



Technical report: Oracle 10g and IBM System Storage N series

Best practices information lifecycle management

• • • • • • • • •

Document NS3556-0

April 11, 2008



Table of contents

Abstract	3
Introduction	3
Regulatory requirements.....	3
What is ILM?	3
Oracle 10g database for Business ILM.....	4
Why IBM N series for Oracle ILM	4
Implementing ILM in four easy steps	5
STEP 1 – Define the data classes	5
STEP 2 – Create storage tiers for the data classes	5
Assigning classes to storage tiers.....	7
STEP 3 – Create data access and migration policies	7
Managing access to data	7
Migrate data between classes.....	7
Step 4 – Define and enforce compliance policies.....	7
Conclusion	9
Trademarks and special notices	10



Abstract

IBM System Storage N series delivers storage solution efficiency and ease in an Oracle environment. Organizations wishing to take advantage of best-practice information lifecycle management can use IBM N series as a way of achieving compliance and new flexibility for Oracle data storage.

Introduction

Although most organizations have long regarded their stores of data as one of their most valuable corporate assets, how this data has been managed and maintained varies enormously, including depending on data use. Originally, data was used and stored to help achieve operational goals, run the business, and help identify future company direction and success. For the most part, data was retained as wanted and for as long as wanted. However, new government regulations and guidelines have been an additional key driving force in how and why data is retained. Some rules require organizations to retain and control information for very long periods of time. Therefore, IT managers today are trying to satisfy a broader, more complex range of data storage objectives: to meet the new regulatory requirements for data retention and protection; to store vast quantities of data; and to do so at lowest possible cost.

Regulatory requirements

Today, many organizations must retain specific data for specified periods of time. Failure to comply with retention rules could result in very heavy fines. Therefore, around the world, a number of regulatory requirements—such as Sarbanes-Oxley, HIPAA, and DOD5015.2-STD in the US and the European Data Privacy Directive in the European Union—are changing how organizations store and manage their data. These regulations specify what data must be retained, whether it can be changed, and for how long it must be retained, which could be for a period of 30 years or even longer.

What is ILM?

Information today comes in a wide variety of types, for example it could be an email message, a photograph, or an order in an online transaction processing system. Once the type of data and how it will be used is known, an understanding of data evolution and final destiny is likely to be known.

Increasingly, now, the challenge before all organizations is to understand how their data evolves, determine how it grows, monitor how its usage change over time, and decide how long it should survive, still while adhering to all the rules and regulations that already apply to that data.

Information Lifecycle Management (ILM) is designed to address these issues, with a combination of processes, policies, software, and hardware so that the appropriate technology can be used for each phase of the lifecycle of the data.

An analysis of your data will most likely reveal that initially it is accessed and maybe updated on a very frequent basis. As the age of the data increases, its access frequency diminishes to almost negligible, if any. Therefore, most organizations find themselves in the situation where most of their users are accessing all of the current data, and very few users are accessing all of the other data. Thus, data can be described as being active, less active, historical, or ready to be archived or purged, as illustrated in Figure 1.

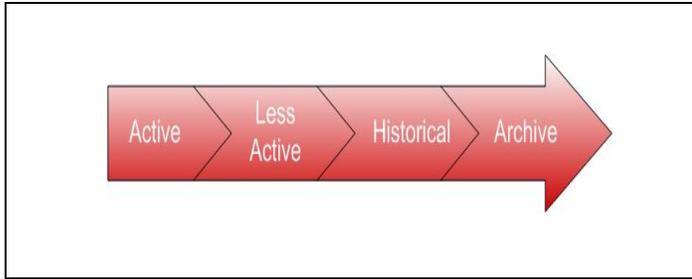


Figure 1)- Data lifecycle.

With so much data being retained, during its lifetime, the data will need to be moved to different physical locations, so it is located on the most appropriate and cost-effective storage device.

Oracle 10g database for Business ILM

Oracle 10g is capable of storing many different types of data, and storing all data in an Oracle database, making data much easier to manage because all data is in one place instead of being stored in many different places. The Oracle database is the ideal platform to implement an ILM policy because it has a number of features that make it very easy to implement an ILM solution:

- Application transparency: Data classification is transparent
- Fine-grained: Data management is at individual row level
- Low-cost: uses low cost to reduce costs
- Enforceable compliance policies: Define and enforce policies.

Why IBM N series for Oracle ILM

A key component of any ILM solution is the storage used to retain all the data, and IBM® System Storage™ N series offers a comprehensive range of devices.

IBM N series systems integrate easily into complex enterprise environments and provide shared access to UNIX®, Microsoft® Windows®, and Linux® operating systems, as well as Web data — while simultaneously supporting a Fibre Channel (FC) storage area network (SAN), IP SAN (iSCSI), and network attached storage (NAS). These high performance systems have a proven ability to continuously serve data at higher than 99.99 percent availability, and can scale from 50 gigabytes to 504 terabytes using Serial ATA or FC disk drives.

All IBM N series storage solutions leverage standard N series features such as the highly optimized and scalable IBM System Storage N series with Data ONTAP® storage operating system, enabling minimized complexity and costs with unequaled storage environment flexibility.

Tightly coupled with Oracle 10g, the IBM System Storage N series with SnapLock® software product family delivers high performance and high-security data permanence to disk-based near-line and primary IBM N series storage. An integrated element of the proven Data ONTAP® operating system, SnapLock software helps ensure the permanence, accuracy, integrity, and security of data by enabling business records to be both unalterable and rapidly accessible online for long periods.



Implementing ILM in four easy steps

Building an ILM solution using Oracle 10g and IBM N series is quite straightforward, and it can be achieved by following these four simple steps.

STEP 1 – Define the data classes

In order to make effective use of ILM, the first step is to look at all the data in your organization, what type of data is it, where is it stored and determine:

- What data is important, where is it and what needs to be retained
- How this data flows within the organization
- What happens to this data over time and is it still needed
- The degree of data availability and protection that is needed
- What is the legal and business requirements for data retention.

STEP 2 – Create storage tiers for the data classes

Since Oracle 10g can take advantage of all the different storage options that are available, the next step is to establish some or all of these storage tiers:

- High performance
- Low cost
- Online archive.

The high-performance storage tier is where all the important and frequently accessed data would be stored. This would utilize the smaller, faster disks on high performance storage devices. IBM System Storage N series N7000 using FC disks are the fastest and most reliable. The N series N7000 servers are enterprise class storage subsystems that provide high capacity and high speed data access.

The low cost storage tier is where the less frequently accessed data is stored. This tier could be built using large capacity ATA disks like those found on IBM System Storage N series N5000 with ATA disks attached. N series N5000 servers with ATA disk are an attractive lower cost option for the low-cost storage tier. The servers can also serve a dual role in that they support FC as well as ATA disks in the same storage subsystem.

The online archive storage tier is where all the data that is never or hardly accessed would be stored. It is likely to be extremely large and to store the maximum quantity of data on the online archive storage tier. This tier could be located in the database or it could be in another database, which serves as a central archive database for all information within the enterprise.

The online archive tier would be comprised of low-cost storage devices like the IBM System Storage N series with NearStore® servers or SnapLock devices. The data would still be online and available, for a cost that is only slightly higher than storing this information on tape, without all the disadvantages that come with archiving data to tape. If the online archive storage tier is identified as read-only, then it would be impossible to change the data and database backups would not be required after it is backed up the first time.

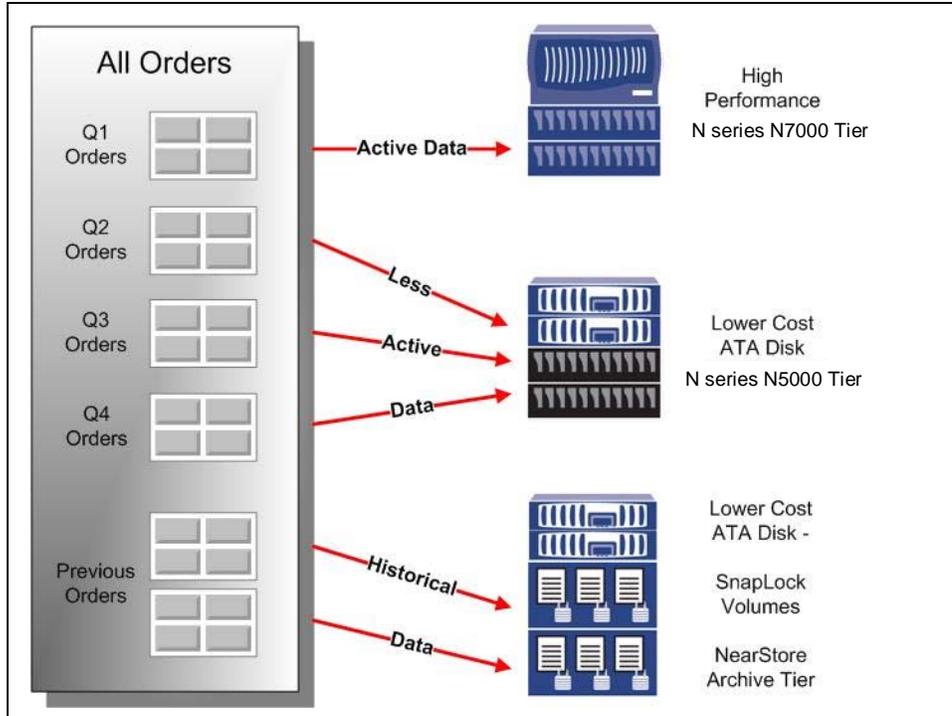


Figure 2) IBM N series tiered storage.

IBM N series SnapLock is available in two versions:

- IBM System Storage N series with SnapLock Compliance: Enables organizations to satisfy strict records-retention regulations such as SEC Rule 17a-4 (broker-dealers), HIPAA (healthcare), Sarbanes-Oxley (public companies), 21CFR Part 11 (life sciences), and DOD 5015.2 (government). Only an act of willful destruction, such as physically removing disks from an IBM N series system, can result in record deletion or alteration prior to the specified retention date.
- IBM System Storage N series with SnapLock Enterprise: Enables adherence to rigorous organizational best practices through functionality similar to that of SnapLock Compliance, but allows administrators to delete entire SnapLock Enterprise volumes. Under no circumstances is it possible for any SnapLock Enterprise user or administrator to delete or modify individual SnapLock Enterprise WORM partitions or undermine SnapLock Compliance WORM volumes.

IBM N series SnapLock complies with regulatory and best-practices records-retention requirements by allowing the creation of nonrewritable, nonerasable write-once-read-many (WORM) volumes on an IBM System Storage N series with NearStore, thereby preventing critical files or Oracle partitions from being altered or deleted until a specified retention date.

IBM N series SnapLock supports open, industry-standard protocols—such as native file system (NFS) and common internet file system (CIFS)—so it offers easy data access and tight integration with Oracle 10g. Consolidating storage on a unified, rapidly scalable IBM N series with NearStore enables increased flexibility in preserving important data using archival and records-management applications.



Assigning classes to storage tiers

Once the storage tiers have been defined, the data classes (partitions) identified can be assigned to the appropriate storage tiers. This provides an easy way to distribute the data across the appropriate storage devices depending on its usage, whilst still keeping the data online and readily available and stored on the most cost-effective device. Since partitioning is completely transparent to anyone accessing the data, no applications changes are required, thus it can be implemented at any time at minimal cost.

STEP 3 – Create data access and migration policies

The next step is to specify who can access the data and the operations they may perform and how to move the data during its lifetime.

Managing access to data

Regulatory requirements are beginning to place exacting demands on how data can be accessed. Using Oracle 10g, ILM, and IBM N series, extremely effective methods for controlling what authorized users of the database may see is to use database views or implement security policies using virtual private database.

A security policy is defined at the database level and it is transparently applied to all database users. The benefit of this approach is that it provides a secure and controlled environment for accessing the data, which cannot be overridden and can be implemented without requiring any application changes.

Migrate data between classes

Within Oracle 10g, there are a number of ways that data can be physically moved within the database to take advantage of the different storage tiers. If the data is partitioned, then the partition containing the orders for Q2 can be moved from the high performance storage tier to the low cost storage tier online. Since the data is being moved within the database, it can be physically moved, without affecting the applications that require it, or cause disruption to regular users. As of Oracle 10g Release 2, the partition can be moved online.

During the lifecycle of the data it will be necessary to move it at various times and this occurs for a variety of reasons, such as:

- For performance, only a limited number of orders are held on the N series N7000 with high performance FC-protocol disks
- Data is no longer frequently accessed and is using valuable high performance storage and needs to be moved to a low-cost N series N5000 storage subsystems with ATA disks
- Legal requirements demand that the information is always available for a given period, and it needs to be held safely at the lowest possible cost, which is when the use of IBM N series SnapLock should be considered.

Step 4 – Define and enforce compliance policies

The fourth step in defining our ILM environment is the creation of policies for compliance. When data is decentralized and fragmented, compliance policies have to be defined and enforced in every data



location, which could easily result in a compliance policy being overlooked. However, using the Oracle Database to provide a central location for storing data means that it is then very easy to enforce compliance policies as they are all managed and enforced from one central location.

When defining compliance policies there are five areas to consider:

1. Retention
2. Immutability
3. Privacy
4. Auditing
5. Expiration.

The retention policy will describe how the data is to be retained, for how long it must be kept, and what happens at life end. Therefore, an example of a retention policy is that a record must be stored in its original form, no modifications are allowed, it must be kept for seven years and then it may be deleted. Using Oracle security and IBM N series SnapLock technology, it is possible to ensure that data remains unchanged and only using authorized processes, can data be removed at the appropriate time.

Immutability is concerned with proving to an external party that data is complete and has not been modified. Digital signatures can be created and held either inside or outside of the database, to show that data has not been altered or tampered in any way. Additionally, the partitions residing on an IBM N series NearStore could be placed on an IBM N series SnapLock volume that would have WORM capabilities to absolutely protect against tampering or accidental deletion.

With so much data being retained today, it is extremely important to maintain privacy of data at all times and the Oracle database can strictly control access to data using security policies defined using Virtual Private Database (VPD), which define exactly which information a user may see. Maintained at the database level, these policies cannot be violated by anyone. As of Oracle 10g Release 2, individual columns can also be encrypted so that anyone looking at the raw data cannot see its contents.

The Oracle Database also has its own auditing capability to track all access and changes to data. These can be defined at the table level or via fine-grained auditing which specifies the criteria for when an audit record should be generated.

Ultimately, data may expire for business or regulatory reasons and its need to be removed from the database. Since that can involve removing vast quantities of data, the Oracle database can remove data very quickly and efficiently by simply dropping the partition that contains the information that resides on an IBM N series SnapLock volume that has been identified for removal.



Conclusion

Oracle 10g provides an ideal platform for implementing ILM because it is simple to use and there are no specialized data stores to manage. It has proven fast performance for the rapid retrieval of data. The security features in the database and on the IBM N series storage utilizing SnapLock technology ensure that data is secure from unauthorized access and data is always transactionally consistent.

The Oracle Database provides total flexibility, therefore it can rapidly adapt to any change in requirements, which is extremely important due to the continuing evolution of the emerging regulations. Finally, the Oracle Database has been around for over 20 years, therefore you know that Oracle will be supported for many years, which is very reassuring.

IBM N series servers are data-center proven and are highly scalable. IBM N series servers are also flexible in their ability to utilize fast FC disk as well as inexpensive ATA disks. IBM N series SnapLock technology provides WORM capabilities that meet stringent government regulations for the protection of data.

ILM enables us to understand our data, which is an extremely valuable business asset, which must be managed properly, to ensure business success and regulatory compliance. Using Oracle 10g, IBM N series storage subsystems and SnapLock technology, a comprehensive ILM solution can be implemented for the lowest possible cost.



Trademarks and special notices

© International Business Machines 1994-2008. IBM, the IBM logo, System Storage, and other referenced IBM products and services are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. All rights reserved.

References in this document to IBM products or services do not imply that IBM intends to make them available in every country.

Data ONTAP, NearStore, Network Appliance, the Network Appliance logo and SnapLock are trademarks or registered trademarks of Network Appliance, Inc., in the U.S. and other countries.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Other company, product, or service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

Information concerning non-IBM products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by IBM. Sources for non-IBM list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. IBM has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-IBM products. Questions on the capability of non-IBM products should be addressed to the supplier of those products.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.