

An Oracle White Paper
November 2009

Optimizing and Protecting Storage with Oracle Database 11g Release 2

INTRODUCTION	1
MANAGE STORAGE MORE EFFICIENTLY	2
OPTIMIZE STORAGE PERFORMANCE.....	4
REDUCE OVERALL STORAGE COSTS	8
PROTECT AGAINST DATA LOSS.....	13

INTRODUCTION

Every organization today must provide fast access to vast amounts of enterprise information for their customers, partners and business users. New, rich applications that combine relational data with XML, spatial information, documents and images increase the amount of information to be managed throughout the enterprise. Integration between applications can lead to widespread information replication, and development and testing operations regularly copy information across many different systems. The net result is that organizations are undergoing an information explosion, and are placing continual pressure on their IT departments to:

- Manage storage more efficiently: Distributed storage silos significantly limit information sharing and provisioning across information management systems.
- Optimize performance: Storage has become a barrier that prevents users from getting quick access to the information they need in order to be successful in their jobs.
- Reduce storage costs: Even while storage prices drop year after year, organizations are under constant pressure to reduce their continually increasing storage costs.
- Mitigate the risk of information loss. Enterprise information must be efficiently protected from loss or inadvertent destruction.

This document focuses on key Oracle Database 11g capabilities that help IT departments better optimize their storage infrastructure, enabling administrators to deliver a cost-effective, scalable information management platform that is easy to manage, and that continues to deliver the performance and availability that today's businesses require.

MANAGE STORAGE MORE EFFICIENTLY

“Oracle Automatic Storage Management (ASM) greatly enhances our ability to manage storage in a database. It allows multiple tiers, so that we can adapt different kinds of storage to different application needs and performance requirements. It balances on-demand for optimum performance.”

Donald Eyberg, Manager of Database Services

Embarq

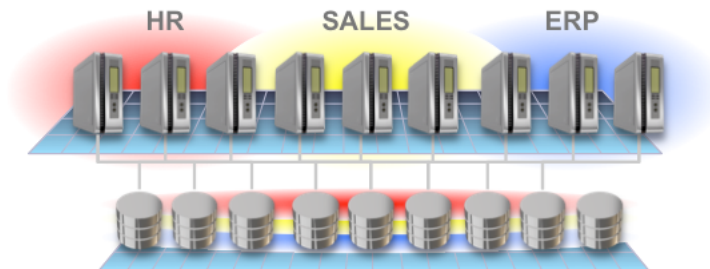
Over the years, IT departments have been plagued with silos of storage deployed under individual database environments. The first stage of effective storage optimization is to consolidate these silos together into a shared storage environment. To this end, many organizations are increasingly deploying shared storage arrays underneath their databases.

While this makes it easier to control and provision storage at the system administrator level, it adds additional complexities for database administrators – how to efficiently lay out their database files across the storage array to get both maximum performance and availability, with the least amount of effort.

In 2000, Oracle published a seminal white paper in the industry called “Optimal Storage Configuration Made Easy”. This white paper postulated a best operating procedure for managing multiple databases on a single shared array called SAME (i.e. Stripe And Mirror Everything). The basic idea is to take all the data stored in enterprise databases and stripe it equally across all disks in a storage array, and mirror that data on at least two of the disks as well. At that time, Oracle offered very little to provide this capability, and many IT departments resorted to 3rd party volume management software to implement this best practice, resulting in additional cost and complexity in their storage environments

Manage data with Automatic Storage Management

In 2004, Oracle remedied this situation with the Automatic Storage Management (ASM) feature of Oracle Database 10g.



Example of Disk Striping with ASM

ASM integrates SAME best practices into the Oracle Database kernel. ASM allows pools of storage called disk groups to be created. These disk groups essentially become containers for database files. Disk groups can house database files from many databases, thus enabling greater storage utilization and providing the foundation for storage consolidation. Database data stored in ASM is always evenly distributed across all disks in a disk group, enabling the optimal balance of I/O workload across all disks in the disk group, removing hot spots and the need for hot spot management

As disk groups become full, or as additional I/O bandwidth is required, additional disks can be added to the disk group. Adding or removing storage from a disk group automatically rebalances the data across the remaining disks in the disk group, always-ensuring even data distribution is maintained. In this manner, ASM provides some key storage management capabilities:

- Enables efficient consolidation of all types of storage underneath multiple databases.
- Ensures that the best I/O performance for database workloads is achieved from the underlying storage array
- Removes the need for 3rd party volume managers and cluster file systems – reducing the cost and complexity associated with integrating products from multiple vendors.
- Simplifies the relationship between database and systems administrators – as databases require additional space, the system administrator simply allocates more disks, and the database then automates management and optimal data placement across disks.

Oracle Database 11g Release 2 improves ASM even further. New installation and management capabilities make it easier to manage, and new intelligent data placement capabilities stripe frequently accessed data on the outer rims of disk platters, further optimizing performance. This level of automation can reduce administration time by up to 50%, and improve performance.

Manage all data with ASM Cluster File System

For many IT professionals, ASM has become the preferred storage management solution for their Oracle Database environments. However, one of their challenges is that there are other file systems outside of the database that also need to be managed. Even in an Oracle Database environment, there are many of these types of files – for example, the files and binaries that make up the Oracle software distribution.

To this end, Oracle Database 11g Release 2 introduces the ASM Cluster File System (ACFS). ACFS is a POSIX/X-OPEN general-purpose file system for files that are stored outside of the Oracle Database. With ACFS, the functionality of ASM has been extended to all files associated with Oracle Database environments. The SAME best practice operating procedure can now be consistently applied across all data without the need for costly 3rd party volume managers and cluster file systems.

OPTIMIZE STORAGE PERFORMANCE

"When it comes to speed, Oracle Exadata technology has changed the game completely....queries that used to take half an hour are now taking less than a minute."

Grant Salmon, CEO

LGR Communications

IT departments invest a great deal of money in their storage management infrastructure. With the ever-increasing speeds of new processing cores, and the availability of more memory in servers, this storage infrastructure is now becoming a major bottleneck that prevents organizations from achieving the required performance for business applications. Oracle Database 11g provides techniques and capabilities that can mitigate these performance bottlenecks.

Size disk arrays according to the workload they need to support

Traditionally, storage arrays used underneath Oracle Databases have been sized based on the amount of data that needs to be stored in corresponding databases. As disk capacity has increased, and processor cores have become faster, this sizing metric is no longer useful. Instead, storage arrays should be size based on the performance they can deliver to database applications.

Oracle Database workloads consist of random I/O (input/output) operations, that are typically caused by data manipulation operations such as inserts, updates and deletes; and sequential I/O operations that are caused by queries. These operations are typically measured as follows:

- Random I/Os are measured in I/O Operations per Second, or IOPs.
- Sequential I/O is measured in the number of megabytes of data that can be scanned from the storage system per second, or MB/s.

Storage Arrays should be sized on the number of IOPs they deliver, as well as the achievable MB/s that they can deliver. The total number of IOPs and MB/s required for existing Oracle Databases can be determined from Automated Workload Repository (AWR) reports.

The total IOPs required is the total of the all the reported **physical read I/O requests** and **physical write I/O requests**, while the total MB/s required can be determined from the total of **physical reads total writes** and **physical read total reads** (When determining these figures, also allow some additional headroom for backup operations).

The following guidelines can be used as an approximate starting point for sizing new systems:

- For OLTP environments, assume every transaction incurs around 5 random IOPs.
- For Data Warehousing environments, assume that a reasonably modern core will need about 200 MB/s in order to be kept fully busy – therefore a 2 processor quad core machine will need around 1.6GB/s scan rate to be fully utilize.

Once the required IOPs or MB/s needed to support the workload have been determined, then the number of disks required in the storage array can be determined. Simply divide the required workload numbers by the specific IOP and MB/s figures obtainable per disk in the storage, and that will determine roughly how many disks are needed for maximum performance (Note that the figures above do not assume any disk mirroring or storage array caching etc, which should then also be taken into account). Work back from this initial disk sizing through the switches and HBAs that will be required to deliver this performance to the overall sizing of the storage required.

Note that disk vendors provide both IOPs and MB/s ratings for the storage arrays they provide. These numbers are often achieved under optimal conditions, and in the real world are a little inflated. Generally, Fibre Channel drives are capable today of delivering around 80-110 IOPs and 20-35 MB/s. SATA drives will provide around 50 IOPs and around 20-30 MB/sec. Oracle provides the ORION (Oracle I/O Calibration Tool), which can be used to test and accurately size storage arrays based on a synthetically created database workload.

Use Automatic Storage Management to maximize I/O bandwidth

Automatic Storage Management provides an efficient realization of storage array capabilities when deployed underneath Oracle Databases. By striping data across all the disks in a disk group, ASM ensures that the maximum possible I/O bandwidth is used for sequential scan operations, and that random I/O is also balanced across the array. With Oracle Database 11g Release 2, frequently accessed data can also be striped on the outside of the physical disks, with less frequently accessed data striped on the inside, ensuring frequently accessed data benefits from the faster transfer rates provided by the operational physics of the disks.

Customers that have moved their Oracle Databases from traditional storage arrays to those same environments enabled with ASM have experienced a 25% performance increase in their database, directly attributable to the better I/O balancing enabled by ASM.

Use Direct and Asynchronous I/O to optimize I/O performance

Oracle Database 11g also has the ability to bypass the operating system file cache, reducing the amount of processor resources and system memory needed for I/O operations. This memory can then be used to increase the size of the Oracle Database buffer cache (SGA), delivering better performance and a more consistent data caching capability. Where supported by the

underlying operating system, direct I/O should be generally used with Oracle Database 11g. In addition, Oracle Database 11g also supports asynchronous I/O, where writes performed by the operating system are done asynchronously from the database processes, reducing the potential for bottlenecks caused by I/O waits. Reads can also be done asynchronously as well, including a read-ahead capability for sequential scans, with often-dramatic performance improvements in query intensive workloads.

Use Direct NFS Client to improve Network Attached Storage performance

Many IT organizations use Network Attached Storage (NAS) appliances underneath their Oracle Database. NAS appliances typically use the NFS protocol as the communication layer between the server and the storage itself. Many of the NFS protocol stacks provided by the operating system are not optimized for the I/O performed by the Oracle Database. To this end, Oracle Database 11g provides a Direct NFS client capability that is built directly into the database kernel. This native capability enables direct I/O with the storage devices, bypassing the operating system file cache and reducing the need to copy data between the operating system and Database memory. Direct NFS client also enables asynchronous I/O on NFS appliances. For Data Warehousing workloads, these capabilities have been shown to increase performance by 40%. For OLTP environments, around a 10% performance improvement has been achieved.

Use Database Smart Flash Cache to reduce physical disk I/O

Oracle Database 11g Release 2 introduces a revolutionary new capability that can dramatically improve Oracle database performance on existing storage areas. New Flash PCI cards are available that can deliver gigabytes of space and support very high numbers of IOPs, often getting close to the I/O characteristics of memory, but at a much lower cost per gigabyte.

Traditionally these cards have been added to a server and used as storage for Oracle Database files such as redo logs. However, a better technique is to have Oracle Database repurpose these cards as a second level cache for data cached in server memory. With the Database Smart Flash Cache feature in Oracle Database 11g Release 2, as data ages out of the memory buffer cache, it is written back to the storage array for long-term data protection. However, at the same time, data blocks are also copied to available space on flash cards. The next time these data blocks are required in the database buffer cache, they are retrieved from the Smart Flash Cache, avoiding the need to perform any physical I/O from the storage array. Using the PCI Express interface also means that these I/O operations bypass any disk controller overhead.

Database Smart Flash Cache can result in a 25% performance improvement using only 1/10 of the existing storage array infrastructure. For storage arrays that are currently maxed out on I/O, the addition of Flash PCI Cards to the hosts, enabled with Database Smart Flash Cache, provides a low cost way to dramatically improve application performance.

Achieve extreme performance with Sun Oracle Exadata Storage Servers



While the above techniques all optimize the performance of existing storage arrays, the grim reality today is that available disk technology has failed to keep up with the performance gains made in core processing speeds. While processor speeds have doubled every 18

months for the last twenty years, disk performance has not increased at the same rate, leading to an imbalance between data processing capability, and the storage array's ability to keep processors busy.

To this end, Oracle, in conjunction with Sun, develops storage that is highly optimized for Oracle Database operations – the Sun Oracle Exadata Storage Server.

This is built on an industry-standard compute server that comes equipped with twelve 600 GB Serial Attached SCSI (SAS) disks that provide up to 7.2 TB of fully redundant, uncompressed user data storage; or twelve 2 TB Serial Advanced Technology Attachment (SATA) disks, that provide up to 24 TB of redundant, uncompressed user data storage.

Sun Oracle Exadata Storage Servers have two Intel Xeon E5540 quad-core processors, which run Exadata Storage Server Software. Oracle Database 11g Release 2 pushes SQL processing to these intelligent Storage Servers, where scans are executed, using all the disks in parallel, and returning only the relevant rows and columns to the database server, reducing database server CPU consumption, and pushing less data between the storage and database servers.

Each Sun Oracle Exadata Storage Server has an effective data scan rate of 1.5 GB/s for uncompressed data on raw SAS disk, rising to 3.6 GB/s for uncompressed data in flash storage. In addition, each Storage Server is enabled with 384 GB of Exadata Smart Flash Cache, providing intelligent data caching for OLTP operations, delivering a 10x increase for random IOPS. 40 Gigabit InfiniBand connections provide network performance that's much faster than traditional storage or server networks. The interconnect protocol uses direct data placement to ensure very low CPU overhead by directly moving data from the wire to database buffers with no extra data copies.

Sun Oracle Exadata Storage Servers are architected to scale-out easily. To achieve higher performance and greater storage capacity, additional Exadata Storage Servers are simply added to the configuration. This, combined with faster InfiniBand interconnect, Exadata Smart Flash Cache and the reduction of data transferred due to offload processing, yields very large performance improvements. A 10x improvement in query performance compared to traditional storage infrastructure is common, with much greater improvement possible, providing an ideal storage solution for extreme performance OLTP and Data Warehousing applications.

REDUCE OVERALL STORAGE COSTS

"Our Chief Financial Officer likes the Advanced Compression option of Oracle Database 11g because with it we won't need anywhere from a third to two thirds of the disc we have right now. So as we grow, we can actually recycle the disc that we're using right now that's going to be saved by compression."

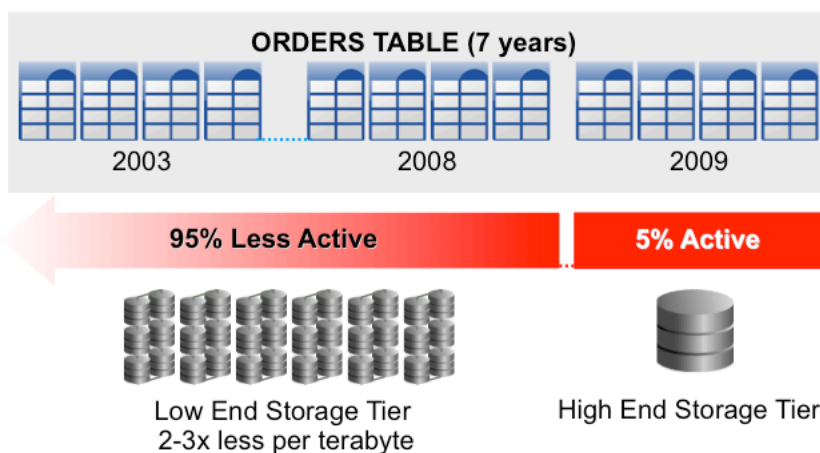
Mike Prince, Chief Technology Officer

Burlington Coat Factory

One of the biggest challenges around information management is reducing the need to provision more storage every year. While storage itself has become cheaper over the years in terms of cost per gigabyte, the cost of provisioning this storage continues to climb – new storage requires additional data center space, more cooling, and more power to run. However, Oracle Database 11g reduces the need to continually buy new storage as databases continue to grow.

Partitioning data across two storage tiers

One way to reduce storage costs involves an innovative use of partitioning underneath large databases. Many organizations today are implementing two tiers of storage – a high end storage array typically used for mission-critical applications, and a lower cost storage array used for less demanding environments. These lower cost storage arrays can often be 2-3 times less cost per terabyte than higher end storage. But, while this approach does save some storage costs, it is not optimal. Entire databases are mapped to either of the two storage tiers. As both mission critical and less mission critical databases continue to grow over time, new storage is eventually required across both tiers. The reality is that most databases contain a mixture of active data – e.g. orders being taken this month, customer call records for current billing cycle; and less active data – e.g. orders taken last year, and call records from prior billing cycles. Typically, active data represents less than 5% of information stored in the database – the other 95% is much less active data.



Example of partitioning used to reduce storage costs

Instead of mapping entire applications to different storage tiers, customers are starting to unlock the lifecycle of data by mapping data within those databases to different storage tiers. Large tables can be easily partitioned on a partition key – typically range partitioned on the key that represents a time component, with current “active” data placed onto high end storage, and as that data becomes less active, its moved via the partitioning capability to the lower cost storage tier. This type of “online archiving” of data means that further storage growth is constrained to the low cost storage tier. Indeed, most enterprise data growth tends to be in the amount of historic data that is maintained online for business intelligence and compliance reasons.

Partitioning data in this way will also improve the performance of the applications that access this data – any accesses to active data that are based on the same partition key (such as sales date), will automatically benefit from the partition pruning performed by the database. In other words, as database tables grow in size, performance to active data will not degrade. For many organizations, this removes the need to regularly archive or purge data to maintain the required performance of their database applications.

However, unlike an offline archive, the historic data is available at any time through the application, and is also maintained throughout database and application upgrades. By partitioning databases based on the lifecycle of the information managed with those databases, IT departments can reduce their dependency on high end storage, reduce their incremental storage costs, keep more data online for longer periods of time, and improve the performance of applications that access large databases.

Compress All Data

Advances in compression technology now mean that it is practical to compress all data within Oracle Databases. Compression can reduce existing storage usage by a factor of 2-4 times, effectively freeing up anywhere from 1/2 to 3/4 of existing storage resources while improving query performance accordingly. In addition, any reduced usage of storage will cascade throughout the data center, via copies made for backup and development and testing purposes. Compression is one of the major enablers of efficient storage management, and Oracle Database 11g provides many multiple techniques for effective data compression.

COMPRESSION TYPE:	SUITABLE FOR:
Basic Compression	Read only tables and partitions in Data Warehouse environments or "inactive" data partitions in OLTP environments.
OLTP Compression	Active tables and partitions in OLTP and Data Warehouse environments.
SecureFiles Compression	Non-relational data in OLTP and Data Warehouse environments.
Index Compression	Indexes on tables in OLTP and Data Warehouse environments.
Backup Compression	All environments.
Hybrid Columnar Compression – Data Warehousing	Read only tables and partitions in Data Warehouse environments.
Hybrid Columnar Compression – Archival	"Inactive" data partitions in OLTP and Data Warehousing environments.

Table showing different compression types supported with Oracle Database 11g.

Basic Compression

Basic compression, first made available with Oracle9i Database Enterprise Edition, is used to compress read-only data that may be found in a data warehouse, or in the non-active partitions of a large partitioned table underneath OLTP applications. The data is compressed as it is loaded into the table (or moved into the partition), with typically around a 2 to 4 times compression ratio achieved.

OLTP Compression

OLTP compression, introduced with the Advanced Compression Option of Oracle Database 11g, compresses active data for OLTP applications. OLTP compression uses a sophisticated algorithm that continuously compresses data as it is written during INSERT or UPDATE operations, while minimizing the overhead of compression operations. This makes it suitable to be used for the tables and partitions of active data involved in the most mission critical OLTP

environments. A 2-4 times compression ratio is typically achieved, and the range of data values compressed is not limited, meaning that all new data is compressed without loss of compression ratio over time.

SecureFiles Compression

The Advanced Compression option of Oracle Database 11g also supports compression of data stored as large objects (LOBs) in the database. This is especially important for environments that mix structured data with unstructured data underneath rich applications – for example, XML, spatial, audio, image or video information stored in Oracle Database 11g can also be compressed. Compression ratios are dependent on the type of media being compressed, however three levels of compression are provided to allow the desired level of compression to be achieved based on available CPU resources

Index Compression

Oracle Database 11g not only compresses row data, but also compresses the indexes associated with these rows. Indexes can be compressed independent of whether the underlying table data is compressed or not. Index compression can significantly reduce the storage associated with large indexes in large database environments.

Backup Compression

Oracle Database 11g can also compress backups of data stored in the rows and indexes in databases, independent of whether the data in these rows or indexes have in turn been compressed.

Hybrid Columnar Compression with the Sun Oracle Exadata Storage Servers

The compression capabilities covered above are enabled with Oracle Database 11g, and associated options, on all hardware and storage platforms. However, with the advent of Sun Oracle Exadata Storage Servers, an additional level of compression capability is provided, called Hybrid Columnar Compression. Hybrid Columnar Compression uses additional capabilities of intelligent Exadata Storage Servers to provide deep compression - anywhere from a 10 to 50 times ratio.

Typical database compression algorithms compress repeating values found within rows of data stored in a database. So for example, all the order dates with an order row will be compressed. However, as the rows are stored on disk in row format, there is a lot of non-relevant information stored between each occurrence of the next value of an order date – between each order date, additional values are found for order id, customer id, product ids, etc.

Hybrid Columnar Compression uses a different technique to store the column values. Instead of storing in a row format, the data is effectively stored by column – for instance, all the order dates

will be stored together, then all the order ids, then all the customer ids, etc. This means that within a unit of compression, a much higher rate of repeating values is found, and a greater compression ratio can be achieved. Two different compression ratios are then provided:

- Data Warehousing – ideal for compressing data that will be used for queries in a data-warehousing environment. Typically around a 10 times compression ratio is achieved, with a corresponding increase in query performance.
- Archiving – this uses a deep compression ratio, and can achieve up to 50 times compression ratios. It is best suited for data that is non-active, typically stored offline, but now delivers an online archive.

Both hybrid columnar compression techniques are best used with data that does not change frequently – the data is compressed on data load, and while updates and inserts are supported against the hybrid columnar compressed tables, the benefits of this type of compression will be lost for the rows impacted by ongoing updates or inserts.

PROTECT AGAINST DATA LOSS

"We utilize SAN arrays and we've got bandwidth, so we've got the ability to use solutions such as remote-mirroring, but for this critical database system, we went with Data Guard. Data consistency and data integrity were the main drivers."

David Willen, Chief Technology Officer

BarnesandNoble.com

Information is the life-blood of successful competitive organizations. Loss of that information for even a few hours or a few days can cost organizations vast sums of money, and can even put them out of business. As such, modern organizations need to protect all their information. Oracle Database 11g provides unique industry leading capabilities to protect against data loss.

Protect against data corruption

Modern disk systems are generally reliable. However, every now and then, things do go wrong, and data is not written successfully to disk, even though the disk subsystem and the operating system believes that it has been done so. These types of events can cause widespread data corruption in a database environment, and can be difficult to detect. Indeed, it may not be until the next access of that data that any inconsistency is noticed, allowing for the problem to spread through the entire database, and possibly through backups and disaster recovery sites as well.

Oracle Database has comprehensive built-in checks to detect and repair data corruptions. Using a single parameter set to a desired protection level, the Oracle Database can detect corruptions in data and redo blocks using checksum validation, detect data block corruptions using semantic checks, and detect writes acknowledged, but actually lost by the I/O subsystem. Specific technologies also provide additional validation – e.g. Recovery Manager (RMAN) can be used to validate data blocks while doing backup and recovery, ASM can be used to recover corrupted blocks on a disk by using the valid blocks available on the mirrored disks and Data Guard can be used to ensure that the standby database is isolated from all data corruptions at the primary database.

The Sun Oracle Exadata Storage Server also prevents corruptions from being written to disk, by incorporating the Hardware Assisted Resilient Data (HARD) technology in its software. HARD uses block checking where the storage subsystem validates the Oracle block contents, which prevents corrupted data from being written to disk. HARD checks with Exadata operate completely transparently and no parameters need to be set for this purpose at the database or storage tier.

Protect against data loss with ASM Mirroring

Many IT Professionals use disk-mirroring techniques such as RAID 5 or RAID 10 to keep multiple copies of data on different disks in their storage arrays, protecting against data loss

through an individual disk failure. Oracle Database 11g works very well with all major disk-mirroring capabilities. However, these disk-mirroring capabilities can often add additional expense to the storage environment being used.

Automatic Storage Management (ASM) can be used to not only stripe data across multiple disks in a storage array, but to also mirror that data across disks as well. ASM's mirroring capability can be set up when disk groups are created. A disk group is divided into failure groups, and each disk is placed in one collection of disks that can become unavailable due to failure of one of its associated components, such as:

- Storage array controllers
- Host bus adapters (HBAs)
- Fibre Channel (FC) switches
- Disks
- Entire arrays

Redundancy for disk groups can be either normal (the default), where files are two-way mirrored (requiring at least two failure groups), or high, which provides a higher degree of protection using three-way mirroring (requiring at least three failure groups)

ASM uses a unique mirroring algorithm that mirrors extents. When ASM allocates a primary extent of a file to one disk in a failure group, it allocates a mirror copy of that extent to another disk in another failure group, ensuring that a primary extent and its mirror copy never reside in the same failure group. Unlike other volume managers, ASM has no concept of a primary disk or a mirrored disk, a disk group only requires spare capacity; a hot spare disk is unnecessary.

When a block is read from disk, it is always read from the primary extent, unless the primary extent cannot be read, in which it will read the secondary extent. When a block is to be written to a file, each extent in the extent set is written in parallel.

In the event of a disk failure in failure group, which will induce a rebalance, the contents (data extents) of the failed disk are reconstructed using the redundant copies of the extents from partner disks. If the database instance needs to access an extent whose primary extent was on the failed disk, then the database will read the mirror copy from the appropriate disk. After the rebalance is complete and the disk contents are fully reconstructed, the database instance returns to reading primary copies only

Efficiently backup and restore data

While storage mirroring provides an important element of data protection, all databases should be regularly backed up. However, as databases grow in size, new optimized techniques are required to constrain both the time required to backup the database, and the time required to restore and recover (if necessary).

Traditionally, backup has been to streaming devices such as tapes. Oracle Database 11g supports many tape backup and vaulting environments through the Oracle Backup Solutions Program, a cooperative program designed to facilitate tighter integration between Oracle's backup products and those of third-party media management vendors.

Oracle Secure Backup

In addition to supporting third party backup products, Oracle also provides its own tape backup solution, called Oracle Secure Backup. This is a centralized tape backup management solution providing high-performance data protection in UNIX, Linux, Windows and Network Attached Storage (NAS) environments.

Oracle Secure Backup provides a complete tape backup solution for protecting both file system and Oracle database data, and is fully integrated with the Oracle Database backup utility, Recovery Manager (RMAN). This tight integration means that Oracle Secure Backup can provide optimized tape backup for the Oracle Database, by backing up only currently used blocks and eliminating backup of committed undo – both of which help increase backup performance by 25 – 40% over comparable products. In addition, Oracle Secure Backup offers backup encryption and key management, ensuring that any confidential information stored on backups tapes sent offsite are protected.

Fast Recovery Areas and Incremental Backups

One of the challenges of a tape based backup strategy is that writing backups to tape and subsequent restores can take a long time. More and more organizations are utilizing low-cost disks as their preferred medium for Oracle Database backups. One of the advantages of this approach is that random I/O can be performed on the backup images stored on disks. Oracle Database 11g takes full advantage of this capability with revolutionary new disk based backup and recovery technologies.

Starting with Oracle Database 10g, administrators have been able to define a disk-based Fast Recovery Area for Oracle Databases. This is a group of disks, typically separate from the storage array used for the database environment that is dedicated to hold database backup images. The disks themselves are fully managed by Oracle's Recovery Manager utility (RMAN), and can also take advantage of ASM for backup striping etc. Once a Fast Recovery Area is setup, then the Oracle Databases will, by default, back themselves up automatically to these areas during pre-defined backup windows. More frequent backup periods can be defined, and the timing and length of the backup window can be defined. If additional space in the Fast Recovery Area needs to be reclaimed for new backups, RMAN will automatically delete files that are obsolete or have already been backed up to tape.

For large databases, backing up the entire database every 24 hours may take a long time. So many IT organizations enable incremental change tracking on their Oracle Databases. During the 24-

hour period since the last backup, a record is kept of which data blocks in the Oracle database have been changed during OLTP operations. Then during the backup window, only the changed blocks are stored in the Fast Recovery Area, as an incremental backup against a complete backup image. This means that the backup window itself is only ever a function of the number of changes made within a 24-hour period, independent of the size of the actual database.

However, having too many incremental backups can extend the time taken to subsequently restore and recover the database if required. To this end, RMAN incrementally updates the complete database backup image with the incremental changes from the previous 24-hour period, eliminating the need to apply multiple incremental backups on recovery. This reduces overall recovery time, in addition to reducing the need to take full backups.

Backup images in the Fast Recovery Area can be in turn backed up to tape using Oracle Secure Backup and other tape backup solutions. Such backups can also be compressed to further save time and space needed.

Read only Tables and Tablespaces

Read Only Tables and Tablespaces can be further used to reduce backup and restore times. Large tables that have been partitioned using a life cycle management strategy described in the section on lowering storage costs can have their older data partitions put into read only format by the Oracle Database, meaning that transactions cannot change the data stored within these partitions. These partitions can then also be compressed, and placed on low-cost storage tiers. As they are read-only, RMAN knows that they do not need to be backed up beyond the initial backup; further reducing the requirements for both backup and restore operations.

Using low-cost disk for the Fast Recovery Area, and an incremental backup strategy provides an alternative to expensive snapshotting technology deployed at the storage array level. The Sun Oracle Exadata Storage Servers make use of this technique, and come pre-configured with a Fast Recovery Area for the databases that are built on them.

Protect against data loss caused by human error

It's an unfortunate truism, but most data loss is actually caused by human error, and not the failure of a disk or a storage array. Database Administrators (DBAs) make mistakes – they log onto development databases to clean up tables and indexes, only to find that they had logged onto the production environment by mistake, and had dropped critical tables and indexes that were in use at the time. Or a DBA will quickly update a customer record based on a service request, only to later discover that they had inadvertently given many customers the same telephone number.

When these problems occur, the traditional approach has been to revert to a backup. The production system is stopped, and a point in time recovery is performed to restore the data back to the point just before the error occurred. There are many problems with this approach – firstly,

additional storage is required, secondly, restore and recovery can take a long time, during which time the production system is unavailable, and thirdly, any 'good' transactions made from the time the error occurred until it was later discovered are lost, resulting in the loss of good data.

To overcome this type of problem, Oracle Database 11g provides unique flashback capabilities. Flashback allows operations that were inadvertently performed online to be undone online. For example, if a DBA does drop a table or an index, the table and index are marked as unavailable in the data dictionary, and are no longer used by the application. However, the actual data extents that were present in the table or index are kept on disk. If at a later stage the DBA identifies that the table or index was dropped inadvertently, they can simply undo the drop operation and the table or index is immediately made available again to the application.

Similar steps can be taken if a transaction invalidly changes one or more rows in a table. Flashback query operations allow the DBA to see the earlier versions of the rows, and also identify the transactions that caused the error. Then the offending transactions can be flashed back online, meaning that all changes caused by the transactions are immediately undone. Alternatively, all changes made to one or more tables since a specified period of time can be undone, also online.

If the entire database has become logically corrupt due to a large number changes, then the entire database can be flashed back in time, rather like running a video in reverse order. While flashing back the database is not an online operation, it is far quicker to unwind the database this way than restoring a database and doing a point in time recovery. In addition, the flashback database operation can be performed multiple times, allowing incremental flashback and roll forward to the exact moment in time required.

Flashback transaction and flashback table operations use the existing undo information that is collected by the Oracle Database during normal operations. Most Oracle DBAs will size the collection of this undo information to give themselves a 24-hour window in which to catch and undo any human error. The ability to flashback the entire database requires additional information to be collected. This additional information is managed as part of the Fast Recovery Area, and is also typically sized to provide a 24-hour window.

These type of flashback capabilities supplant the need for additional snapshotting at the storage level, further reducing storage costs, and provide a much finer granularity of operation, with the ability to undo any individual transaction to any point in time.

Protect against data loss through disaster

The above techniques protect against data loss within the data center. However, the data center itself can be lost through fire, earthquake, other disaster, or even something as mundane as power loss. To fully protect themselves against data loss, many IT organizations are also investing in secondary data centers that house standby databases, which are then synchronized with the changes being made in the production environment.

The traditional method of doing this synchronization is to use expensive remotely mirrored storage solutions. There are however a couple of problems with this method:

- Firstly, the remotely mirrored storage solution is often expensive.
- Secondly, the storage replicates every write performed on the production system to the standby system. This means that expensive, high-bandwidth networks are required between the data centers, incurring additional cost and also limiting the supportable distances between the data centers.
- Thirdly, the remotely mirrored standby solution is idle, in that it is only useful in the event of the loss of the production data center. This decreases the value of the investment made in the standby environment, and also has an unintended consequence that administrators are often loathe to failover to the standby environment, as they are not familiar with operating it on a day-to-day basis.
- Fourthly, remote mirrored storage introduces another layer in the infrastructure where problems can be introduced, especially around propagation of data corruption between the production and standby sites.

The Oracle Database works well with remotely mirrored storage solutions. But Oracle also provides a built in standby solution, called Oracle Data Guard, which addresses the problems identified above with remote mirrored storage.

Oracle Data Guard only transmits the writes to the redo logs from the production databases to the standby databases. This can mean a 1/7th reduction on the network of the volume, and up to 1/27th of I/O operations allowing a much smaller capacity network to be utilized between the data centers, and can also prevent the need to move to expensive network solutions.

In addition, network latency is less of an issue, allowing synchronous operations to be used across the network. Independent tests have shown that a network with a round trip time of 10-15 milliseconds can cause up to a 25 to 40% latency impact on the primary database when using remotely mirrored storage with synchronous writes; with Oracle Data Guard, the measured latency impact on the primary database was around 4%. This means that Oracle Data Guard can provide greater data protection across longer distances between data centers.

Oracle Data Guard also ensures much better data protection and data resilience than remote mirroring solutions, since corruptions introduced on the production database are more likely to be propagated by remote mirroring solutions to the standby site, but are intelligently eliminated by Data Guard.

Businesses also have to ensure that they are getting as much value as possible from their IT investments. Remotely mirrored storage systems basically sit idle until a disaster happens. Oracle Data Guard has been architected to allow businesses to get useful work from their investment in the standby site. Using the Oracle Active Data Guard option in Oracle Database 11g) the standby databases can be open read-only for reporting and queries while changes are being

applied from the production system, allowing IT organizations to offload resource intensive reporting operations to their standby databases, improving the performance of their production database. Oracle Active Data Guard is well-integrated with Recovery Manager (RMAN), allowing fast incremental backups to be off-loaded to the physical standby database, saving critical system resources at the primary site and enabling efficient utilization of system resources on the standby site.

Oracle Data Guard offers an integrated switchover capability that can be used to address planned maintenance, e.g. hardware or OS upgrades, minimizing the cost of downtime associated with such scheduled outages.

Remote Mirroring solutions conceptually appear to offer simple and complete data protection. However, for data resident in Oracle databases, Oracle Data Guard, with its in-built zero data loss capability, is more efficient, less expensive and better optimized for data protection and disaster recovery than traditional remote mirroring solutions.

CONCLUSION

Many IT organizations today use inefficient storage management techniques underneath their Oracle Databases that are based on traditional practices that are no longer the industry's best operating procedure. By re-evaluating and using many of the advanced capabilities provided for storage optimization by the Oracle Database, IT Professionals can reduce their overall storage costs by a factor of 10x. And, optimize their storage management strategy to deliver the performance and availability that today's modern enterprises require.



Optimizing and Protecting Storage
with Oracle Database 11g Release 2

Nov 2009

Author: Mark Townsend

Contributing Authors: Willie Hardie, Ron Weiss,
Ara Shakian, Ashish Ray, Kevin Jernigan, Nitin
Vengurlekar

Oracle Corporation
World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.

Worldwide Inquiries:
Phone: +1.650.506.7000
Fax: +1.650.506.7200
oracle.com



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

Copyright © 2009, Oracle and/or its affiliates. All rights reserved. This document is provided for information purposes only and the contents hereof are subject to change without notice. This document is not warranted to be error-free, nor subject to any other warranties or conditions, whether expressed orally or implied in law, including implied warranties and conditions of merchantability or fitness for a particular purpose. We specifically disclaim any liability with respect to this document and no contractual obligations are formed either directly or indirectly by this document. This document may not be reproduced or transmitted in any form or by any means, electronic or mechanical, for any purpose, without our prior written permission.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates. Other names may be trademarks of their respective owners.

0109