Software does this by performing predicate filtering, which entails evaluating database predicates to optimize the performance of certain classes of bulk data processing.

Oracle Database can optimize the performance of queries that perform table and index scans to evaluate selective predicates in Oracle Exadata Storage Server. The database can complete these queries faster by pushing the database expression evaluations to the storage cell. These expressions include simple SQL command predicates, such as `amount > 200`, and column projections, such as `SELECT customer_name`. For example:

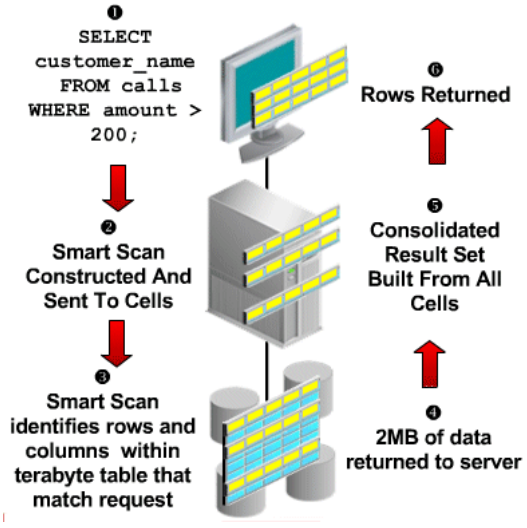
```
SQL> SELECT customer_name FROM calls WHERE amount > 200;
```

In the preceding example, only rows satisfying the predicate, specified columns, and predicated columns are returned to the database server, eliminating unproductive data transfer to the database server.

Oracle Exadata Storage Server Software uses storage-side predicate evaluation that transfers simplified, predicate evaluation operations for table and index scans to the storage cell. This brings the table scan closer to the disk to enable a higher bandwidth, and prevents sending unmatched rows to hosts.

Figure 1-1 shows an example of the query process.

Figure 1-1 Offloading Data Search and Retrieval



Description of "Figure 1-1 Offloading Data Search and Retrieval"

Offloading of Incremental Backup Processing

To optimize the performance of incremental backups, the database can offload block filtering to Oracle Exadata Storage Server. This optimization is only possible when taking backups using Oracle Recovery Manager (RMAN). The offload processing is done transparently without user intervention. During offload processing, Oracle Exadata Storage Server Software 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, making backups significantly faster.

See Also:

- "Using V$BACKUP_DATAFILE with Oracle Exadata Storage Server"
- Oracle Database Backup and Recovery User's Guide

Protection Against Data Corruption

Oracle Exadata Storage Server Software is compliant with the Oracle Hardware Assisted Resilient Data (HARD) initiative, a joint initiative between Oracle and hardware vendors to prevent data corruptions from being written to disks. Data corruptions, while rare, can have a catastrophic effect on a database, and therefore on a business. Oracle Exadata Storage Server Software takes data protection to the next level by protecting business data, not just the physical bits.

The key approach to detecting and preventing corrupted data is block checking in which the storage subsystem validates the Oracle block contents. Oracle Database validates and adds protection information to the database blocks, while Oracle Exadata Storage Server Software detects corruptions introduced into the I/O path between the database and storage. It stops corrupted data from being written to disk, and validates data when reading the disk. This eliminates a large class of failures that the database industry had previously been unable to prevent.

Oracle Exadata Storage Server Software implements all the HARD checks, and because of its tight integration with Oracle Database, additional checks are implemented that are specific to Oracle Exadata Storage Server Software. Unlike other implementations of HARD checking, HARD checks with Oracle Exadata Storage Server Software operate completely transparently. No parameters need to be set at the database or storage tier. The HARD checks transparently handle all cases, including Oracle ASM disk rebalance operations and disk failures.

Storing the Server Parameter File on HARD-enabled Storage

The server parameter file (SPFILE) is compliant with the HARD specifications implemented by Oracle Exadata Database Machine. To fully enable HARD protection for the data in the SPFILE, the SPFILE must reside on Oracle Exadata Storage Server.

The HARD-compliant SPFILE can be stored on non-HARD-enabled storage. In this case, the SPFILE format supports only detection of corrupt SPFILE data. Storing the SPFILE on HARD-enabled storage prevents corrupt data from being written to storage.

See Also:

- Oracle Database High Availability Overview for additional information about the Hardware Assisted Resilient Data (HARD) Initiative
- Oracle Maximum Availability Architecture (MAA) Web site for additional information about the Hardware Assisted Resilient Data (HARD) initiative at

  http://www.oracle.com/technetwork/database/features/availability/maa-090890.html?ssSourceSiteId=ocomen

Fast File Creation

File creation operations are offloaded to Oracle Exadata Storage Servers. Operations such as `CREATE TABLESPACE`, which can create one or more files, have a significant increase in speed due to file creation offload.

Storage Index

Oracle Exadata Storage Servers maintain a storage index which contains a summary of the data distribution on the disk. The storage index is maintained automatically, and is transparent to Oracle Database. It is a collection of in-memory region indexes, and each region index stores summaries for up to eight columns. Storage indexes work with any non-linguistic data type, and work with linguistic data types similar to non-linguistic indexes.There is one region index for each 1 MB of disk space.

The content stored in one region index is independent of the other region indexes. This makes them highly scalable, and avoids latch contention. For each region index, the storage index maintains the minimum and maximum values of the columns of the table. The minimum and maximum values are used to eliminate unnecessary I/O, also known as I/O

filtering. The cell physical IO bytes saved by storage index statistic, available in the V$SYS_STAT view, shows the number of bytes of I/O saved using storage index.

Queries using the following comparisons are improved by the storage index:

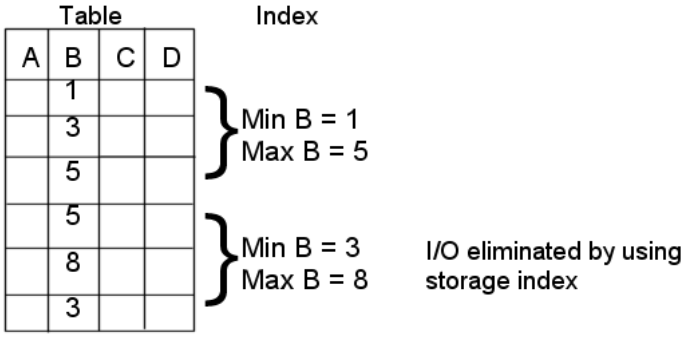
- Equality (=)
- Inequality (<, !=, or >)
- Less than or equal (<=)
- Greater than or equal (>=)
- IS NULL
- IS NOT NULL

Storage indexes are built automatically after Oracle Exadata Storage Server Software receives a query with a comparison predicate that is greater than the maximum or less than the minimum value for the column in a region, and would have benefited if a storage index had been present. Oracle Exadata Storage Server Software automatically learns which storage indexes would have benefited a query, and then creates the storage index automatically so that subsequent similar queries benefit.

Note:
The effectiveness of storage indexes can be improved by ordering the rows based on columns that frequently appear in WHERE query clauses.

Example 1-1 Elimination of Disk I/O with Storage Index

The following figure shows a table and region indexes. The values in the table range from one to eight. One region index stores the minimum 1, and the maximum of 5. The other region index stores the minimum of 3, and the maximum of 8.



Description of the illustration examples.png

For a query such as SELECT * FROM TABLE WHERE B<2, only the first set of rows match. Disk I/O is eliminated because the minimum and maximum of the second set of rows do not match the WHERE clause of the query.

Example 1-2 Partition Pruning-like Benefits with Storage Index

As shown in the following figure, there is a table named Orders with the columns Order_Number, Order_Date, Ship_Date, or Order_Item. The table is range partitioned by Order_Date column.

Orders Table

Order#	Order_Date	Ship_Date	Item
1	2010	2010	
2	2011	2011	
3	2012	2012	

Description of the illustration table.png

The following query looks for orders placed since January 1, 2012:

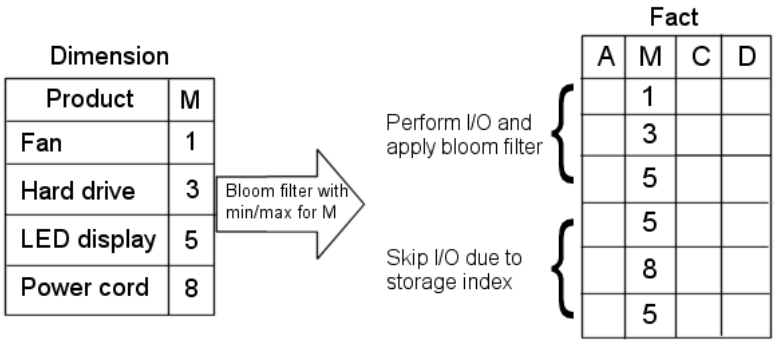
```
SELECT count (*) FROM Orders WHERE Order_Date >= to_date ('2012-01-01', \
'YYY-MM-DD')
```

Because the table is partitioned on the Order_Date column, the preceding query avoids scanning unnecessary partitions of the table. Queries on Ship_Date do not benefit from Order_Date partitioning, but Ship_Date and Order# are highly correlated with Order_Date. Storage indexes take advantage of ordering created by partitioning or sorted loading, and can use it with the other columns in the table. This provides partition pruning-like performance for queries on the Ship_Date and Order# columns.

Example 1-3 Improved Join Performance Using Storage Index

Using storage index allows table joins to skip unnecessary I/O operations. For example, the following query would perform an I/O operation and apply a Bloom filter to only the first block of the fact table.

```
SELECT count(*) from fact, dim where fact.m=dim.m and dim.name="Hard drive"
```



Description of the illustration bloom.png

The I/O for the second block of the fact table is completely eliminated by storage index as its minimum/maximum range (5,8) is not present in the Bloom filter.

Note:
The storage index is maintained during write operations to uncompressed blocks and OLTP compressed blocks. Write operations to Exadata Hybrid Columnar Compression compressed blocks or encrypted tablespaces invalidate a region index, but not the storage index. The storage index for Exadata Hybrid Columnar Compression is rebuilt on subsequent scans.

See Also:
"Using V$SYSSTAT with Oracle Exadata Storage Server Software"

Oracle Exadata Storage Server Software Components

This section provides a summary of the following Oracle Exadata Storage Server Software components:

- About Oracle Exadata Storage Server Software
- About Oracle Automatic Storage Management
- About Grid RAID
- About Cell Security
- About iDB Protocol
- About Cell Software Processes
- About Cell Management
- About Database Server Software
- About Oracle Enterprise Manager for Oracle Exadata Database Machine

About Oracle Exadata Storage Server Software

Oracle Exadata Storage Server is a network-accessible storage device with Oracle Exadata Storage Server Software installed on it. The software communicates with the database using a specialized iDB protocol, and provides both simple I/O functionality, such as block-oriented reads and writes, and advanced I/O functionality, including predicate offload and I/O resource management. Each storage cell has a physical disk. The physical disk is an actual device within the storage cell that constitutes a single disk drive spindle.

Within the storage cells, a logical unit number (LUN) defines a logical storage resource from which a single cell disk can be created. The LUN refers to the access point for storage resources presented by the underlying hardware to the upper software layers. The precise attributes of a LUN are configuration-specific. For example, a LUN could be striped, mirrored, or both striped and mirrored.

A cell disk is an Oracle Exadata Storage Server Software abstraction built on the top of a LUN. After a cell disk is created from the LUN, it is managed by Oracle Exadata Storage Server Software and can be further subdivided into grid disks, which are directly exposed to the database and Oracle ASM instances. Each grid disk is a potentially noncontiguous partition of the cell disk that is directly exposed to Oracle ASM to be used for the Oracle ASM disk group creations and expansions.

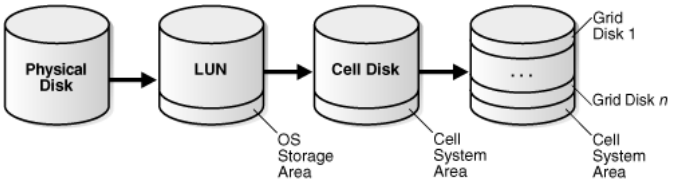
This level of virtualization enables multiple Oracle ASM clusters and multiple databases to share the same physical disk. This sharing provides optimal use of disk capacity and bandwidth. Various metrics and statistics collected on the cell disk level enable you to evaluate the performance and capacity of Oracle Exadata Storage Servers. I/O Resource Management schedules the cell disk access in accordance with user-defined policies.

Figure 1-2 illustrates how the components of a cell are related to grid disks.

- A LUN is created from a physical disk.

- A cell disk is created on a LUN. A segment of cell disk storage is used by the Oracle Exadata Storage Server Software system.
- Multiple grid disks can be created on a cell disk.

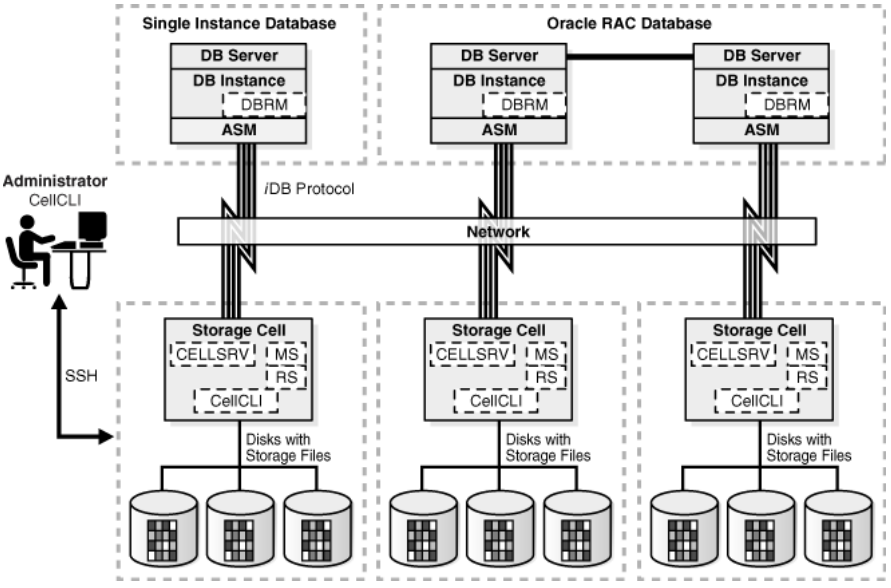
Figure 1-2 Oracle Exadata Storage Server Components



Description of "Figure 1-2 Oracle Exadata Storage Server Components"

Figure 1-3 illustrates software components in the Oracle Exadata Storage Server environment.

Figure 1-3 Software Components in the Oracle Exadata Storage Server Environment



Description of "Figure 1-3 Software Components in the Oracle Exadata Storage Server Environment"

The figure illustrates the following environment:

- Single-instance or Oracle RAC databases access storage cells using the iDB protocol over an InfiniBand network.
- The database server software includes Oracle Exadata Storage Server Software functionality.
- Storage cells contain cell-based software.
- Storage cells are configured on the network, and are managed by the Oracle Exadata Storage Server Software CellCLI utility.

About Oracle Automatic Storage Management

Oracle Automatic Storage Management (Oracle ASM) is the cluster volume manager and file system used to manage Oracle Exadata Storage Server resources. Oracle ASM provides enhanced storage management by:

- Striping database files evenly across all available storage cells and disks for optimal performance.
- Using mirroring and failure groups to avoid any single point of failure.
- Enabling dynamic add and drop capability for nonintrusive cell and disk allocation, deallocation, and reallocation.
- Enabling multiple databases to share storage cells and disks.

> **See Also:**
> Oracle Automatic Storage Management Administrator's Guide for additional information about Oracle ASM

Automatic Storage Management Disk Group

An Oracle ASM disk group is the primary storage abstraction within Oracle ASM, and is composed of one or more disks.

Oracle Exadata Storage Server grid disks appear to Oracle ASM as individual disks available for membership in Oracle ASM disk groups. Whenever possible, grid disk names should correspond closely with Oracle ASM disk group names to assist in problem diagnosis between Oracle ASM and Oracle Exadata Storage Server Software.

The Oracle ASM disk groups are as follows:

DATA is the data disk group.

RECO is the recovery disk group.

DBFS (Oracle Database File System) is the file system disk group.

SPARSE is a sparse disk group to keep snapshot files.

To take advantage of Oracle Exadata Storage Server Software features, such as predicate processing offload, the disk groups must contain only Oracle Exadata Storage Server grid disks, and the tables must be fully inside these disk groups.

Note:
The database and grid infrastructure must be release 12.1.0.2.0 BP3 or later when using sparse grid disks.

> **See Also:**
> Oracle Automatic Storage Management Administrator's Guide for additional information about Oracle ASM

Automatic Storage Management Failure Group

An Oracle ASM failure group is a subset of disks in an Oracle ASM disk group that can fail together because they share the same hardware. Oracle ASM considers failure groups when making redundancy decisions.

For Oracle Exadata Storage Servers, all grid disks, which consist of the Oracle ASM disk group members and candidates, can effectively fail together if the storage cell fails. Because of this scenario, all Oracle ASM grid disks sourced from a given storage cell should be assigned to a single failure group representing the cell.

For example, if all grid disks from two storage cells, A and B, are added to a single Oracle ASM disk group with normal redundancy, then all grid disks on storage cell A are designated as one failure group, and all grid disks on storage cell B are designated as another failure group. This enables Oracle Exadata Storage Server Software and Oracle ASM to tolerate the failure of either storage cell.

Failure groups for Oracle Exadata Storage Server grid disks are set by default so that the disks on a single cell are in the same failure group, making correct failure group configuration simple for Oracle Exadata Storage Servers.

You can define the redundancy level for an Oracle ASM disk group when creating a disk group. An Oracle ASM disk group can be specified with normal or high redundancy. Normal redundancy double mirrors the extents, and high redundancy triple mirrors the extents. Oracle ASM normal redundancy tolerates the failure of a single cell or any set of disks in a single cell. Oracle ASM high redundancy tolerates the failure of two cells or any set of disks in two cells. Base your redundancy setting on your desired protection level. When choosing the redundancy level, ensure the post-failure I/O capacity is sufficient to meet the redundancy requirements and performance service levels. Oracle recommends using three cells for normal redundancy. This ensures the ability to restore full redundancy after cell failure. Consider the following:

- If a cell or disk fails, then Oracle ASM automatically redistributes the cell or disk contents across the remaining disks in the disk group as long as there is enough space to hold the data. For an existing disk group using Oracle ASM redundancy, the `USABLE_FILE_MB` and `REQUIRED_FREE_MIRROR_MB` columns in the `V$ASM_DISGKROUP` view give the amount of usable space and space for redundancy, respectively.
- If a cell or disk fails, then the remaining disks should be able to generate the MBPS and IOPS necessary to sustain the performance service level agreement.

After a disk group is created, the redundancy level of the disk group cannot be changed. To change the redundancy of a disk group, you must create another disk group with the appropriate redundancy, and then move the files.

Each Exadata Cell is a failure group. A normal redundancy disk group must contain at least two failure groups. Oracle ASM automatically stores two copies of the file extents, with the mirrored extents placed in different failure groups. A high redundancy disk group must contain at least three failure groups. Oracle ASM automatically stores three copies of the file extents, with each file extent in separate failure groups.

System reliability can diminish if your environment has an insufficient number of failure groups. A small number of failure groups, or failure groups of uneven capacity, can lead to allocation problems that prevent full use of all available storage.

> **See Also:**
> Example 3-1, "Creating Oracle ASM Disk Groups for Oracle Exadata Storage Server" for an example of creating an Oracle ASM disk group

About Oracle ASM for Maximum Availability

Oracle recommends high redundancy Oracle ASM disk groups, and file placement configuration which can be automatically deployed using Oracle Exadata Deployment Assistant. High redundancy can be configured for DATA, RECO or any other Oracle ASM group with a minimum of 3 storage cells. The voting disks can reside in a normal redundancy DBFS disk group when there are less than 5 storage cells or the first high redundancy disk group containing 5 or more storage cells.

Maximum availability architecture (MAA) best practice uses three Oracle ASM disk groups, DATA, RECO, and DBFS. The disk groups are located as follows:

- The disk groups are striped across all disks and Oracle Exadata Storage Servers to maximize I/O bandwidth and performance, and simplify management.
- The DATA disk group is located on the outer section of all disks.
- The RECO disk group is located on the outer/inner section of all disks.
- The DBFS disk group is located on the inner section of all disks.
- The DATA and RECO disk groups are configured for high redundancy.

The preceding attributes ensure optimal file placement in the different Oracle ASM disk groups. In addition, all operations have access to full I/O bandwidth, when needed. To avoid excessive resource consumption, use I/O Resource Management, Oracle Database Resource Manager, and instance caging.

The benefits of high redundancy disk groups are illustrated by the following outage scenarios:

- Double partner disk failure: Protection against loss of the database and Oracle ASM disk group due to a disk failure followed by a second disk failure of a partner disk. If the voting disk resides on normal redundancy disk group, then the database cluster will fail and the database has to be restarted. If the voting disk also resides in high redundancy disk group, the cluster and database will remain available.
- Disk failure when Oracle Exadata Storage Server is offline: Protection against loss of the database and Oracle ASM disk group when a storage server is offline and one of the storage server's partner disks fails. The storage server may be offline because of planned maintenance, such as rolling storage server patching. If the voting disk resides on normal redundancy disk group, then the database cluster will fail and the database has to be restarted. If the voting disk also resides in high redundancy disk group, the cluster and database will remain available.
- Disk failure followed by disk sector corruption: Protection against data loss and I/O errors when latent disk sector corruptions exist and a partner storage disk is unavailable either due to planned maintenance or disk failure.

Oracle recommends High Redundancy for ALL (DATA and RECO) disk groups because it provides maximum application availability against storage failures. In contrast, if all disk groups were configured with normal redundancy and two partner disk fails, all clusters and databases on Exadata will fail and you will lose all your data (normal redundancy does not survive double partner disk failures). Other than better storage protection, the major difference between high redundancy and normal redundancy is the amount of usable storage and write I/Os. High redundancy requires more space, and has three write I/Os instead of two. The following table describes that redundancy option, as well as others, and the relative availability trade-offs.

Redundancy Option	Availability Implications	Recommendation
High Redundancy for ALL (DATA and RECO)	Zero application downtime and zero data loss for the preceding storage outage scenarios if voting disks reside in high redundancy disk group. If voting disks reside in normal redundancy disk group, the database has to be restarted but zero data loss to the database is maintained.	Use this option for best storage protection for mission-critical applications. Requires more space for higher redundancy.
High Redundancy for DATA only	Zero application downtime and zero data loss for preceding storage outage scenarios if voting disks reside in high redundancy disk group. If voting disks reside in normal redundancy disk group, the database has to be restarted but zero data loss to the database is maintained. This option requires an alternative archive destination.	Use this option for better storage protection for DATA while trading RECO protection for more available space.
High Redundancy for LOG and RECO only	Zero data loss for the preceding storage outage scenarios.	Use this option when longer recovery times are acceptable for the preceding storage outage scenarios. Recovery options include the following: • Restore and recover: - Recreate DATA disk group - Restore from RECO and tape-based backups, if required - Recover database • Switch and recover:

		- Use RMAN switch to copy Recover database
Normal Redundancy for ALL (DATA and RECO) **Note:** Cross-disk mirror isolation limits an outage to a single disk group when two disk partners are lost in a normal redundancy group that share physical disks and storage servers.	The preceding storage outage scenarios resulted in failure of all Oracle ASM disk groups. However, using cross-disk group mirror isolation the outage is limited to one disk group. **Note:** This option is not available for quarter racks.	Use the Normal Redundancy for ALL option when the primary database is protected by an Oracle Data Guard standby database deployed on a separate Oracle Exadata Database Machine. Oracle Data Guard provides real-time data protection and fast failover for storage failures. If Oracle Data Guard is not available and the DATA or RECO disk groups are lost, then leverage recovery options described in My Oracle Support note 1339373.1

The following table describes the optimal file placement for setup for MAA:

File Type	**Location**
Oracle Database files	DATA disk group.
Flashback log files, archived redo files, and backup files	RECO disk group.
Redo log files, and control files	First high redundancy disk group. If no high redundancy disk group exists, then redo log files are multiplexed across the DATA and RECO disk groups.
Control file	First high redundancy disk group. If no high redundancy disk groups exist, the use one control file in the DATA disk group. The backup control files should reside in the RECO disk group, and `RMAN CONFIGURE CONTROLFILE AUTOBACKUP ON` should be set.
Server parameter files (SPFILE)	First high redundancy disk group. If no high redundancy disk group exists, then SPFILE should reside in the DATA disk group. SPFILE backups should reside in the RECO disk group.
Oracle Cluster Registry (OCR) and voting disks for Oracle Exadata Database Machine Full Rack and Oracle Exadata Database Machine Half Rack	First high redundancy disk group. If no high redundancy disk group exists, then the files should reside in the DATA disk group.
Voting disks for Oracle Exadata Database Machine Quarter Rack or Eighth Rack	Normal redundancy disk group[Foot 1]
Temporary files	First normal redundancy disk group. If the High Redundancy for ALL option is used, then the use the first high redundancy disk group.
Staging and non-database files	DBFS disk group

Footnote 1 To place voting disks in a high redundancy Oracle ASM disk group, a minimum of five Exadata Cells are needed. If the voting disks reside in normal redundancy and you use lose two partner disks, then Oracle RAC cluster will fail because there's no longer a quorum. The database can be restarted if the database files reside in high redundancy disk groups.

See Also:
Oracle Exadata Database Machine Installation and Configuraton Guide

About Grid RAID

A grid Redundant Array of Independent Disks (RAID) configuration uses Oracle ASM mirroring capabilities. To use grid RAID, you place grid disks in an Oracle ASM disk group with a normal or high redundancy level, and set all grid disks in the same cell to be in the same Oracle ASM failure group. This ensures that Oracle ASM does not mirror data extents using disks within the cell. Using disks from different cells ensures that an individual cell failure does not cause the data to be unavailable.

Grid RAID also provides simplified creation of cell disks. With grid RAID, LUNs are automatically created from available physical disks because Oracle software automatically creates the required LUNs.

About Cell Security

Security for Exadata Cell is enforced by identifying which clients can access cells and grid disks. Clients include Oracle ASM instances, database instances, and clusters. When creating or modifying grid disks, you can configure the Oracle ASM owner and the database clients that are allowed to use those grid disks.

About iDB Protocol

The iDB protocol is a unique Oracle data transfer protocol that serves as the communications protocol among Oracle ASM, database instances, and storage cells. General-purpose data transfer protocols operate only on the low-level blocks of a disk. In contrast, the iDB protocol is aware of the Oracle internal data representation and is the necessary

complement to Exadata Cell-specific features, such as predicate processing offload.

In addition, the iDB protocol provides interconnection bandwidth aggregation and failover.

About Cell Software Processes

Oracle Exadata Storage Server Software includes the following software processes:

- **Cell Server (CELLSRV)** services iDB requests for disk I/O and advanced Exadata Cell services, such as predicate processing offload. CELLSRV is implemented as a multithread process and should be expected to use the largest portion of processor cycles on a storage cell.
- **Management Server (MS)** provides standalone storage cell management and configuration.
- **Restart Server (RS)** monitors the Cell Server and Management Server processes and restarts them, if necessary.

About Cell Management

Each cell in the Oracle Exadata Storage Server grid is individually managed with Cell Control Command-Line Interface (CellCLI).The CellCLI utility provides a command-line interface to the cell management functions, such as cell initial configuration, cell disk and grid disk creation, and performance monitoring. The CellCLI utility runs on the cell, and is accessible from a client computer that has network access to the storage cell or is directly connected to the cell. The CellCLI utility communicates with Management Server to administer the storage cell.

To access the cell, you should either use Secure Shell (**SSH**) access, or local access, for example, through a **KVM switch** (keyboard, video or visual display unit, mouse) switch. SSH allows remote access, but local access might be necessary during the initial configuration when the cell is not yet configured for the network. With local access, you have access to the cell operating system shell prompt and use various tools, such as the CellCLI utility, to administer the cell.

You can run the same CellCLI commands remotely on multiple cells with the dcli utility.

See Also:

- Chapter 8, "Using the CellCLI Utility" for additional information about CellCLI commands
- Chapter 9, "Using the dcli Utility" for additional information about managing multiple cells with a centralized management tool

---

About Database Server Software

Oracle Exadata Storage Server Software works seamlessly with Oracle Database. The software on the database server includes:

- Oracle Database instance, which contains the set of Oracle Database background processes that operate on the stored data and the shared allocated memory that those processes use to do their work.
- Oracle Automatic Storage Management (Oracle ASM), which provides storage management optimized for the database and Oracle Exadata Storage Servers. Oracle ASM is part of Oracle Grid Infrastructure.

  The Oracle ASM instance handles placement of data files on disks, operating as a metadata manager. The Oracle ASM instance is primarily active during file creation and extension, or during disk rebalancing following a configuration change. Run-time I/O operations are sent directly from the database to storage cells without passing through an Oracle ASM instance.
- The Database Resource Manager (DBRM), which ensures that I/O resources are properly allocated within a database.
- The iDB protocol is used by the database instance to communicate with cells, and is implemented in an Oracle-supplied library statically linked with the database server.

See Also:

- Chapter 6, "Managing I/O Resources" for additional information about DBRM
- "About iDB Protocol" for additional information about iDB protocol

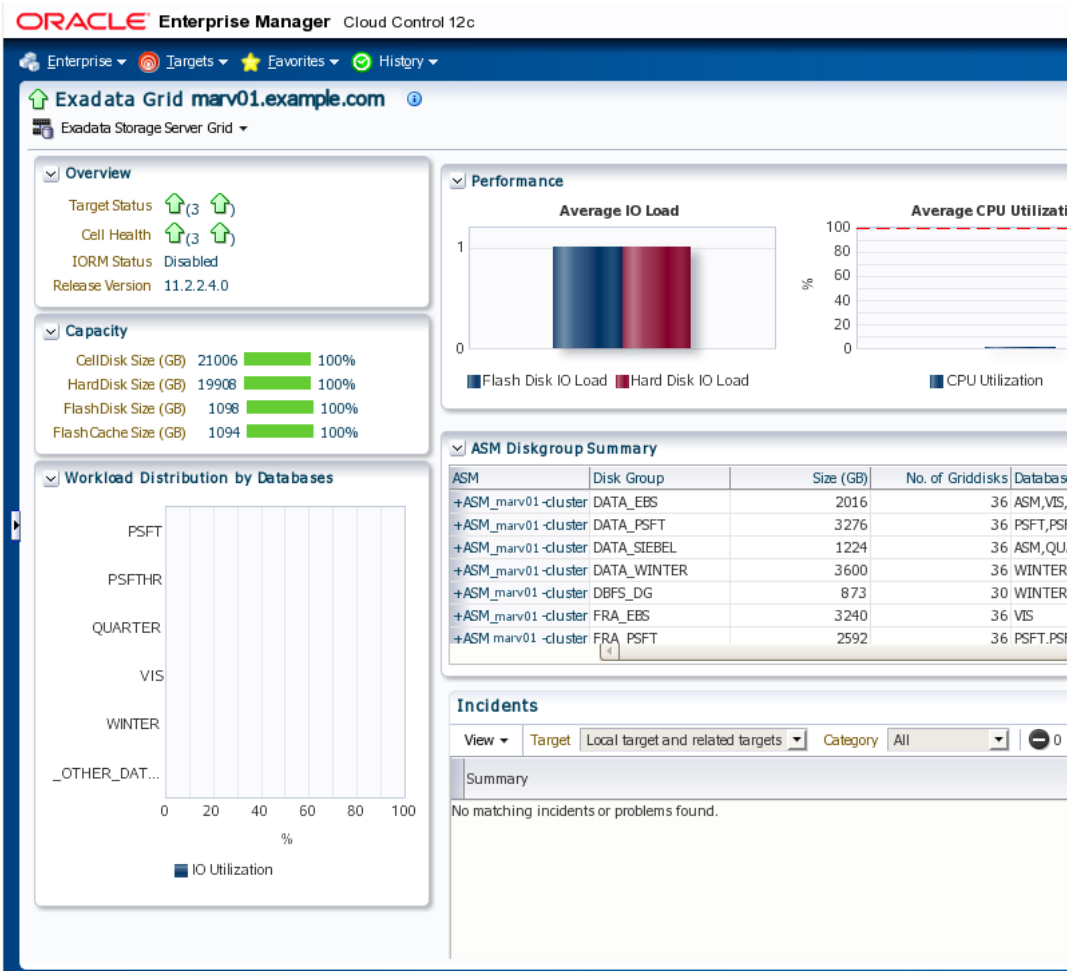
---

About Oracle Enterprise Manager for Oracle Exadata Database Machine

Oracle Enterprise Manager 12c provides a complete target that enables you to monitor Oracle Exadata Database Machine, including configuration and performance in a graphical user interface (GUI).

Figure 1-4 shows Oracle Exadata Storage Server Home Page. Viewing this page, you can quickly see the health of the storage cells, key cell performance characteristics and resource utilization of storage by individual databases.

Figure 1-4 Oracle Exadata Storage Server Home Page in Oracle Enterprise Manager



Description of "Figure 1-4 Oracle Exadata Storage Server Home Page in Oracle Enterprise Manager"

In addition to reports, Oracle Enterprise Manager for Oracle Exadata Storage Server enables you to set metric thresholds for alerts and monitor metric values to determine the health of a storage cell.

See Also:
Oracle Enterprise Manager Exadata Management Getting Started Guide for information about Oracle Enterprise Manager Exadata Target