WHITE PAPER

Controlling Storage Costs with Oracle Database 11g

By Brian Babineau
With Bill Lundell

February, 2008

Table of Contents

Executive Summary

Recent ESG research indicates that the average database is expected to grow 25% annually. More resources are needed to retain and manage corporate data. Expenses increase due to poor resource utilization—most organizations think 50% storage utilization is acceptable—as companies buy more systems instead of leveraging existing systems more efficiently.

One answer to this information explosion is Information Lifecycle Management (ILM), which can help organizations align the value of information with its underlying IT infrastructure by balancing infrastructure costs with the accessibility and availability requirements of corporate data. ILM can help curb storage costs by utilizing a variety of storage tiers with different performance, reliability and cost characteristics. As data ages and becomes less active, it can be moved to a more appropriate storage tier where it will remain online and accessible. When combined with storage optimization solutions such as compression and data de-duplication, ILM can lead to measurable cost savings.

The ILM capabilities of Oracle Database 11g complement a tiered storage strategy. Features such as the ILM Assistant, Partitioning and Advanced Compression provide a solution that will relocate data to an optimal storage medium based on age and business value. One of the key compression enhancements in Oracle Database 11g is that it works with regular data manipulation operations, such as 'INSERTs' and 'UPDATEs.' As a result, data is compressed on an ongoing basis—as opposed to only during bulk load operations, as was the case previously. Oracle Database 11g also includes performance optimizations that minimize compression overhead, making it ideal for both transaction processing and data warehousing workloads.

In ESG's opinion, an 85% reduction in storage-related costs is possible thanks to the ILM capabilities in the Oracle database. Customers can also save in terms of power, cooling and floor space reductions. Tiered storage and compression are clearly crucial factors to consider when implementing successful ILM strategies. Oracle Database 11g allows users to extract maximum financial value from their data.

Introduction

The explosion of digital content can be traced to several factors. Organizations have embraced computing as an integral contributor to their hopes for success. Whether it's just a few critical applications or an entire networked infrastructure, everything from communications to business documents has gone digital.

Exponential data growth can also be attributed to the plethora of government and industry regulations regarding extensive data retention policies facing organizations today. In the past, data that had exceeded its useful life would be erased or retained—often in a manner that made it difficult to be found and retrieved. Now, this information is being retained indefinitely and IT must be able to locate it within a reasonable period of time.

Clearly, there are cost implications associated with generating more information and keeping it longer. If you're creating information at a high rate and retaining the majority of it, you're going to need additional storage and management resources. However, increased retention comes with its own benefits. Organizations can now leverage extensive pools of information, which may translate to a competitive edge and expanded revenue opportunities.

Organizations must also factor in the operational impact that increased retention presents to IT departments. The proliferation of new hardware will take up valuable data center real estate and consume additional power and cooling resources.

While it may be hard for organizations to avoid incremental storage capacity purchases, there are opportunities currently available that can make these investments more cost effective. One such approach is deploying an Information Lifecycle Management (ILM) strategy that enables organizations to manage information according to its business value using different storage tiers. With a tiered storage implementation in place, ILM solutions can help identify a piece of data's relative business value and then automatically move it to the most appropriate storage tier. ILM solutions can also optimize these storage tiers through compression technology. Together, compression and storage tiering helps organizations store information more cost effectively.

Handling Information Growth

Digitization has increased the value of information—data is easier to retrieve, share and analyze. New data is constantly being created as organizations look to capitalize on these benefits and extract as much value as possible from their information. A large portion of this information, especially critical business data, is managed by databases and continues to grow at a healthy pace. **According to ESG research[1], 85% of respondents expect to see double digit annual growth among their primary database instances.** This data represents information captured as part of everyday activities such as booking an airline ticket or capturing a new customer's contact information.

In database environments, IT often creates multiple secondary instances that are full copies or subsets of the primary data set—for a variety of reasons. Secondary instances are used in test and development environments, data warehouses and reporting operations. These additional instances further propagate the amount of database information that must be stored and managed.

Database information growth is also spurred by data protection and disaster recovery processes. Copies of primary data must be made in order to prevent any information from being lost in the event of corruption or accidental deletion. While helping to foster higher data availability, these solutions also require storage capacity—across a variety of media formats such as tape or optical systems.

[1] Source: ESG Research Report: *2007 Database Archiving Survey*, December 2007

IT must also deal with regulatory, corporate governance and legal mandates that require data to be retained for specified periods of time. From a regulatory standpoint, organizations must keep certain business records—many of which reside in databases. In legal scenarios, organizations must preserve potential evidence and make it readily accessible to corporate counsel when necessary. These requirements often lead to copies of database records being moved to lower cost disk systems or tape media. Approximately 60% of organizations that have been through an electronic discovery have had to produce database records and as a result, storing this data requires additional resources that must be segregated for preservation and security purposes.

Understanding the Cost of Storage

Whether it is creating secondary database instances to load a data warehouse or copying data for litigation support purposes, organizations must acquire additional storage capacity to handle database information growth. While many IT departments concern themselves with the storage system's acquisition cost and associated management or protection software, there are several other expenses that must be considered when attempting to understand the complete cost of data retention.

Capital Expenses

When it comes to addressing information growth, organizations respond by acquiring more and more of the same types of storage capacity. On the surface, this may appear to be the best approach because database administrators and storage managers want to avoid any 'surprises.' However, using the same storage system configurations with the same performance and availability capabilities for the majority of database information, regardless of type and business value, is inefficient at best.

Tertiary storage devices are needed to back up information and in some cases, to archive data that must be kept for compliance purposes. Until recently, organizations relied on tape systems as backup targets and shipped the media offsite for disaster recovery purposes. **Due to database data growth and its importance to the organization, disk systems are being used for backup and business continuity operations more frequently.** Information may be replicated between two systems for ultra-available configurations, while denser storage systems are deployed to enable more accessible backups. As information grows, so does the number of tertiary systems required to retain and protect it.

Organizations must also buy storage networking components, including host bus adaptors (HBAs) and switches when capacity requirements expand. Some organizations may still believe that server-attached or direct attached storage meets their needs, but the benefits of centralized backups and management make networked storage a much better choice. Additionally, backup and replication software solutions must also be factored into the purchasing decisions, especially those that are licensed based on storage capacity.

Managing Database Storage

While it is true that the cost of storage hardware is decreasing, IT must still manage all underlying infrastructure. Storage management tasks go well beyond configuring and maintaining systems—IT must find ways to optimize storage system performance while guaranteeing availability, ensuring that information is always accessible to the appropriate business users. **ESG research[2] has shown that some organizations spend up to 30% of their storage budgets on labor.**

Often Overlooked Operational Expenses

Storage devices require appropriate power supplies, as do data center cooling systems. Currently, power is an expense of increasing significance to data center operations. **While data center real estate is certainly valuable, many facilities simply do not have access to enough power resources to run all of the**

[2] ESG research estimates and may vary by organization. Source: Enterprise Strategy Group, 2007

hardware. The density of both servers and storage systems is also creating power issues as these configurations require incremental cooling, which drains even more electricity.

Organizations may be content to solve the information growth problem by simply buying more hardware; however, this is becoming less feasible due to power issues and a lack of data center floor space. This vicious cycle does not seem to be subsiding—given the power constraints in many major metropolitan areas—and should lead organizations to implement solutions that will help them better utilize the capacity they already have.

Developing an Information Lifecycle Management Strategy

Creating a Multi-Tier IT Infrastructure

ILM can be deployed via tiered storage architectures to help control costs. As organizations gain a better grasp on the costs of storing and managing more information, they will soon realize that buying more of the same types of hardware is not the answer. Information Lifecycle Management (ILM) is a strategy, not a product. It assists organizations in understanding the value of information and aligning this value with the appropriate IT infrastructure. In order to achieve this goal, organizations must deploy a varied IT infrastructure—each tier having its own cost, performance and reliability metrics, in addition to a way of identifying and moving the information across these different tiers.

One of the simplest facets of an IT infrastructure to apply a tiered strategy to is storage because of the various system configurations that organizations can choose to deploy. There are also different ways to connect to a storage system, as well as various types of storage media—including several disk drive options that provide different levels of availability and performance. Further, there have been advancements in optimizing storage system power utilization by spinning down disk drives when they are not regularly accessed (some refer to these systems as MAID, which stands for Massive Array of Independent Disks). Several storage systems support multiple drive types within the same device, enabling an organization to build different tiers without having to buy and manage multiple boxes. These configuration choices result in different price points and cost benefits, allowing organizations to make investments with a variety of criteria in mind.

Permutations can vary; however, there are common configurations that organizations use to build different tiers. For example, higher performing disk drives (that also cost more) are commonly used for mission critical applications. Denser, less expensive drives that do not provide the same level of application response time can be used for backups or information being kept for compliance reasons. Organizations may deploy various drive types in a single system or in separate systems. Additionally, MAID or other power-efficient solutions can be used to store less active data, helping control usage.

Most organizations deploy separate tape storage, which is one of the least expensive storage media formats in terms of acquisition cost. Tape is portable and allows organizations to meet disaster recovery and long-term archive requirements, but information stored on tape is not immediately accessible.

The value of information changes over time, making it prohibitively expensive for organizations to keep all of their data on primary, high performing and more costly disk drives. Organizations can choose to move information to less expensive disk drives so that it remains available, albeit at a lower cost. Tiered storage helps organizations balance information access and availability requirements with the cost of storage. These are critical considerations that can help wrangle in IT departments in constant buying mode—usually the more expensive configurations.

Efficiently Reducing Data

A tiered IT infrastructure is just one tactical step towards Information Lifecycle Management. With information, both relational and structured data, traditionally stored in databases and unstructured information—files, web

pages, etc.—growing exponentially, organizations must find ways to use their IT resources more effectively. Eliminating duplicate data and applying compression across the entire IT environment can have a profound impact on capital and operating budgets.

De-duplication and compression technology make transparent data volume reduction (without deletion) a possibility. Data de-duplication ensures that only one copy of a database record or file is stored, backed up or copied for disaster recovery purposes. Compression decreases the size of a database or file, further reducing the amount of information to be stored or managed. In addition to reducing the amount of disk space required to store data, compression technologies also improve disk I/O performance—resulting in more efficient use of memory and maximizing network bandwidth. Compressing data within the primary database magnifies storage savings as it reduces the size of data volumes copied for backup, disaster recovery and test and development purposes. Less data means fewer systems and people are required to run the infrastructure.

Oracle Database 11g Optimizes ILM Storage Strategies

Leveraging Tiered Storage with Oracle Database Partitioning

ILM strategies also need solutions that can move and manage information. When it comes to managing databases, organizations must ensure data access and integrity.

Organizations need software that can automate the process of migrating information to the most appropriate storage tier. Oracle Database 11g facilitates database partitioning, which enables customers to store a variety of data types across multiple storage tiers based on defined business rules. This allows customers to reduce database storage costs by moving older information to less expensive disk drives where performance is less of an issue. The level of granularity ranges widely, with administrators having the flexibility to partition based on time intervals of their choosing and further refining their criteria through virtual columns and composite partitions. The Oracle ILM Assistant enables customers to establish ILM requirements for data placement on the most cost effective storage tier.

Since information movement is performed at the database level, it is transparent to the application. Therefore, application owners or users do not have to make any configuration changes to maintain data access. All data movement is done non-disruptively. Additionally, organizations keep all database information online—and at a lower cost. Customers can query the data with any tool and, by using Virtual Private Databases (VPD), these operations are only executed against data that the user is authorized to see.

Oracle Database 11g, with its ILM Assistant, facilitates cost effective information access without adding any complexity to the application environment. The Oracle ILM Assistant is a tool for guiding administrators through the process of implementing and managing an ILM environment. It provides the ability to create lifecycle definitions; advises on when to move, archive or delete data and illustrates associated storage requirements and potential cost savings.

Improving System Utilization with Compression

In addition to new capabilities that allow customers to take advantage of a tiered storage infrastructure, Oracle Database 11g improves table compression capabilities from previous generations, which were geared more toward data warehouse environments. Advanced Compression in Oracle Database 11g targets transactional applications by providing compression during data manipulation operations such as update and insert operations. Using an algorithm specifically designed for relational data, duplicate values can be eliminated across both database blocks and multiple columns.

The compression technology uses symbol tables, which store a single instance of a duplicate value as a form of metadata. Additional copies are replaced with pointers that reference the appropriate symbol table entry. The

blocks can be read in their compressed format with no impact. In addition, query performance can be significantly improved because queries on a compressed table need to access fewer blocks to retrieve the same amount of data when compared to an uncompressed table. Tests conducted by Oracle—whose results were presented at Oracle OpenWorld 2007—show compression speeding full table scan operations by over 2.5x with a corresponding 3.5x reduction in the number disk I/Os. The same tests showed an almost negligible (less than 3%) overhead—especially considering all the benefits compression provides—on write operations. While ESG has not validated these claims, customers should nonetheless expect performance improvements. Another advantage of reading data in a compressed format is that it increases memory efficiency as Oracle Databases can now fit more data in a given amount of cache memory.

Advanced Compression in Oracle Database 11g also helps control the explosion of unstructured information. ESG estimates that unstructured information consumes 70-80% of storage within a given organization, increasing the need for solutions such as SecureFiles. A new technology introduced in Oracle Database 11g, SecureFiles stores unstructured information in a database where files can be compressed using standard algorithms. To further minimize unstructured information storage costs, customers can employ SecureFiles's de-duplication functionality to ensure that only one copy of a file is stored, replacing any other occurrences with pointers (to the single copy). In addition to reducing storage costs, SecureFiless's de-duplication functionality can also significantly speed up the performance of write and copy operations.

Oracle's Advanced Compression helps minimize the impact of database backups on the storage infrastructure as it is seamlessly integrated with Oracle Recovery Manager (RMAN), Oracle's backup and recovery utility, and compresses backup data as it is written to disk or tape. With typical storage capacity reductions in the range of 50-75%, Recovery Manager with Oracle 11g Advanced Compression can help organizations experience significant savings in their IT budgets while ensuring that backups are completed quicker.

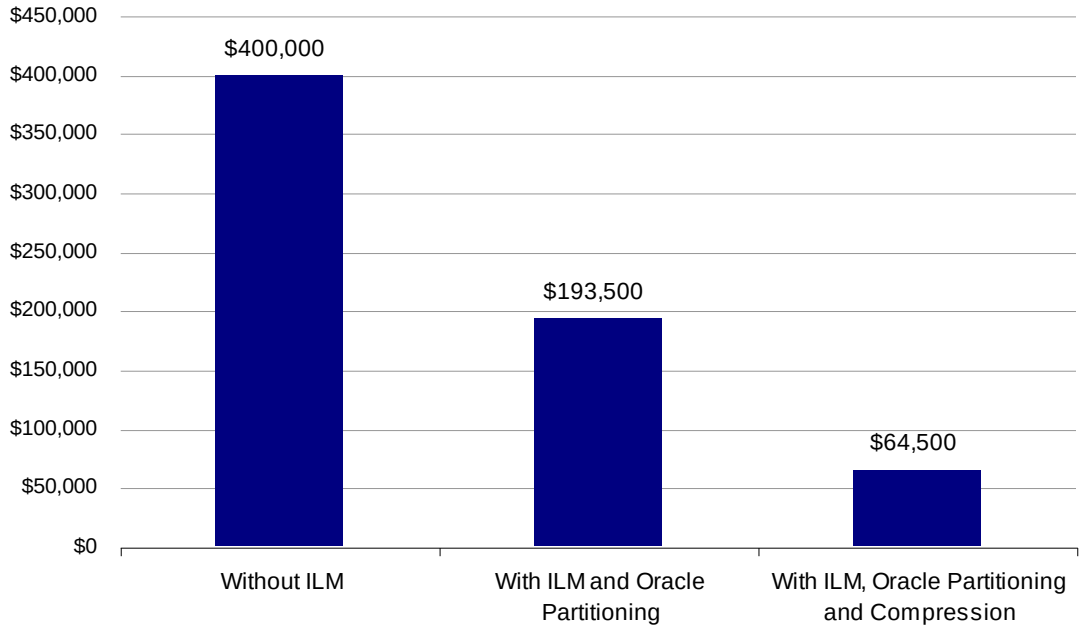
Saving Money with Oracle Database 11g

The economic impact—as it relates to storage expenses—of Oracle Database 11g's ILM features becomes even more obvious when you compare the total cost of storage for pre-ILM and post-ILM strategies. In this scenario, 10 TB will serve as the amount of capacity. ESG market analysis indicates that the average cost on a dollar per gigabyte ($/GB) basis—including acquisition and management costs (power and cooling, as well as floor space costs, are additional costs that can be considered but were not used in this example)—is: $40 for primary storage, $25 for secondary storage and $12 for dense, archival storage. Customers can utilize their own respective storage cost metrics when performing a similar analysis. Primary storage refers to high performance Fibre Channel drives, secondary storage is slower, higher capacity Fibre Channel drives and archival storage refers to high capacity, lower performing ATA drives.

In the scenario, one option is to keep all database records on primary storage, regardless of access activity levels, for a one year period. Apply the aforementioned price per GB for primary storage and the total cost of database storage is $400,000. This makes sense if your databases are exclusively supporting applications that perform frequent data inserts, refreshes or deletes.

Another method to reduce storage costs is to assume that 10% of the 10 TB is active, 35% is less active and the remaining 55% is historical. One TB of data would remain on primary storage to accommodate frequent changes. Organizations may vary in the way that they handle the other 90% of the data, but typically less active and historical data is moved to secondary drives for the next 6 months and then to denser drives for the rest of the year (and subsequent years, if it needs to be retained for an extensive period of time.). As a result, there would be 3.5 TB of less active data and 5.5 TB of inactive or historical data in a system. Applying the appropriate price per GB costs, this would translate to $40,000 for primary storage, $87,500 for secondary storage and $66,000 for historical storage—for a grand total of $193,500. This represents a cost savings of $206,500 over a one-year period, which is over a 50% reduction in saving all data on the primary tier. These savings are depicted in Figure 2.

FIGURE 1. IMPACT OF ORACLE'S ILM ON THE COST OF STORAGE IN ONE YEAR



Based on10 TB of storage, All per GB storage costs are based on ESG research estimates and may vary by organization.
Source: Enterprise Strategy Group, 2007

The potential savings can be greater if customers choose to use Oracle's Advanced Compression. This may allow for a 3x (or greater) reduction in the amount of data stored, regardless of the underlying hardware. The 10 TB that originally cost $400,000 to maintain would now cost just over $60,000—a difference of more than $335,000 and almost an 85% reduction in storage system expenses.

Conclusion

Organizations must realize just how inefficient current database storage management practices are—namely, adding more of the same hardware to accommodate data growth and dealing with duplication as well as uncompressed information. Simple steps such as tiering storage, moving data across these tiers and reducing the amount of data to be managed, can dramatically reduce capital and operating expenses. Organizations would be remiss if they did not, at a minimum, evaluate solutions such as Oracle Database11g, which can help make these savings a reality without requiring any application changes.

20 Asylum Street
Milford, MA 01757
Tel: 508-482-0188
Fax: 508-482-0218

www.enterprisestrategygroup.com