Oracle GRID Center

Oracle Database ILM Solution based on Fujitsu ETERNUS DX/SPARC Enterprise

— Lower storage costs and power consumption for long-term data storage —

Creation Date: March 3, 2009
Last Update: April 16, 2010

Version: 1.1

ORACLE

FUJITSU

Contents

1. Introduction

The rapidly changing economic climate has generated a pressing need to reappraise corporate investments. IT investment is no exception, and there is growing pressure for a clear, detailed awareness of the specifics of IT costs and ways to lower such costs. At the same time, with corporate business becoming increasingly computerized, the data volumes handled by corporate systems have continued to grow. Exacting requirements for long-term storage of business data to satisfy revised laws and compliance requirements have dramatically increased the data volumes retained and stored on various systems. The volume of data handled by systems is said to have increased three-fold over the most recent two-year period, and a major focus in the quest to reduce IT costs has been minimizing rising storage costs associated with storing this data.

Fujitsu Limited and Oracle Corporation Japan have developed a joint Information Lifecycle Management (ILM) solution to help customers manage their data both in the short term and long term. This solution combines the Fujitsu ETERNUS DX storage system (hereinafter called "ETERNUS DX") with the Oracle Database 11g database to handle growing data management and storage costs generated by rising data volumes. This ILM solution is based on a business model that assumes access frequencies for business data fall off over time. It optimizes management and storage costs over data lifecycles by sequentially transferring archival data with lower access frequencies to lower cost storage areas, thereby making optimal use of storage facilities.

This ILM solution deploys a hierarchical storage system that combines a high-speed fibre channel (FC) disk and a high-capacity low-cost nearline serial ATA (SATA) disk into a single ETERNUS DX package. The database partitions the data tables covered by the ILM in chronological order, using Oracle Database 11g's Oracle Partitioning. This system reduces and optimizes overall storage costs by allocating partitions covered by the ILM to optimized storage areas.

This whitepaper examined partition transfers using Oracle Database MOVE PARTITION and data transfers using the RAID migration function specific to ETERNUS DX to move data within storage facilities and to establish procedures for the database ILM using a customer's system environment. We examined two methods to establish database ILM best practices for various situations, considering factors such as impact on business operations and load on server CPUs.

The design and development of Fujitsu's SPARC Enterprise and ETERNUS DX provide numerous energy-conserving features. This validation demonstrated that storage system energy consumption can be cut without affecting system performance by shutting down the disk used for the database backup area (except during backup acquisition) with the ETERNUS DX energy-saving "Eco-mode" function. This reduces power and IT system operating costs.

This document discusses best practices for database ILM solutions based on a combination of the Fujitsu ETERNUS DX disk storage system, SPARC Enterprise Unix server, and Oracle Database 11g to efficiently store and manage database data over extended periods, optimize storage costs, and cut energy consumption.

2. Validation Objectives

If the vast amount of corporate data handled is examined in closer detail, we see that the certain data has the following characteristics.

- Recent data is accessed frequently and is often subject to high response requirement processing (e.g., for online business operations).
- Access frequency decreases as data becomes older, and demand for performance decreases accordingly.

This is true for data such as order history tables, where records are added and then stored as long-term records. Information lifecycle management (ILM) is a term that refers to managing data with such characteristics by optimizing costs and service levels to suit access frequency, processing requirements, and data retention period.

ILM involves optimizing data storage costs by moving data accessed less frequently from costly high-speed and limited capacity disks to lower-performance, high-capacity, low-cost disks. ILM makes it possible to provide the storage capacity needed to store all of the data that needs to be available, regardless of access frequency, at reduced cost. Since nearline disks generally offer greater capacity per drive than high-speed disks, they make it possible to use fewer actual disk drives and lower potential energy requirements.

Oracle Database includes the Oracle Partitioning function for partitioning tables based on the date on which data was added. Although each table is normally stored in one table space, partitioning allows the table to be divided into partitions and each partition assigned to a specific tablespace. ILM is applied to the database by partitioning corresponding tables within the database in accordance with data storage policies and assigning them to the table space corresponding to the service levels required for the respective partition.

The Fujitsu ETERNUS DX disk storage system offers the advantage of allowing the incorporation of a high-speed, high-reliability FC disk together with a high-capacity, low-cost nearline SATA disk within the same package. ILM can then be achieved within a single storage package by using Oracle Database to partition tables and allocating recent, frequently accessed data to the high-speed FC disk and less frequently accessed archival data to the nearline SATA disk. ETERNUS DX also includes a RAID migration function for transferring logical volumes to separate RAID groups, allowing logical volumes on a RAID group consisting of an FC disk to be transferred within the storage space to a nearline SATA disk. No database server resources are used, since data transfers based on RAID migration operate entirely within the storage space, minimizing impact on database. While MOVE PARTITION is generally used to move data in Oracle Database, using RAID migration in conjunction with ETERNUS DX offers a choice of methods for moving data.

The following validation was performed by Fujitsu and Oracle Japan using the GRID Center joint validation center to establish design and operating procedures for ILM using ETERNUS DX and Oracle Database.

- Validation of ILM data transfers from high-speed disks to nearline disks

- Validation of data processing performance on nearline disks
- Validation of the impact on ILM backup and backup performance

3. Validation Equipment

This section describes the equipment used in this validation.

3.1. Fujitsu SPARC Enterprise

SPARC Enterprise is available in the form of the SPARC Enterprise M9000, M8000, M5000, M4000, and M3000 to offer mainframe-comparable reliability for mission-critical operations to meet the various operational issues companies face, such as speed, continuity, TCO reductions, and environmental obligations. Other SPARC Enterprise formats include SPARC Enterprise T5440, T5240, T5220, T5140, and T5120, which offer high throughput ideal for Web front-end operations and application servers.

3.1.1. SPARC Enterprise M4000

[SPARC Enterprise M4000 features]

- Provides mainframe reliability via the SPARC64 VI/VII high-performance processor.
- Achieves up to 16-core, 32-thread functionality in a single package using multi-core, multi-thread technology.
- High performance using increased bus bandwidth and PCI Express.
- Improved unit availability through comprehensive data protection and redundancy.
- Flexible server operations based on hardware partitioning, DR, and COD functions.

The SPARC Enterprise M4000 is a mid-range class server that incorporates the functions traditionally offered by high-end servers, including high performance, high reliability, and virtualization technology. It provides the reliability of mainframes, using a SPARC64 VI/VII high-performance processor to achieve up to 16-core and 32-thread multi-core, multi-thread configurations in a single device.

Mission-critical operations, including those using database and batch processing, typically have high loads per transaction and sometimes require specific account processing sequences. The SPARC64 VI/VII high-performance processor was developed to handle such high-speed processing of high-load transactions. High performance is achieved using new multi-core multi-thread technology in addition to powerful parallel command processing capabilities and precision command branching prediction technologies. Overall system performance is further increased with a strengthened system bus and use of PCI Express.

By definition, mission critical operations have serious business repercussions if interrupted. The development of SPARC Enterprise is based on mainframe design concepts. It incorporates a number of technologies to minimize the risk of system stoppages involving server failures or other issues, including functions that prevent

unforeseen errors, functions that allow operations to continue by correcting or reducing functionality when problems arise, as well as component redundancy and hot-swapping functions.

SPARC Enterprise M4000 also enables dynamic reconfiguration by partitioning resources within the package using virtualization technologies in the form of hardware partitioning and dynamic reconfiguration functions—functions previously available only with high-end servers. The specific operations that place loads on the server vary over time—for example, during the daytime, during the night, and at the beginning or at the end of the month. While servers have traditionally been provided individually to match peak workloads, this technology enables the necessary resources to be added and removed as required, providing a flexible response to variable workloads.

3.1.2. SPARC Enterprise M3000

[SPARC Enterprise M3000 features]

- Offers leading entry-class processor performance for SPARC/Solaris.
- Offers mid-range class high-reliability technology in an entry-level product.
- Offers Green Policy Innovation features (energy and space savings).

The SPARC Enterprise M3000 incorporates a SPARC64 VII to offer multi-core, multi-thread configurations with up to four cores and eight threads. Systems can address up to 64 GB of memory, enabling all the functionality required for business operations with a 2U (2-unit) space. Standard configurations include one SAS port and four PCI Express slots. Offering class-leading performance for entry-level models, the SPARC Enterprise M3000 is ideal for a wide range of applications, including database servers and application servers.

The SPARC Enterprise M3000 also features the high level of reliability associated with the M4000 to M9000 models, ensuring high reliability through system reliability at the LSI, unit, and system levels.

The SPARC Enterprise M3000 is also a Super Green environmentally-friendly product, as specified by Fujitsu. With its 2U (2-unit) dimensions, it offers low weight and space savings of 50% over the PRIMEPOWER450 (4U). Its maximum power consumption of 505 W (at 100 V) represents a 54% reduction. Combined with performance improvements, these improvements cut annual CO_2 emissions by approximately 65%. The SPARC Enterprise M3000 also boasts a low-noise design with operating noise levels of 47 dB at a standard server site ambient temperature of 25°C, making it the leading environmentally-friendly server for power consumption and noise levels among competing 4-core servers.

The standard Solaris 10 configuration includes the virtualization technology known as

Solaris Container, which allows resources to be centralized through server integration, even with the SPARC Enterprise M3000, for improved system efficiency.

3.2. ETERNUS DX Disk Storage System

Fujitsu provides storage systems that meet wide-ranging customer requirements. The ETERNUS DX disk storage system was developed to meet the following three requirements:

1. Ensure business continuity and scalability for large data volumes without delays.
2. Ensure data integrity and security for correct data storage.
3. Ensure appropriate and flexible corporate-level access to large-volume data while minimizing TCO.

Capable of responding to changes in operating configurations to achieve customer business goals, the ETERNUS DX provides storage infrastructure and services that ensure access to required data and storage resources at any time from business applications and processes assigned the appropriate authority.[1]

3.2.1. RAID Migration

RAID migration refers to transferring logical volumes to different RAID groups while guaranteeing data integrity and permitting reallocation of RAID logical volumes to suit customer operating requirements.

The ETERNUS DX permits logical volume reallocation in real-time, with no penalty on server CPU performance. It also allows rebuilding to different RAID levels—for example, from RAID5 to RAID1+0.

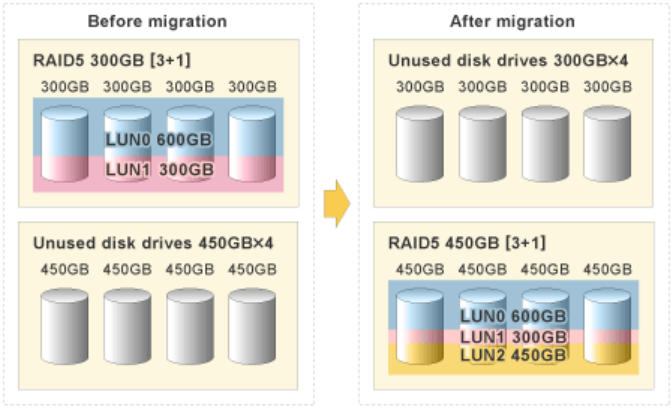


Figure 3-1 Example of 300GB Disks RAID5(3+1) configuration moving to different capacity of 450GB Disks RAID5(3+1) , and add other logical volume(LUN2) in free space

1

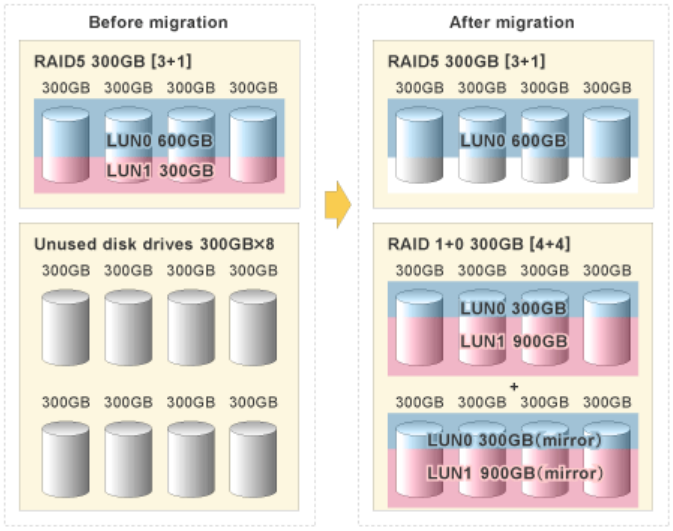


Figure 3-2 Example of 300GB Disks RAID5(3+1) configuration moving to different level of RAID Group RAID1+0

3.2.2. Eco-mode

The ETERNUS DX features an “Eco-mode” (MAID technology[2]) for controlling disk drive rotation only when necessary to suit customer system requirements.

Eco-mode is a mode for reducing power consumption by shutting down, for specific periods, disks to which access is restricted to certain timeframes. Shutdown scheduling involves setting the individual disk and time for individual RAID groups. This can also be set to coincide with operations such as backup.

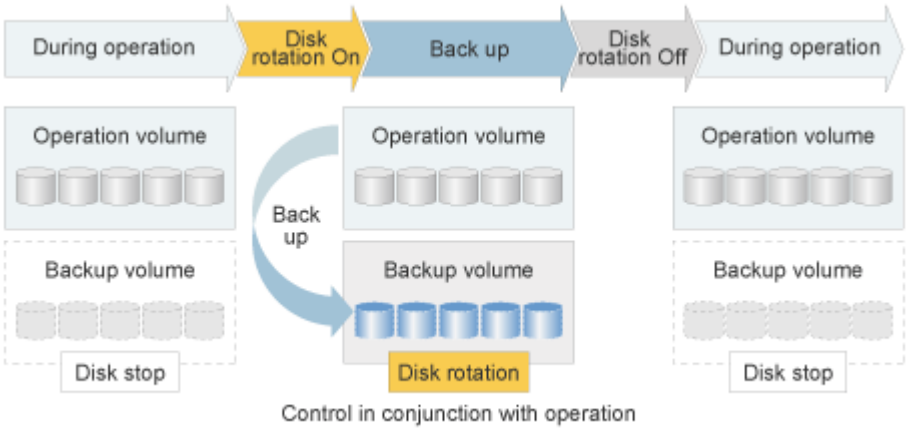


Figure 3-3 MAID Technology

Disk drives utilizing Eco-mode are shut down if no access occurs for more than a variable preset duration. If a disk access is attempted when a disk is shut down, it takes approximately one minute to restart the disk to permit access.

[2] MAID (Massive Array of Idle Disks) technology. Technology for reducing power consumption and extending the service life of disk drives by shutting down rarely-accessed disk drives.

The example configuration shown below based on an ETERNUS DX400 series (hereinafter called ETERNUS DX400) results in annual energy savings of approximately 4,720 kWh and CO_2 emission reductions of approximately 1,830 kg when backup volumes (50 disk drives) are shut down for 20 hours per day using Eco-mode. This reduces overall power consumption (and environmental burden) by approximately 15%.[3]

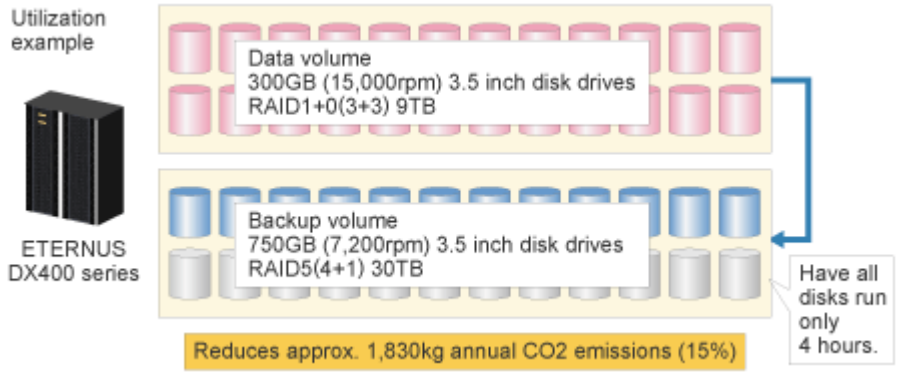


Figure 3-4 Example of Eco-mode usage[4]

3.2.3. ETERNUS SF Storage Cruiser and ETERNUS SF AdvancedCopy Manager

ETERNUS SF Storage Cruiser manages operations by linking resources such as business server file systems, connector paths, mirror disks, and databases from ETERNUS DX disk drives. The system allows ready identification of the correlation between individual resources and enables storage system expansion, fault recovery, and accurate information acquisition and display.

In conjunction with the ETERNUS DX , ETERNUS SF AdvancedCopy Manager provides high-speed backup/restore and replication operations based on the Advanced Copy function. The Advanced Copy function rapidly copies working volumes at set times to a different volume (copy volume) within the same disk array. The copy volume can then be used to perform a backup to a tape device. ETERNUS SF AdvancedCopy Manager tape server options can be used to make easy backups from disks to tape, eliminating complex procedures such as copy completion and tape backup start scheduling and management of multiple disks and tapes. The copy volume is separate from the working volume so that no overwriting occurs even if these steps are performed while actual work proceeds.

3.2.4. ETERNUS Multipath Driver

The ETERNUS multipath driver uses multiple paths from a single server to improve

[3] Fujitsu comparison figures for configuration shutdown for 20 hours daily and normal backup volume disk operations.

[4] For more information about Eco-mode , please refer to the following white paper:
"Energy-savings using the Fujitsu ETERNUS disk array MAID technology"
http://storage-system.fujitsu.com/jp/products/diskarray/download/pdf/MAID_whitepaper.pdf

sustained operations and system performance with a single system.

This driver is software that ensures continued access to a disk array even when a path fails, based on a configuration that multiplexes physical access paths between a server and disk array. Operating paths and standby paths can be set individually for each logical volume.

The load balancing function balances loads on paths (distributing I/O load) by simultaneously using multiplexed physical access paths between the server and disk array to improve system performance.

3.2.5. Features of the Nearline Disk in the ETERNUS DX

The flexible ETERNUS DX can include a mixture of online disk drives, that offer high-capacity and high-reliability, and near line disks, that offer low-cost drives, within the same package. This meets the requirements of customers whose data handling policies entail managing infrequently accessed data on nearline disk drives.

Nearline disk drives can be effectively deployed in the following cases:

Destination disk in disk-to-disk backup

The backup volumes used for disk-to-disk backups are generally used with multiple generation management for primary storage until data has been backed up to tape, or for emergency high-speed recovery. With their high capacity, high reliability, and low cost, nearline disk drives are best suited to such high-volume backup duties.

The ETERNUS DX Advanced Copy function creates a backup configuration offering high speed and outstanding cost performance.

For storing reference data

Data such as e-mail archives, video, images, audio, CAD, and R&D, or stored data covered by the Electronic Document Law is generally rarely accessed but nonetheless must be readily available for reference when needed. This type of rapidly proliferating data must also be stored for long periods of time, making nearline disk drives, with their outstanding cost performance, an ideal storage option.

Table 1 Differences between online disk and nearline disk

	Online disk drive (Fibre-channel disk drive)	Nearline disk drive (Nearline SATA disk drive)
Capacity	146 GB, 300 GB, 450 GB	500 GB, 750 GB, 1 TB
Rotational speed	15,000 rpm	7,200 rpm
Interface speed	FC (4 Gbit/s)	FC (4 Gbit/s)
Supported RAID level	RAID1, RAID1+0, RAID5, RAID6	RAID5, RAID6
Recommended use method	• For storing high-usage-frequency data	• For storing low-usage-frequency data (i.e. for backing-up and archiving data)

4. Oracle Database 11g Functions

This section describes the functions of Oracle Database 11g used in this validation.

4.1. Oracle Partitioning

Oracle Partitioning (a function available from Oracle 8) allows tables, indexes, and index configuration tables to be divided in specific ways to access these database objects. From an application viewpoint, partitioned tables require no changes even for SQL access, whether partitioned or unpartitioned.

Each partition has a fixed name and can be given various storage settings, including options to enable table compression, save partitions to a different table space, or save partitions to different ASM disk groups.

Oracle Partitioning provides a range of data distribution methods to control the transfer of data allocated to individual partitions. In ILM, range partitioning is effective for dividing data into different partitions based on a date value range.

In range partitioning, data is distributed based on the key value range. For example, if the table partition key is a date sequence and the “January 2009” partition is specified as containing data for January 2009, the partition will include data with key sequence values ranging from “01-JAN-2009” to “31-JAN-2009.” Data is distributed continuously without breaks. The lower limit of the range is automatically defined by the upper limit of the previous range.

This validation used Oracle Database 11g with range partitioning. Oracle Database 11g also expanded functions to handle large numbers of partitions. For more detailed information, refer to the white paper, “Partitioning with Oracle Database 11g” [5]

4.2. Automatic Storage Management (ASM)

ASM is a feature provided in Oracle Database to manage the disks used for a database. In addition to multiple disk striping and mirroring functions, it also provides a rebalancing function capable of dynamically altering disk configurations. This makes it possible to add disks without having to take the database offline. It also offers potential disk management cost benefits and performance improvements through dispersed disk loads.

4.3. Real Application Clusters (RAC)

RAC is a shared-everything cluster database that distributes information on disks and memory across multiple nodes and uses parallel processing for all nodes. RAC offers the following advantages:

1. Allows efficient use of server resources based on active-active configurations.

[5] http://otndnld.oracle.co.jp/products/database/oracle11g/pdf/twp_partitioning_11gR1.pdf

2. Facilitates flexible system expansion (simply by adding nodes).
3. Shared-everything configuration reduces system downtimes and switchover times in the event of faults.

5. Validation Setup

This validation was performed with an RAC configuration composed of two database servers and one storage device.

5.1. System Configuration

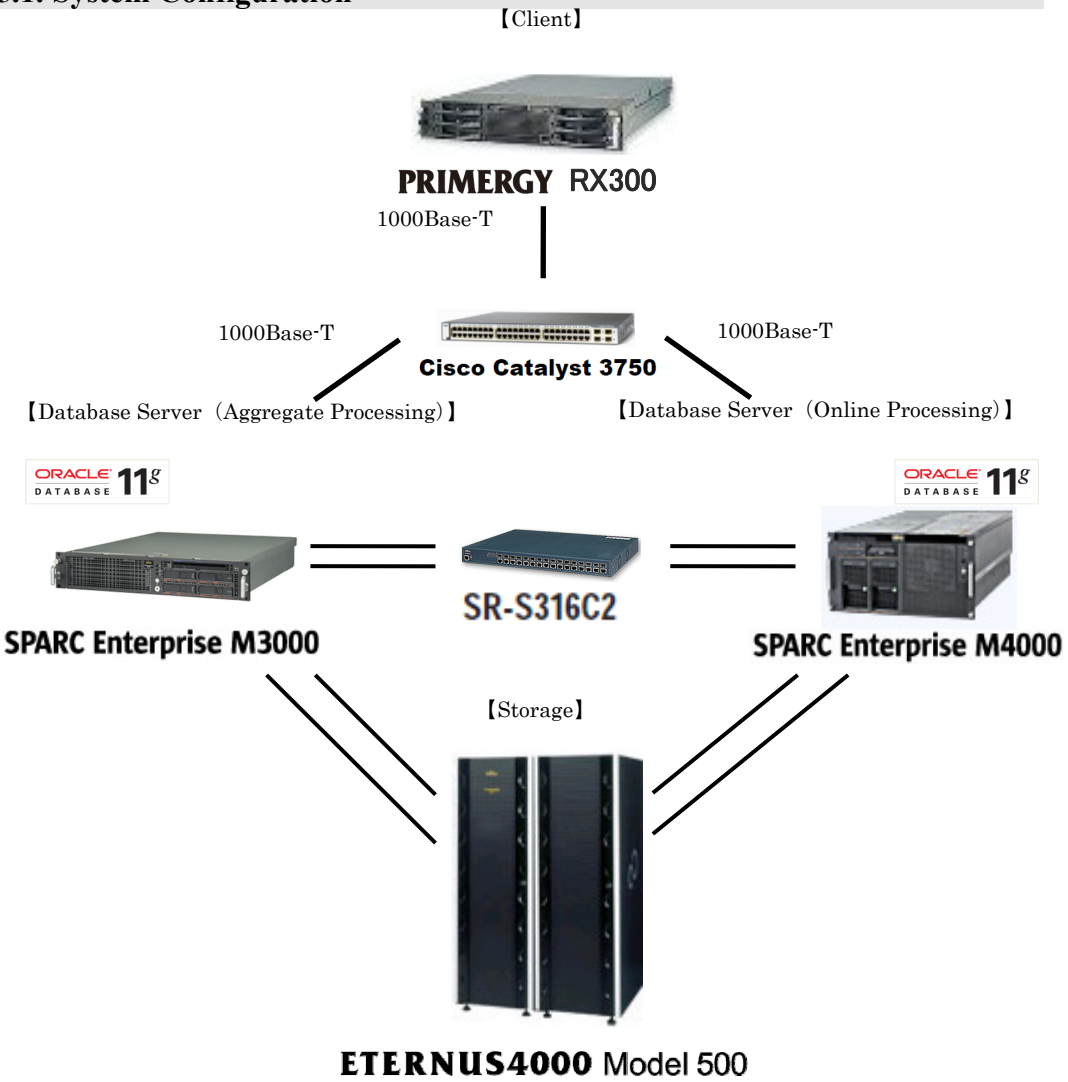


Figure 5-1 System Summary

Note*: ETERNUS DX440 is the successor of ETERNUS4000 Model 500.

5.1.1. Database Server (Online Business Server)

Hardware

Model	Fujitsu SPARC Enterprise M4000
CPU	SPARC64 VII 2.40 GHz/5 MB cache 2-CPU/8-core/16-thread
Memory	32 GB
Internal HDD	73 GB SAS Disk × 2

Software

OS	Solaris™ 10 Operating System (Generic_137137-09)
Database	Oracle Database 11g (11.1.0.7) Enterprise Edition
Storage management	ETERNUS SF AdvancedCopy Manager

5.1.2. Database Server (Tabulation Processing Server)

Hardware

Model	Fujitsu SPARC Enterprise M3000
CPU	SPARC64 VII 2.52 GHz/5 MB cache 1-CPU/4-core/8-thread
Memory	12 GB
Internal HDD	146 GB SAS Disk × 3

Software

OS	Solaris™ 10 Operating System (Generic_137137-09)
Database	Oracle Database 11g (11.1.0.7) Enterprise Edition
Storage management	ETERNUS SF AdvancedCopy Manager

5.1.3. Storage

Model	Fujitsu ETERNUS4000 Model 500
Disk drive	Fibre channel disk 146 GB (15,000 rpm) × 38 73 GB (15,000 rpm) × 12 Nearline SATA disk 750 GB (7,200 rpm) × 10

Note*: ETERNUS DX440 is the successor of ETERNUS4000 Model 500.

5.1.4. Client

Hardware

Model	Fujitsu PRIMERGY RX300
CPU	Xeon E5540 (Quad-core, 2.53GHz)
Memory	8GB (4GBx2)
Internal HDD	SAS 300GB(15K) x 3 (RAID-5)

Software

OS	Windows Server 2003 R2
Storage management	ETERNUS SF AdvancedCopy Manager
	ETERNUS SF Storage Cruiser

For detailed information on the hardware and software used by Fujitsu, please contact your Fujitsu sales representative.

5.2. Schemer Configuration

The schemer configuration used in this validation is shown below.

Table name	Number of entries	Summary
BUYERS	500,000	Customer table: Contains customer IDs and representatives.
DEPARTMENTS	11,220	Department table: Contains department IDs and names.
ORDERSFACT	62 million per month × 72 months	Order history table: Approx. 80 GB per year. 6 years' data created for 2004 to 2009.
PRODUCTS	10 million	Product table: Contains product IDs and product names.

Table name	Index	Summary
BUYERS	Idx_buyers	
DEPARTMENTS	Idx_departments	
PRODUCTS	Idx_products_prodid	
PRODUCTS	Idx_products_prodname	
ORDERSFACT	Idx_ordersfact_orderid	Approx. 50 GB per year
ORDERSFACT	Idx_ordersfact_spid	

5.3. Application Model

This validation used an application representing an in-house purchasing management system. Online processing was assumed to involve product order processing, and tabulation processing was assumed to involve monthly sales tabulation. Online processing is performed on one node and tabulation processing is performed on the other node.

5.3.1. Online Processing

This validation involved multiple execution of the two transactions shown below using Java applications.

1. Order transaction

    1. Acquire necessary information such as department ID and region ID from employee ID performing order processing in Departments table.
    2. Acquire product ID from name of product ordered in Products table.
    3. Enter details such as number of orders and insert.

    The tables searched in this transaction are read into the database buffer cache, after which internal memory processing ends.

2. Order change transaction

    1. Search for data for changing orders using Idx_ordersfact_orderid with the conditions of order date (timeid) and order number (orderid) from ordered data for January to March 2010.
    2. Change the number of orders for the data searched.

These two transactions are executed in multiples of 150. One process executes 10 transactions, with order transactions executed at a ratio of nine to every order change transaction.

5.3.2. Tabulation Processing

This validation involved executing the following two queries assuming tabulation processing.

1. Monthly sales tabulation queries

    Calculates the relative sales figures for a month compared to the previous month for data for a specific year.

2. Sales tabulation queries by employee

    Calculates the sales figures for each employee for January in a specified year and returns the names and sales figures for the top 100 sales.

In “6.1 ILM Using Oracle Standard Functions” and “6.2 ILM Using RAID Migration,” “1. Monthly sales tabulation queries” were performed repeatedly in serial form on data for 2008. Queries were performed in sequence two months at a time—for example, with sales comparisons for January and February 2008, with tabulation queries repeated for the year (2008) being searched once December was completed.

In “6.3 Effects of Disk Performance Differences on Tabulation Processing,” only monthly sales tabulation queries and sales tabulation queries by employee for 2009 data were used.

6. Validation Details and Results

This section describes the details and results of the ILM validation using MOVE PARTITION and RAID migration.

6.1. ILM Using Oracle Standard Functions

ILM using Oracle database functions uses MOVE PARTITION statements. Using MOVE PARTITION enables ILM tailored to customer system requirements and configurations.

6.1.1. ILM and Physical Design Using Oracle Standard Functions

This section describes efficient ILM operation procedures using MOVE PARTITIONand physical database design for such ILM operation.

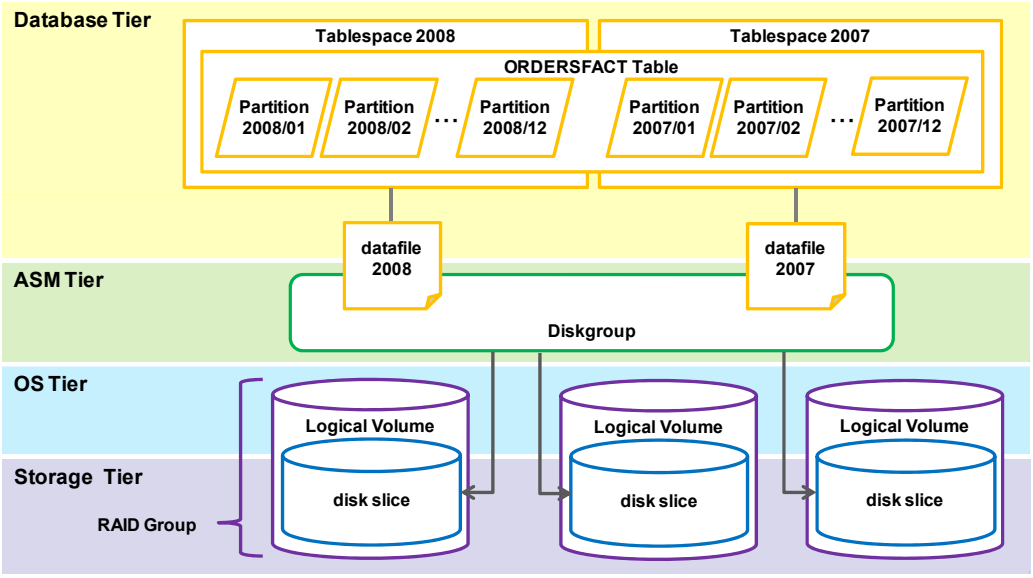


Figure 6-1 Physical Design of Oracle ILM by standard Function

Figure 6-1 shows an example with 2007 and 2008 data model used in this validation.In the above example, 12 months of data stored in a single tablespace and a single data file. While 1 DiskGroup to multiple data files are stored. DiskGroupis composed of multiple logical units.

Operations after introducing ILM are expected to involve periodic operations to move older partitions to disks to suit the current access frequency. Periodically transferring data with lower access frequency to low-cost disks prevents the increase of data on high-performance disks and enables the overall increase in data to be handled by adding low-cost disks, reducing the cost associated with adding disks.

ILM based on Oracle standard functions is considered for the ORDERSFACT table used in this validation. The tables covered by operations in ILM are shown below together with the transfer source and destination table space.

Tables covered by ILM operations: ORDERSFACT table

Partitions moved: P200901 to P200912 (approx. 80 GB)

Transfer source table space: TS_2009 (on FC disk)

Transfer destination table space: TS_2009_OLD (on SATA disk)

The data storage period was assumed to be 5 years, with data deleted after this period is exceeded. The following table space was therefore deleted when moving the 2009 partition.

Deleted table space: TS_2004_OLD

The following index partitions were also moved at the same time in this validation by rebuilding the index using the UPDATE INDEXES statement when executing the MOVE PARTITION statement (data for Jan to Dec 2009 only, approx. 50 GB).

- idx_ordersfact_orderid
- idx_ordersfact_spid

Note that this procedure is only an example. The actual procedure used should be selected to suit the customer's particular configuration and system requirements. The procedure is explained below.

(i) A new partition is added before the year changes from 2009 to 2010, and the index default table space is changed to the 2010 table space. (Figure 6-2)

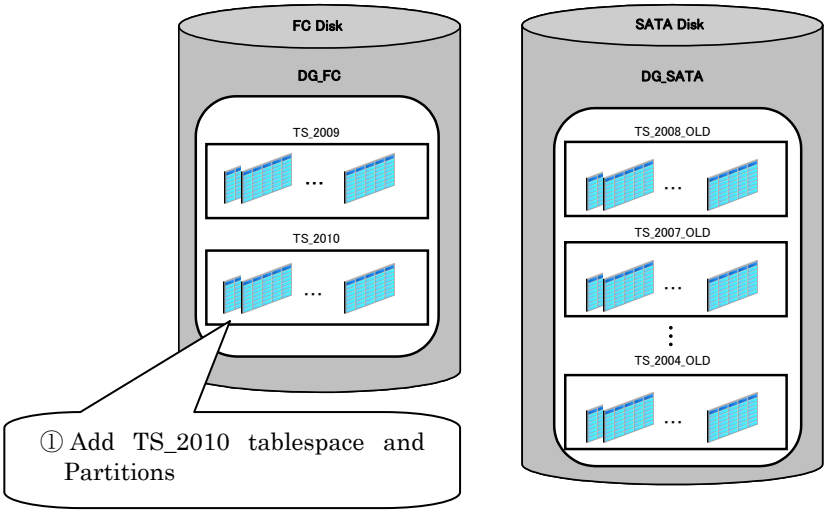


Figure 6-2 Adding Table Space and Partitions

The partition added excludes tabulation information, which could change the execution schedule and degrade performance. Similarly, newly added partitions have very few data entries compared to other partitions, and the optimizer statistics may differ significantly. This may also change the execution schedule and degrade performance. Operations that copy the 2009 partition optimizer tabulation information to the 2010 partition prevent performance degradations due to execution schedule changes.

(ii) The partition and table space (2004 partition and table space here) are deleted to remove partitions containing data that no longer needs to be retained. (Figure 6-3)

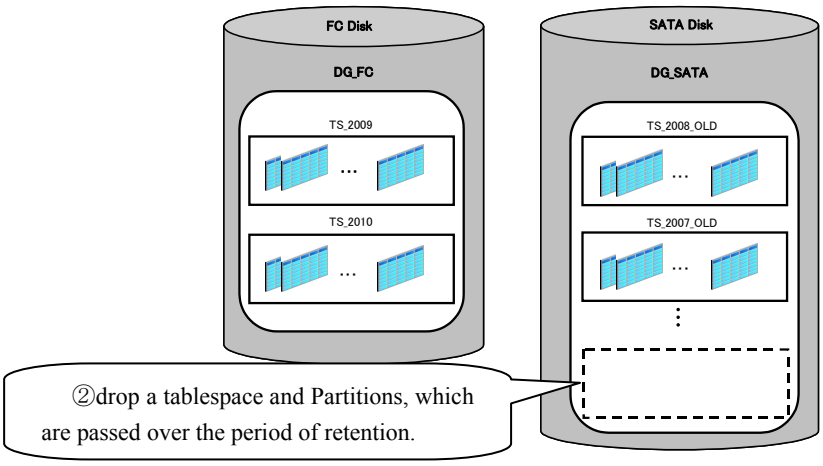


Figure 6-3 Drop Table Space and Partitions

(iii) The transfer destination table space is created on the SATA disk. (Figure 6-4)

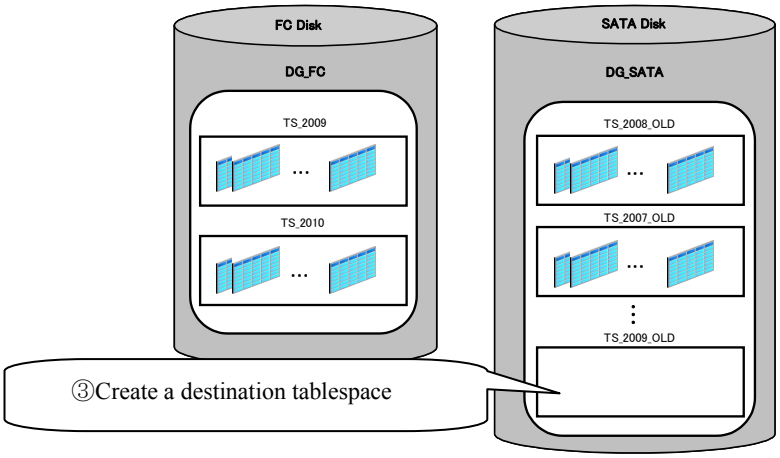


Figure 6-4 Create a Destination Table space

(iv) The transfer source table space is set to “READ ONLY,” and the ALTER TABLE … MOVE PARTITION statements are used to transfer the partition from the transfer source table space to the transfer destination table space. (Figure 6-5)

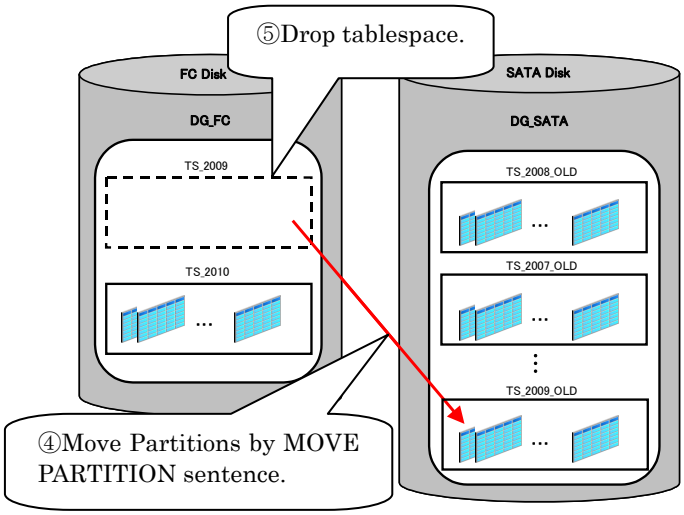


Figure 6-5 Moving Data by MOVE PARTITION

6.1.2. Resource Use Status when Executing MOVE PARTITION

The resource load status is first checked when transferring one year's data (approx. 130 GB with index) from the FC disk to the SATA disk using the MOVE PARTITION statement after the system is halted. In this validation, the MOVE PARTITION statement is executed as a single process. Thus, CPU utilization is for one thread (approximately 12.5%).

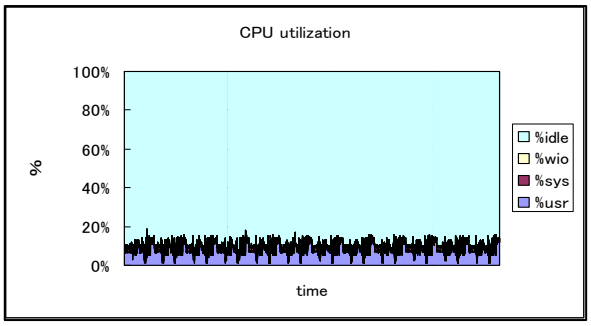


Figure 6-6 CPU usage while MOVE PARTITION

If the MOVE PARTITION statement is executed while other operations are underway, CPU usage will increase by roughly 10% while MOVE PARTITION is being executed. Precautions are needed to avoid CPU shortfalls when executing MOVE PARTITION.

FC and SATA disk loads are checked next. Figure 6-7 shows the FC and SATA disk loads.

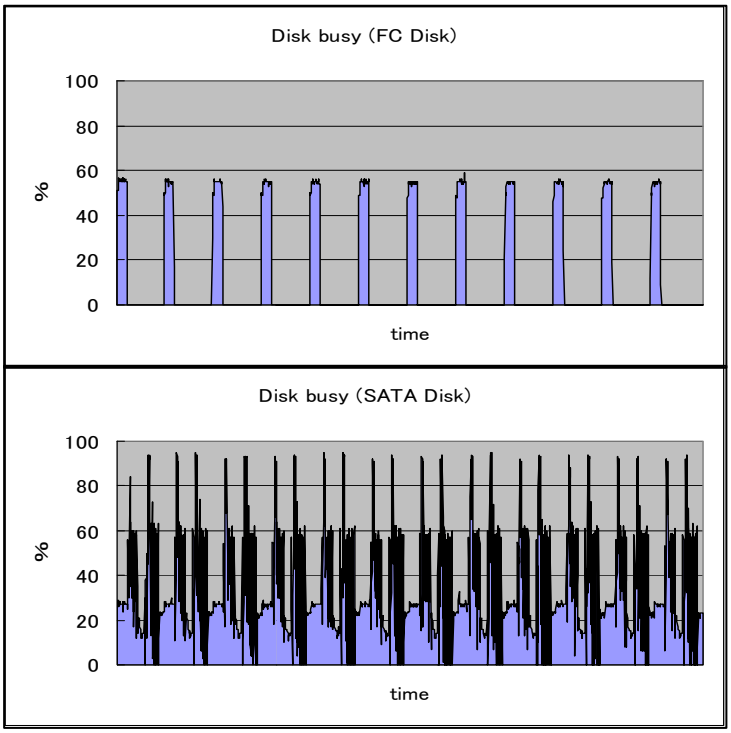


Figure 6-7 Disk Busy Parcent while executing MOVE PARTITION

Executing the MOVE PARTITION statement reads the data stored in the FC disk in partition units and writes it to the SATA disk. The index is rebuilt once the data has been transferred. Since the data has already been transferred to the SATA disk, it is read in from the SATA disk, sorted, then written to the SATA disk.

For this validation, partitions were created for each month and partitions for each year's data were transferred together. This is why the graph shows 12 peaks for each partition.

Figure 6-8 illustrates the processing for individual partitions.

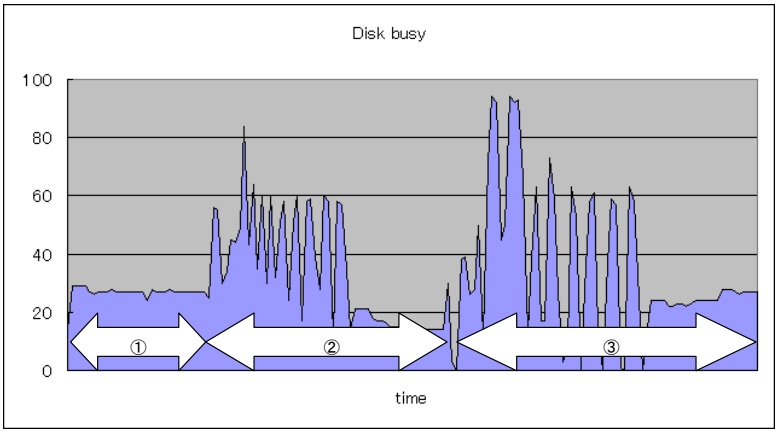


Figure 6-8 Conduct of each Partitions

Data is transferred as in (i) and the index is rebuilt as in (ii) and (iii). The SATA disk load for index rebuilding is three times that for the data transfer. Based on these results, we recommend determining the SATA disk load status and executing the MOVE PARTITION statement when the disk load is low.

The graph also shows that index rebuilding is a lengthy process. The time required to execute the MOVE PARTITION statement depends on the size and quantity of the indexes.

6.1.3. Impact of MOVE PARTITION on Operations

We will examine what happens when the MOVE PARTITION statement is executed during online processing or tabulation processing.

We first confirm the impact of MOVE PARTITION on online processing. For information on online processing specifics, refer to “5.3.1 Online Processing.”

Online operations are performed using a single node. We examined the impact of executing the MOVE PARTITION statement for each node. For more information on the validation procedure, refer to “6.1.1 ILM and Physical Design Using Oracle Standard Functions.”

6.1.3.1 Executing MOVE PARTITION with a Node not Performing Online Operations

We first confirm the throughput and response time for normal online processing (i.e., online processing only). Figure 6-9 shows the relative values with respect to normal mean throughput with response time set to 1.

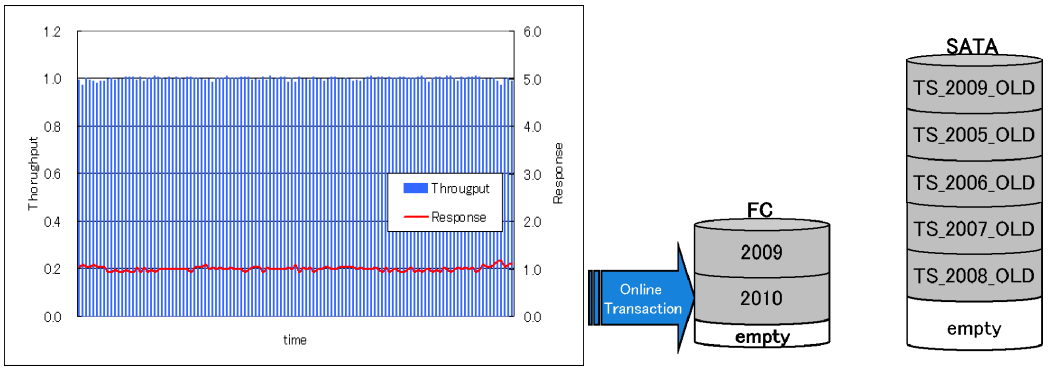


Figure 6-9 Throughput and response time for normal online processing

Figure 6-10 shows throughput and response times when 2009 data is transferred to the SATA disk using the MOVE PARTITION statement with a node not used by online processing during online processing. It shows relative values with respect to normal mean throughput with response time set to 1.

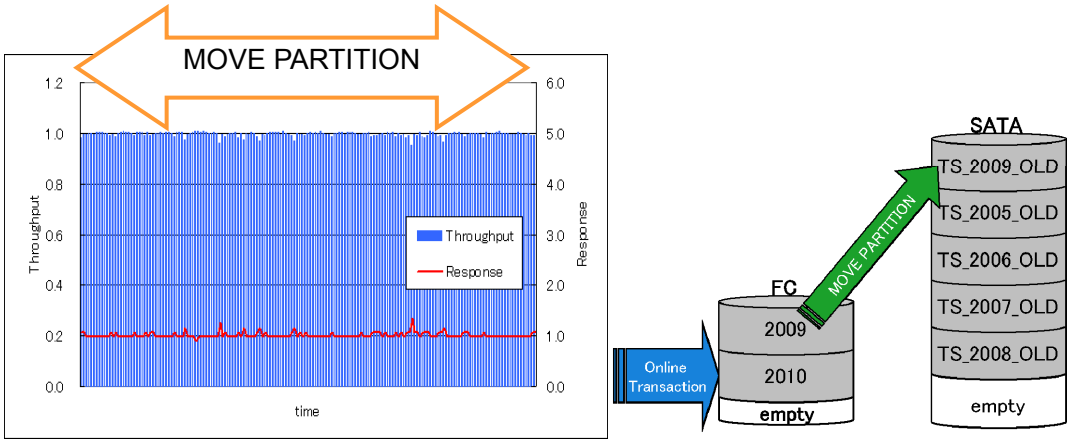


Figure 6-10 Transaction Performance of execution MOVE PARTITION with a Node not performing Online Operations

Executing the MOVE PARTITION statement with a node not used for online processing during online processing will have virtually no impact on online processing. As shown in Figure 6-11, there is virtually no difference in execution times for the MOVE PARTITION statement.

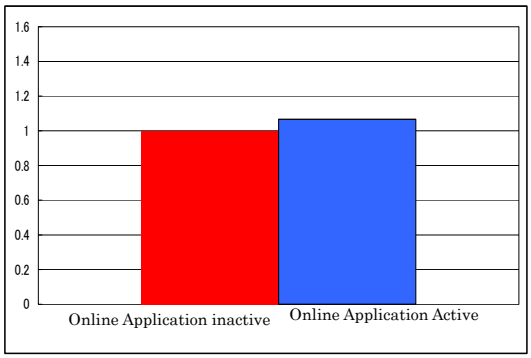


Figure 6-11 Time comparison of execution MOVE PARTITION

As shown above, the impact on online processing can be minimized when dividing operations in an RAC setup and by executing the MOVE PARTITION statement on a node with low load.

6.1.3.2 Executing MOVE PARTITION with a Node Performing Online Processing

This validation assumes that the MOVE PARTITION statement is executed in a single DB setup rather than an RAC setup.

Figure 6-12 shows throughput and response times when 2009 data is transferred to the SATA disk using the MOVE PARTITION statement with a node used for online processing. It gives values as relative values, with normal mean throughput and response time defined as 1.

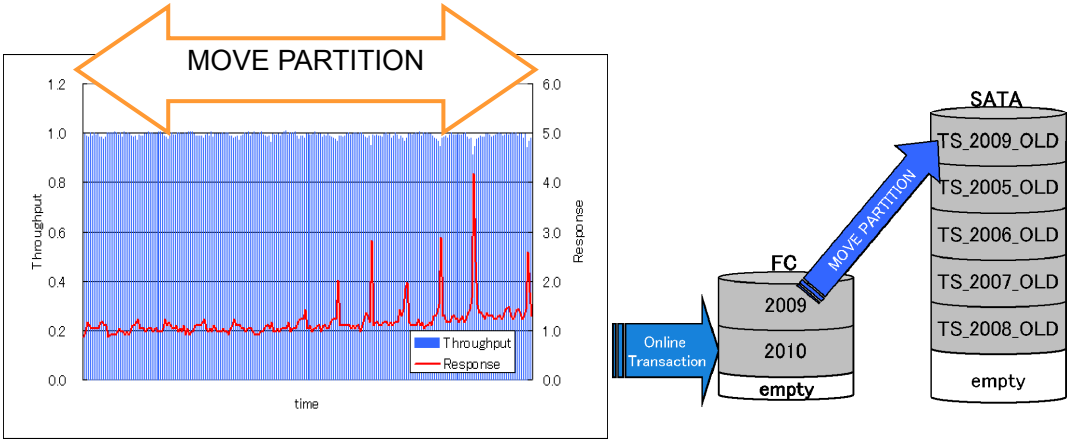


Figure 6-12 Transaction Performance of execution MOVE PARTITION with a Node Performing Online Operations

The graph shows how response times degrade at certain points. This is believed to be due to past product search query delays occurring for the order change transaction due to increased FC disk load when executing the MOVE PARTITION statement, especially when transferring data. Figure 6-13 is a graph showing the FC disk busy rate when executing the MOVE PARTITION statement during online processing.

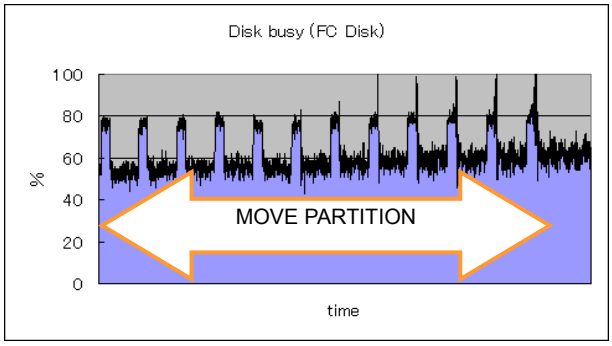


Figure 6-13 Disk Busy Parcent while execution online processing

The graph shows the time for which the disk busy rate reaches 100%. Of the order change transactions, SQL statements searching for already ordered data appear to be affected at these points.

We will now look at CPU use. Figure 6-14 is a graph showing CPU usage when only online processing is performed and when the MOVE PARTITION statement is executed at the same time. CPU usage is approximately 10% greater when executing the MOVE PARTITION statement simultaneously than when performing online processing alone.

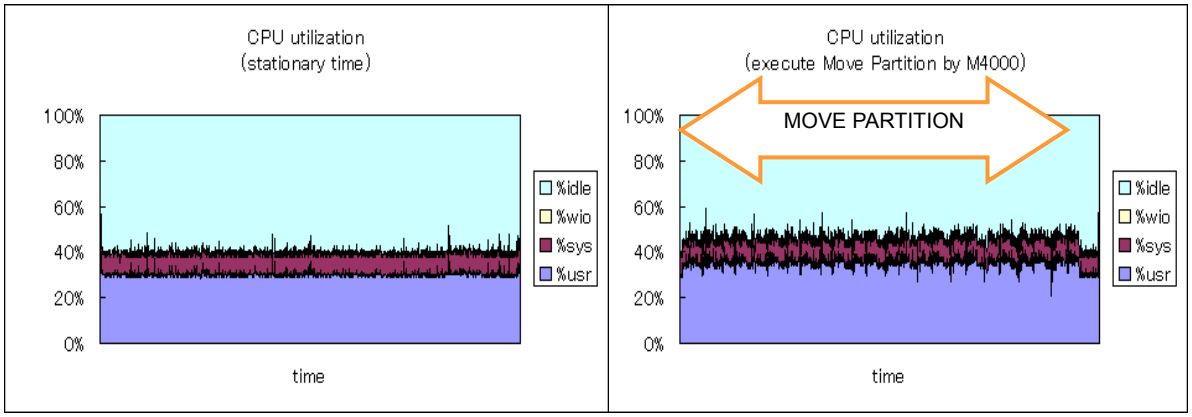


Figure 6-14 Compare to CPU usage execution online process and execute MOVE PARTITION with online process

These results show that when considering operations with a single DB setup, FC and SATA disk loads are more important factors than CPU use. Online processing is likely to be affected if the FC disk is subject to high loads.

6.1.3.3 With Partitions Divided Among Different FC Disks

The validation above indicates the likelihood of increased loads on the FC disk affecting operation processing in the case of ILM with a single DB setup. However, it appears possible to minimize the impact on order change transactions by allocating the 2010 and 2009 partitions to different RAID groups on the FC disk to distribute the load on the FC disk. Figure 6-15 shows throughput and response times for each transaction when the 2009 and 2010 partitions are allocated to different RAID groups. It gives values as relative values, with normal mean throughput and response time defined as 1.

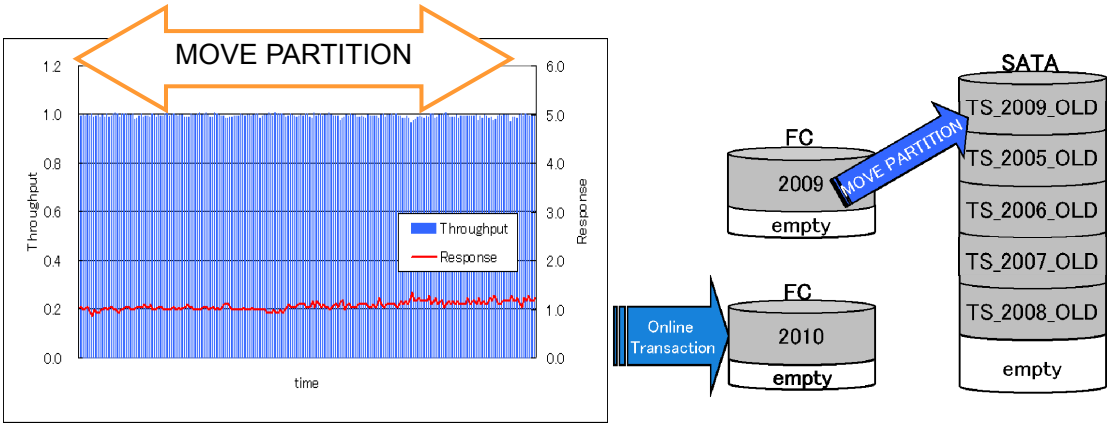


Figure 6-15 Relative value of the transaction performance if divide the FC Disk

As can be seen in Figure 6-15, no major differences are apparent in performance for each transaction compared with normal operations.

The individual FC disk loads are shown in Figure 6-16.

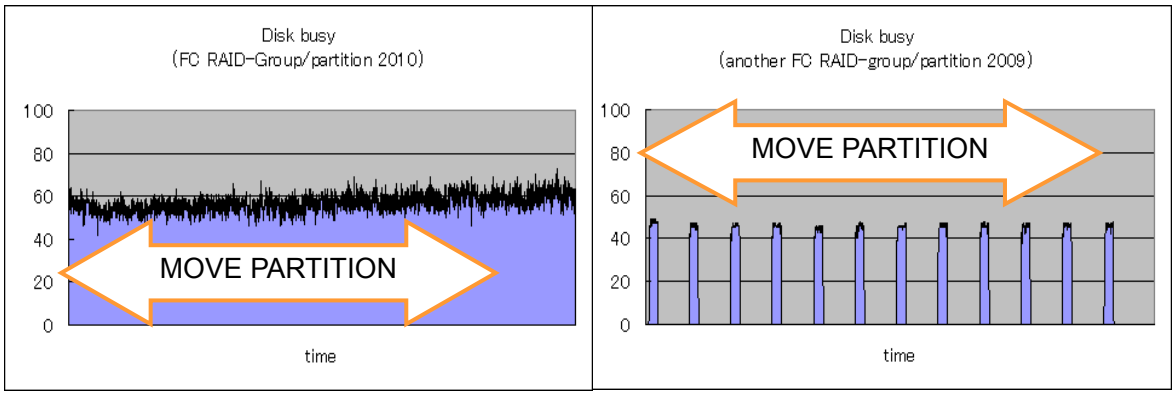


Figure 6-16 Disk Busy Percent of each FC Disk

If the FC disks have high loads, the impact on transactions can be minimized by allocating the partitions accessed by online processing and the partitions moved using the MOVE PARTITION statement respectively to different RAID groups to distribute the disk I/O.

6.1.3.4 Impact on Tabulation Processing

We will now examine the impact of executing the MOVE PARTITION statement for a node used for tabulation processing. For more information about tabulation processing, refer to “5.3.2 Tabulation Processing.”

Tabulation operations are performed using a different node to the node used for online processing. For more information on the validation procedure, refer to “6.1.1 ILM and Physical Design Using Oracle Standard Functions.”

The query response time for tabulation processing is shown in Figure 6-17. It shows the relative values for 2008 data with respect to the mean query response time as 1.

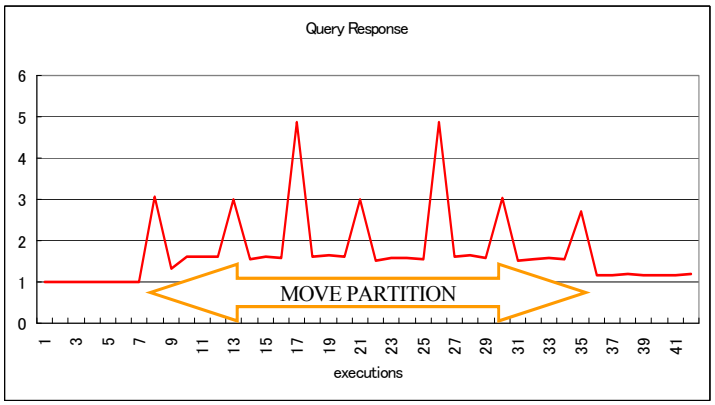


Figure 6-17 Response ratio of aggretate queries

The response time is worsened by up to five-fold when executing the MOVE PARTITION statement.

We next examine CPU usage and disk busy rate.

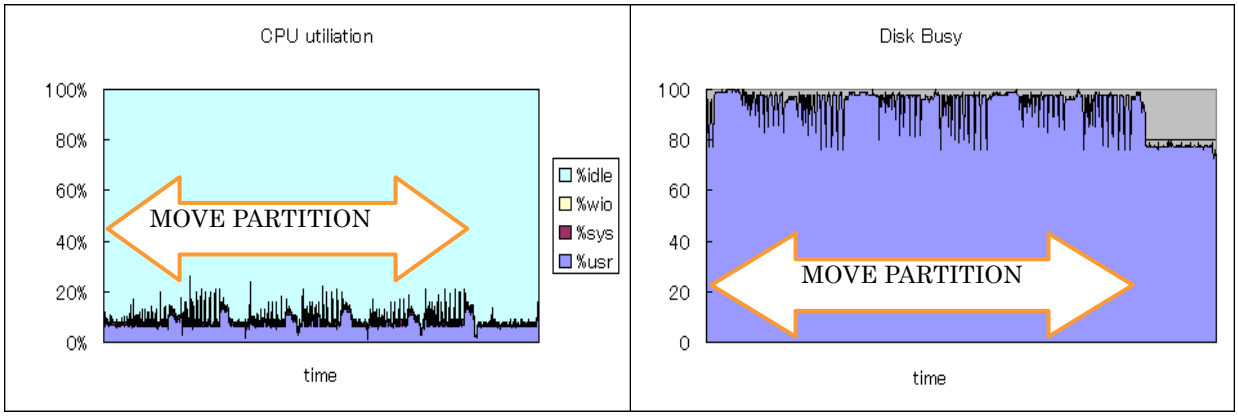


Figure 6-18 CPU usage and Disk Busy Percent when running queries

Figure 6-18 shows how the busy rate for the SATA disk increases by approximately 20% when executing the MOVE PARTITION statement compared to when executing tabulation queries only. There are also many times at which the disk busy rate reaches 100%. It is believed that this leads to delays in query data reading, impacting transactions.

As mentioned earlier, the disk load increases for index rebuilding compared to data transfer. There is therefore a possibility of delays to MOVE PARTITION processing also if index writing takes place at the same time as data reading using queries for the SATA disk. Figure 6-19 shows the processing time as relative values when executing the MOVE PARTITION statement with a node using tabulation processing with respect to the normal MOVE PARTITION statement processing time of 1.

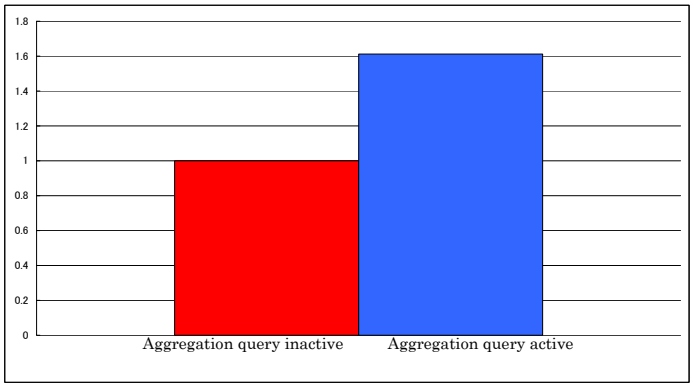


Figure 6-19 Time comparison when running aggregate queries

The graph reveals that the time taken is approximately 1.6 times longer compared to normal.

These results indicate the need to pay attention to access for the transfer destination data when executing the MOVE PARTITION statement.

6.1.4. Summary of ILM using MOVE PARTITION

ILM using the MOVE PARTITION statement does not require any particular attention to physical storage design. The period for data transfer can also be altered to suit the customer's system requirements and setup. It is also possible to compress data simultaneously while defragging tables or moving partitions. In other words, ILM is compatible with a wide range of requirements.

ILM using the MOVE PARTITION statement, however, requires attention to the following points.

- FC disk load

  As described in "6.1.3.2 Executing MOVE PARTITION with a Node Performing Online Processing," the FC disk load increases for data transfers if MOVE PARTITION is executed simultaneously with a node used for online processing. We recommend executing the MOVE PARTITION statement or distributing the FC disk RAID groups as described in "6.1.3.3 With Partitions Divided Among Different FC Disks" after analyzing the disk load status using a tool such as iostat.

- SATA disk load

  As described in "6.1.3.1 Executing MOVE PARTITION with a Node not Performing Online Operations," the SATA disk load is extremely high when rebuilding indexes if indexes are moved while executing MOVE PARTITION. SATA disks tend to be poorly suited to multi-processing, and so attention must be paid to access to SATA disks when executing MOVE PARTITION.

- CPU use

  As described in "6.1.3.2 Executing MOVE PARTITION with a Node Performing Online Processing," CPU use increases if the MOVE PARTITION statement is executed during normal online operations for a single DB setup. This makes it necessary to monitor CPU usage for online processing using tools such as sar or vmstat and to execute MOVE PARTITION only when online operations will not be affected.

ILM schedules can be established using MOVE PARTITION based on these points to enable more efficient data management.

Note that, as described in "6.1.2 Resource Use Status when Executing MOVE PARTITION," the time required for MOVE PARTITION depends on data volumes and whether index rebuilding is used.

A white paper on data segment compression function validation is available at the following URLs.[6]

[6] [SPARC Enterprise and Oracle Database 11g Performance Validation]
http://primeserver.fujitsu.com/sparcenterprise/news/article/08/0527/
http://www.oracle.co.jp/solutions/grid_center/fujitsu/

6.2. ILM Using RAID Migration

Oracle Database ILM is achieved by transferring data by changing the table space using MOVE PARTITION and allocating data with low access frequency to high-capacity, low-cost disks. As shown by the results of validating MOVE PARTITION described earlier, we confirmed that MOVE PARTITION subjects the database server to CPU loads, since Oracle process data is read from the disk and data is written to the transfer destination disk before indexes are rebuilt. This means operations are affected to some extent when implementing ILM using MOVE PARTITION.

The ETERNUS DX RAID migration function described in “3.2.1 RAID Migration” can be used to move logical volumes created on a RAID group within other disk storage systems. This makes it possible to minimize effects on ILM operations by transferring data from high-speed disks to low-speed disks without using database server resources.

This section describes the procedures for ensuring efficient ILM by combining Oracle Database ILM with the ETERNUS DX RAID migration function based on the results of this validation.

6.2.1. Efficiency and Physical Design Using RAID Migration

This section describes efficient ILM operation procedures using RAID migration and physical database design for such ILM operation.

With MOVE PARTITION for ILM, the table space can be moved in partition units. This means the configuration of the table space in which data exists before the transfer and how logical volumes are used are not issues from the viewpoint of ILM operations. However, since RAID migration transfers data in storage logical volume units, ILM operations must be taken into consideration for physical design, and design must ensure that dedicated logical volumes are provided for data manipulated by ILM. Figure 6-20 shows an example with the 2007 and 2008 data model used in this validation. In this case, the amount of data for one year forming the unit of ILM is estimated and a logical volume created to contain that data. This logical volume can contain only the data for that one year.

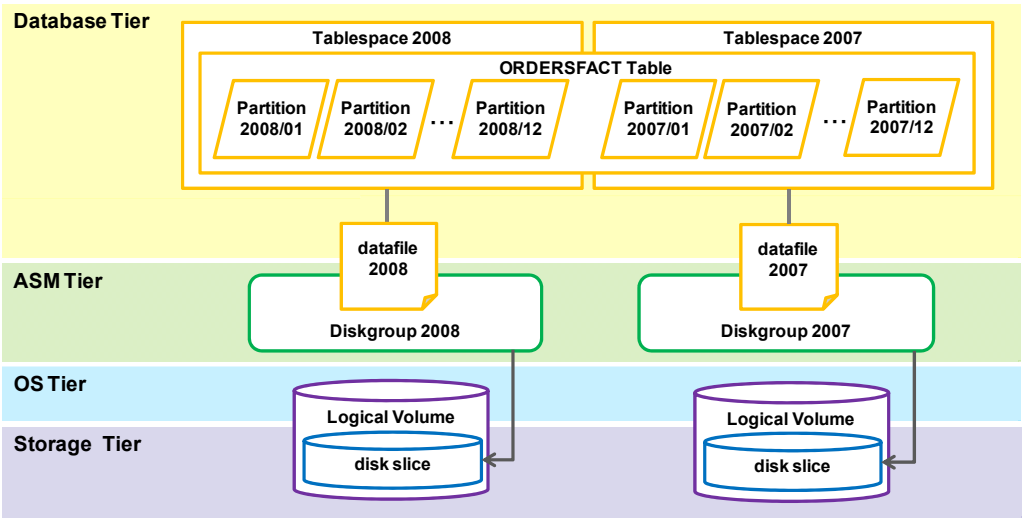


Figure 6-20 Physical Design of using RAID Migration

Although the example above shows data for 12 months in a single tablespace and a single data file, monthly table spaces can be used in actual practice, enabling data files to be created for individual months. Similarly, although ILM is run annually, logical volumes and disk groups can be provided for individual months, allowing operations with RAID migration performed 12 times. However, this requires 60 disk groups (12 months × 5 years) just for ILM disk groups, exceeding the upper limit of 63 ASM disk groups including master tables and disk groups for other areas. We recommend design that minimizes the number of disk groups, with a one-to-one correlation of ILM data transfer units to disk groups, to permit future system expansion.

Next we consider various operation procedures. The model used in this validation involves the retention of the most recent whole year of data for online operations together with 5 years of past data covered by tabulation operations. Since ILM data is moved in 1-year increments, logical volumes are provided for each year, as described above. The data for the most recent year (2009) involves logical volumes on the RAID groups formed using FC disks. The five years of earlier data is allocated to the five logical volumes for each year created on the RAID groups formed using SATA disks.

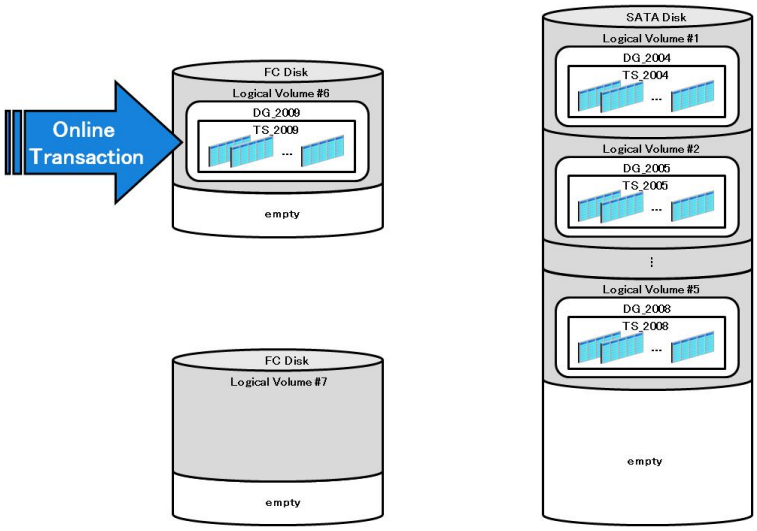


Figure 6-21 Initial placement of ILM target Data

This comprises the initial allocation. The disk group, the table space, and partition for the next year (2010) can be allocated to the logical volume (#7) provided. Online operations increasingly involve the processing of 2010 data, while online operations on 2009 data decline.

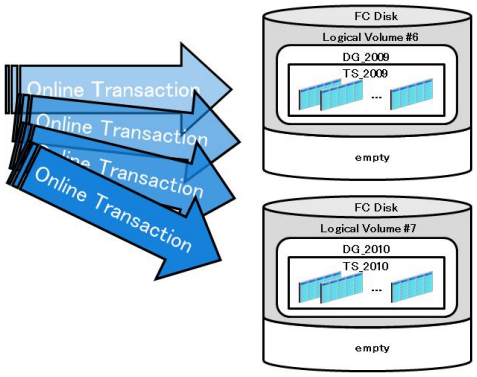


Figure 6-22 Add Disk Group, Table Space and Partitions

Once online operations apply to the 2010 partition and 2009 data is no longer used by online operations, the 2009 data can be transferred to the SATA disk. The logical volume (#6) containing the 2009 data on the FC disk is then transferred to the SATA disk by RAID migration.

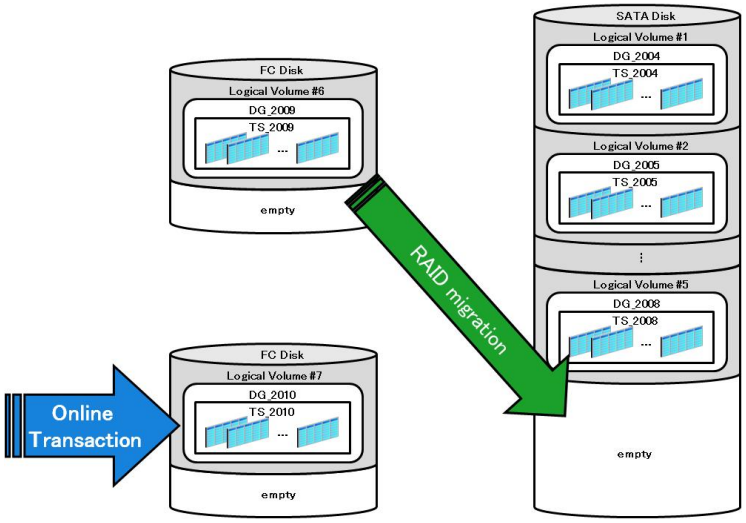


Figure 6-23 Moving 2009 Data

Providing two RAID groups from the FC disk for latest data prevents disk I/O conflicts due to online operations and RAID migration.

Since the data storage policy specifies five years, 2004 data is subject to deletion. Before deleting the 2004 data, however, we first use RAID migration to move the logical volume (#1) to the FC disk previously containing the 2009 data.

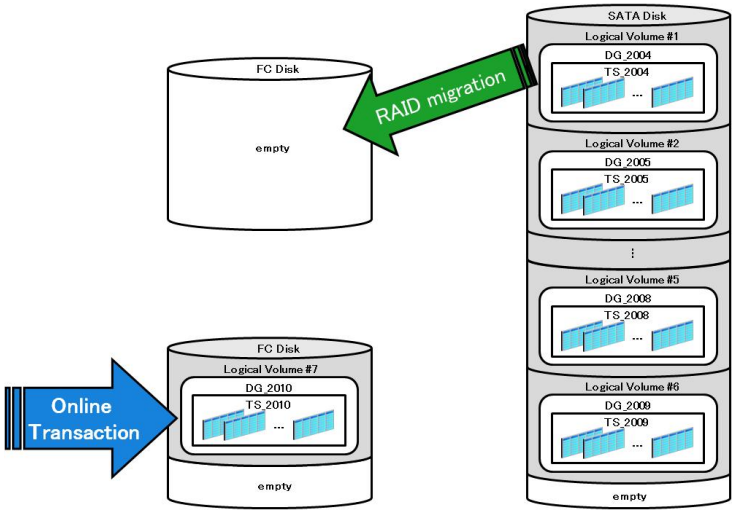


Figure 6-24 Moving Data it is deleted

The logical volume is moved to enable use of this logical volume for the next year's (2011) data after the 2004 data is deleted. If the logical volume used for the 2004 data is not to be moved, the logical volume is deleted on ETERNUS DX after deleting the 2004 data to confirm that the logical volume has been deleted with the ETERNUS multipath driver. A logical volume must then be created on the FC disk to store the 2011 data, requiring the addition of a logical volume, recognition of the logical volume by the OS/driver, and creation of slices. Transferring the logical volume previously containing the 2004 data to the FC disk eliminates the need for these procedures. Additionally,

deleting data from the FC disk is faster than deleting data from the low-speed SATA disk. The 2004 data should therefore be deleted following RAID migration.

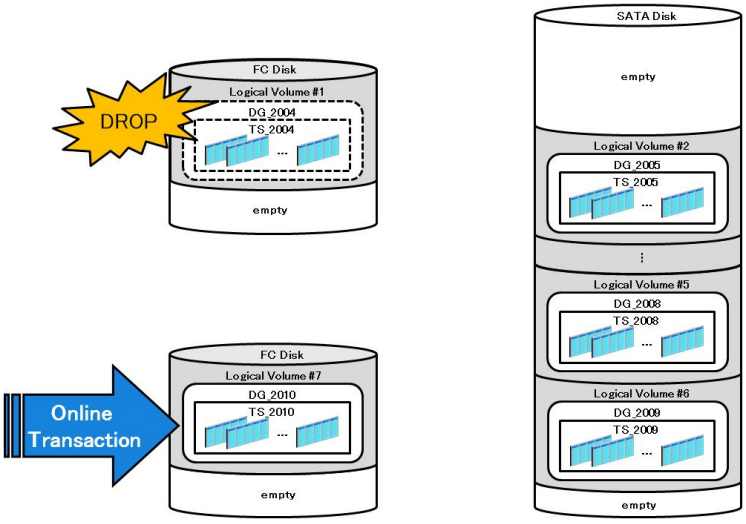


Figure 6-25 Delete the Old Data

The data is ultimately configured as shown in Figure 6-26. Logical Volume #1 is ready to serve as the logical volume for the next year's (2011) data, resulting in a configuration more or less identical to the initial allocation. The ILM can be managed by the same procedures for 2012 and subsequent years.

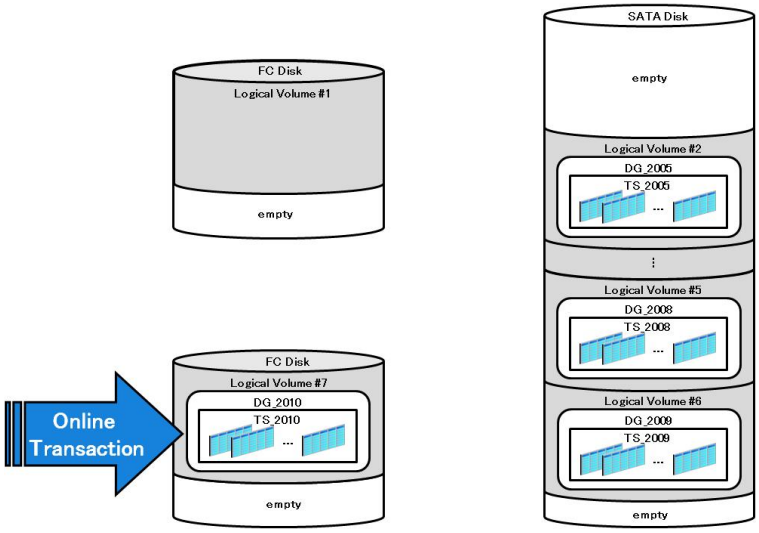


Figure 6-26 The Placement of Data after ILM

This section described the procedures used to achieve ILM using RAID migration and the various issues associated with physical database design for such ILM operation. The differences from ILM using Oracle standard functions can be summarized as shown below.

- Dedicated logical volumes are assigned to the units of data transfer in ILM.
- Two FC disks used alternately are provided to store the most recent data.
- Data exceeding its storage period is transferred to the FC disk by RAID migration before being deleted.

For more information on these procedures, refer to “9.3 ILM Procedures Using RAID Migration.”

6.2.2. Impact of RAID Migration on Operations

We examined the effects of RAID migration on operations. As shown in Figure 6-23, the most recent data was subjected to RAID migration during online operations.

Figure 6-27 shows online operational throughput and response times for normal operations (without RAID migration). It gives values as relative values, with mean throughput and response time defined as 1.

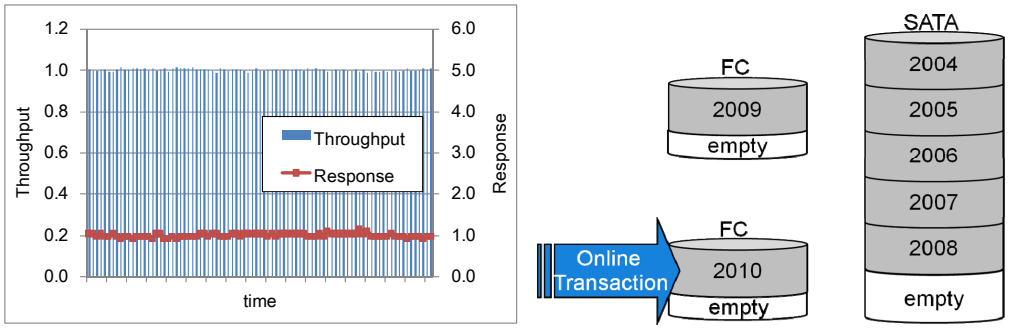


Figure 6-27 Throughput and response time for normal online processing

Figure 6-28 shows online operational throughput and response times during RAID migration to transfer 2009 data on the FC disk to the SATA disk. It gives values as relative values, with normal mean throughput and response time defined as 1.

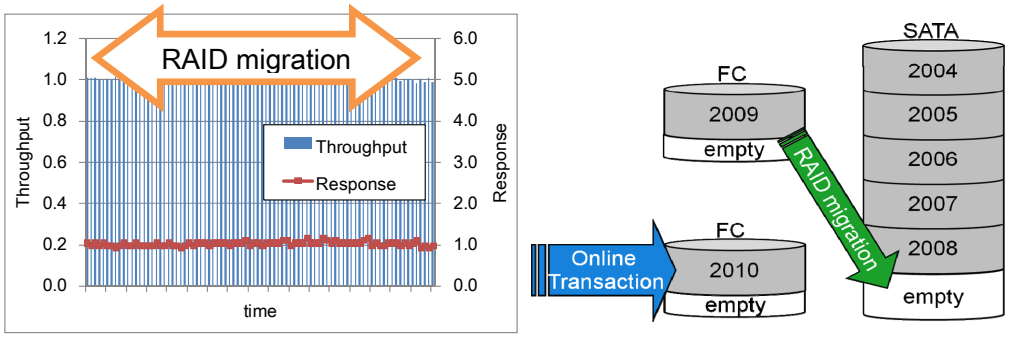


Figure 6-28 Effect for Online Operations when execution RAID Migration to move 2009 data to SATA Disks

We see that RAID migration has no effect on online operations, with both throughput and response time for online operations remaining at a relative value of roughly 1 when using RAID migration.

Figure 6-29 shows online operational throughput and response times during RAID migration to transfer 2004 data on the SATA disk to the FC disk. It gives values as relative values, with normal mean throughput and response time defined as 1.

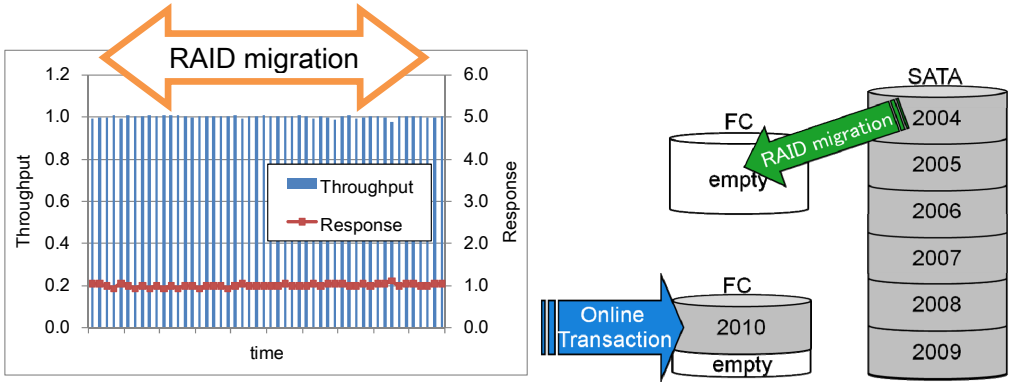


Figure 6-29 Effect for Online Operations when execute RAID Migration to move 2004 data to FC Disks

As when transferring 2009 data to the SATA disk by RAID migration, online operational throughput and response times remain unaffected.

We next examined CPU usage and disk busy rates for the database server. We first examined CPU usage and disk busy rates for normal operations.

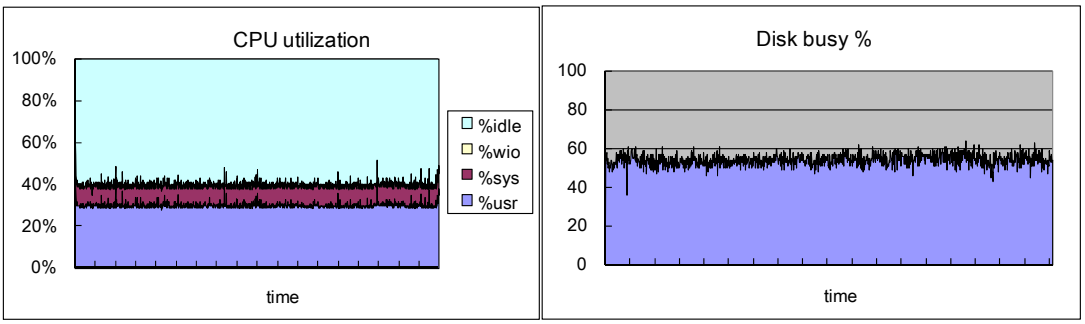


Figure 6-30 CPU usage and Disk Busy Percent of Database Server for normal online processing

CPU usage is approximately 40% overall, while the disk busy rate is around 50% to 60% with a steady load applied.

Figures 6-31 and 6-32 show the database server CPU use and disk busy rates during RAID migration. The load does not differ from normal operations in either case.

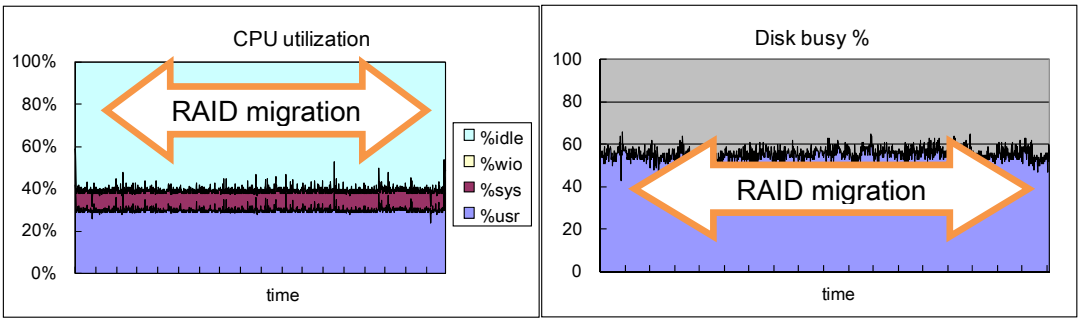


Figure 6-31 CPU usage and Disk Busy Percent when execute RAID Migration to move 2009 Data from FC Disks to SATA Disks

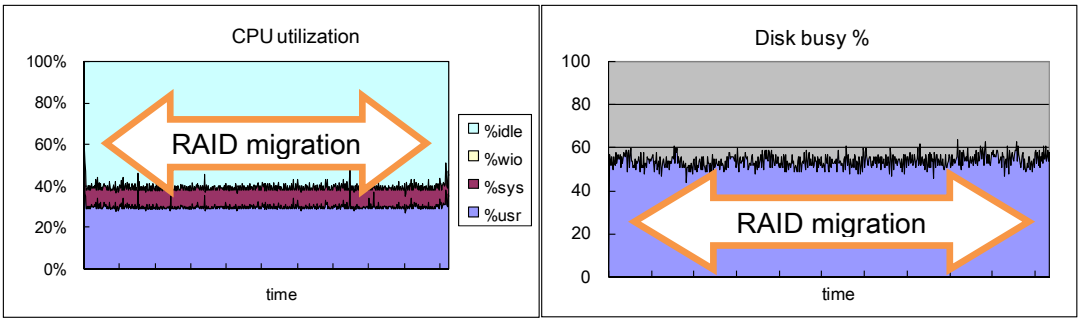


Figure 6-32 CPU usage and Disk Busy Percent when execute RAID Migration to move 2004 Data from SATA Disks to FC Disks

CPU use does not differ from normal operations, since the database server CPU is not used for RAID migration. Similarly, disk busy rates remain the same as for normal operations, since the disk used for online operations is separate from the disk used for RAID migration.

Reference (i): Server CPU use during RAID migration

The database server CPU load was examined when moving one year's worth of data (approximately 130 GB) from the FC disk to the SATA disk by RAID migration with operations suspended. The database server CPU use during RAID migration shown by Figure 6-33 indicates the database server CPU is not used during RAID migration.

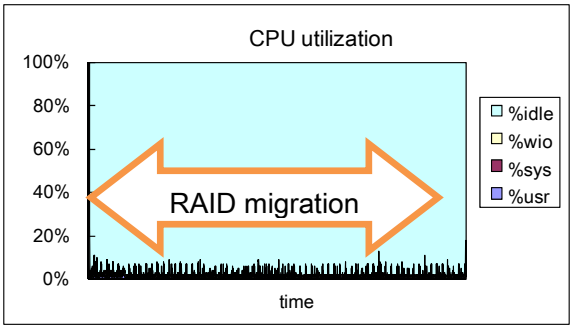


Figure 6-33 CPU usage of Database Server while execute RAID Migration

Reference (ii): With only one RAID group on FC disk

Using a single FC disk will result in an impact on online operations, due to disk I/O conflicts between online operations and RAID migration. In the validation model used here, disk busy rates are around 50% to 60% for normal operations, but reach 100% during RAID migration, confirming that disk bottlenecks affect operations.

Based on these results, we recommend using at least two FC disk RAID groups, as in this validation.

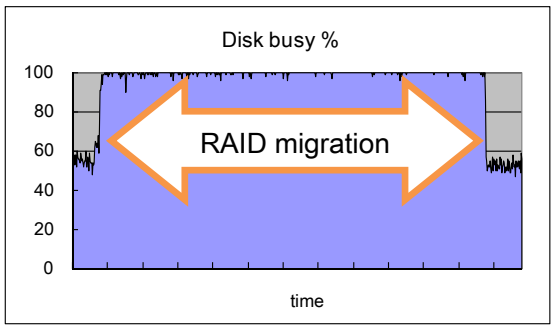


Figure 6-34 Influence of executing RAID Migration if you have a RAID group

We next examined the effects on tabulation operations. We examined individual query performance for use as tabulation processing base datum values. Figure 6-35 shows CPU usage and disk busy rates when executing individual queries.

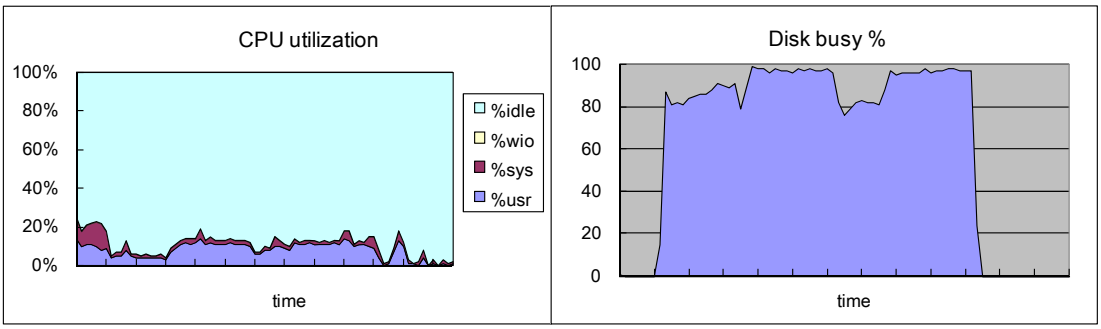


Figure 6-35 Single aggregate query performance

The SPARC Enterprise M3000 used for tabulation processing has a total of eight threads, since it uses 1 CPU × 4 cores × 2 threads. Tabulation processing uses single multiplexing. Thus, while CPU use is approximately 10%, the load more or less fully uses one thread. The disk busy rate is at least 80%, indicating extremely high disk loads.

We examined the impact on operations of moving the 2009 data from the FC disk to the SATA disk by RAID migration while using this tabulation processing in sequence on the 2008 data. Figure 6-36 shows the relative values for query response when the query response for individual use is defined as 1.

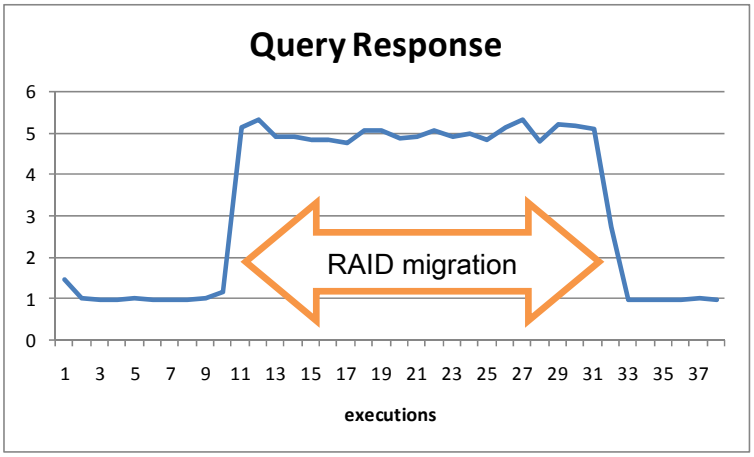


Figure 6-36 Aggregate query response while execute RAID Migration

Clearly, query response suffers serious delays during concurrent RAID migration. The query response returns to its original value once RAID migration ends. We next examined CPU usage and disk busy rate.

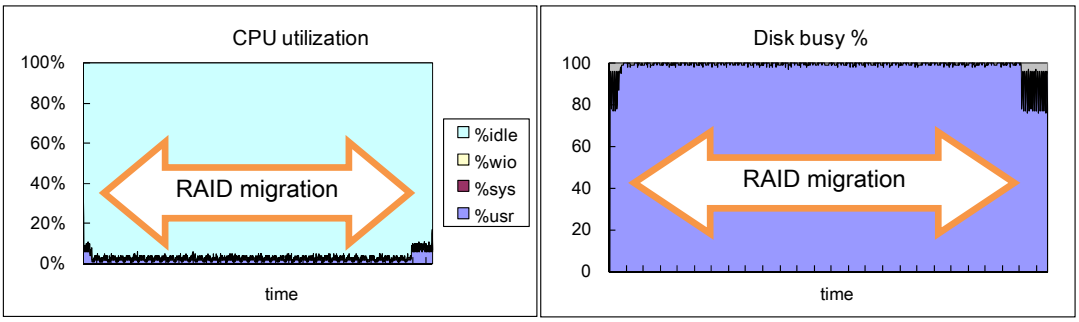


Figure 6-37 Influence for aggregate query by execution RAID Migration

CPU use drops immediately after the start of RAID migration, returning to original levels after RAID migration is complete. The disk busy rate reaches 100% immediately on starting RAID migration. These results are believed to point to a disk bottleneck; reading from the disk is suspended at the database due to the load imposed by tabulation processing reading from the disk and writing from data using RAID migration. The time required for the RAID migration here is approximately 40% longer.

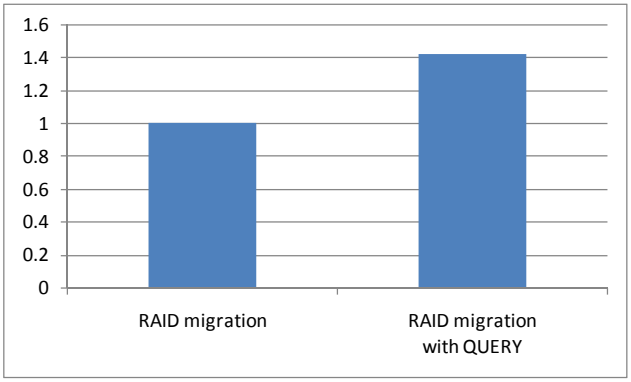


Figure 6-38 Time required execution of RAID Migration

There is no impact on online operations, since the RAID group used for RAID migration is separate from the RAID group used for online operations. Tabulation processing is affected by RAID migration, since the RAID group used for RAID migration is not distinct from the RAID group used for tabulation. Additionally, the RAID migration itself takes longer. We recommend performing RAID migration from the FC disk to the SATA disk when the SATA disk is not being accessed or when operations accessing the SATA disk are suspended.

6.2.3. Time Taken for RAID Migration

Here we explain the time taken for RAID migration. In this validation, RAID migration was used to transfer 2009 data from the FC disk to the SATA disk and to transfer 2004 data from the SATA disk to the FC disk. Figure 6-39 shows the relative values for the time taken for RAID migration.

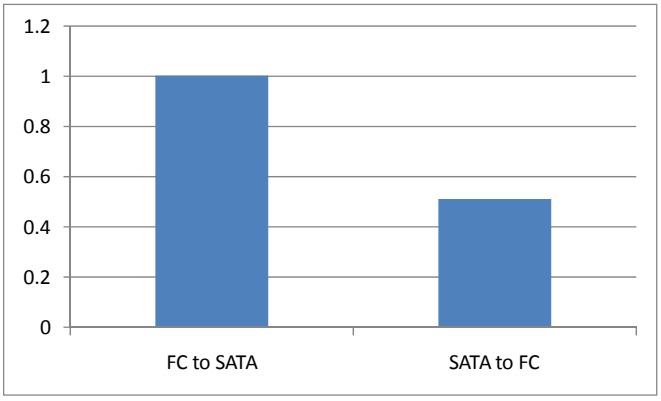


Figure 6-39 Time required execution of RAID Migration

RAID migration from the SATA disk to FC disk, which involves writing to the FC disk, takes approximately half the time required for migration from the FC disk to the SATA disk, which entails writing to the SATA disk. Performance requirements for SATA disks tend to be lower than for FC disks due to capacity-related benefits. The time required for RAID migration therefore varies significantly, depending on the type of disk to be transferred to.

"6.1.2 Resource Use Status when Executing MOVE PARTITION" shows that the time required for ILM operations varies with the configuration and number of indexes involved in the MOVE PARTITION. For RAID migrations, since logical volumes are moved within storage, the time required for RAID migration is determined by the RAID configuration and volume size, regardless of the internal logical volume configuration, provided the RAID migration disk I/O does not overlap that of operations.

6.2.4. Summary of ILM Using RAID Migration

To minimize the impact on operations when moving data, move data using the ETERNUS DX RAID migration function. RAID migration does not use the database server CPU, since all processing takes place within storage. Disk I/O conflicts between online operations and RAID migration can be eliminated by deploying and alternating between two FC disk RAID groups. Tabulation processing reading of data allocated to a SATA disk will generate conflicts due to writes associated with RAID migration.

With ILM, the data allocated to SATA disks is data with a low access frequency. This means RAID migration should be scheduled to avoid concurrent operations with tabulation processing or that operations involving SATA disk access should be suspended while RAID migration occurs.

RAID migration is clearly effective in moving data in ILM, but with the following caveats:

- The system must be designed specifically for ILM operations.

  For more information on design procedures, see “6.2.1 Efficiency and Physical Design Using RAID Migration.”

- Space cannot be reduced using storage functions.

  For more information on methods for reducing space, refer to “9.3.2 Support for Data Reduction.”

- LUN concatenation[7] cannot be used to expand space.

  For more information on ways to expand space, refer to “9.3.1 Support for Increased Data Volumes.”

6.3. Effects of Disk Performance Differences on Tabulation Processing

ILM involves moving data accessed less frequently from high-speed FC disks to SATA disks. Given below are assessments of potential effects on tabulation processing response.

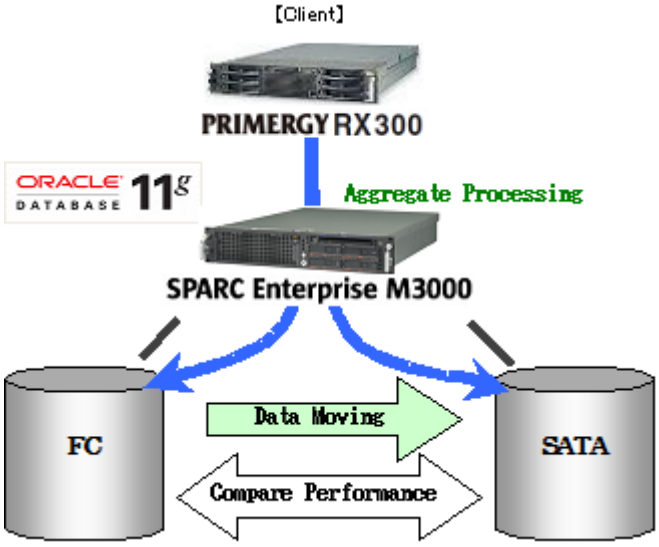


Figure 6-40 Summary of Aggregate Processing Performance validation

[7] LUN concatenation: A technique for increasing existing logical volume capacity by separating unused space to create a new logical volume and linking it to an existing logical volume.

Figure 6-41 is a graph comparing processing times for tabulation processing on FC and SATA disks for a previous month sales comparison (Q1) and sales by employee (Q2).

The tabulation processing for previous month sales comparison (Q1) on the SATA disk takes approximately 1.2 times the time required with the FC disk.

However, for tabulation processing for sales by employee (Q2), there is no significant difference in time taken between the FC or SATA disks, likely because processing times for sales by employee involve CPU-intensive queries such as GROUP BY and ORDER BY, minimizing the effects of differences in disk performance.

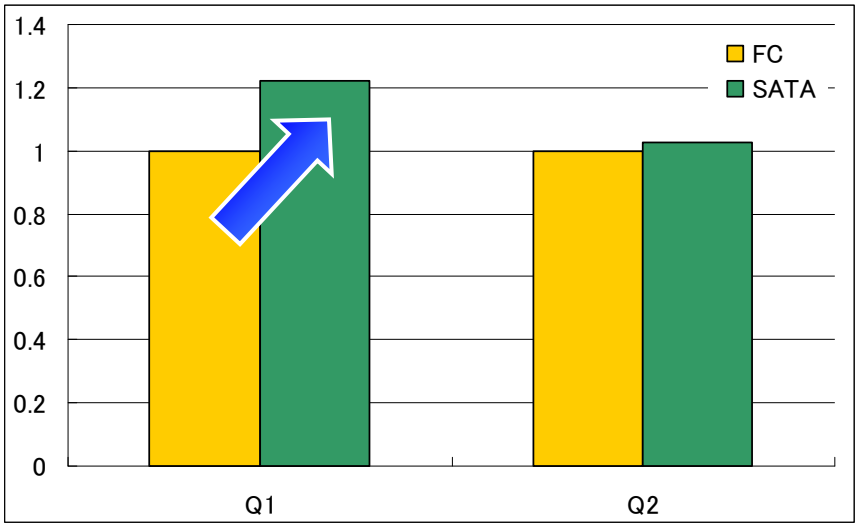


Figure 6-41 Compare to aggregate query performance of each Disk type

6.4. Summary of ILM Validation

ILM may use RAID migration (ETERNUS DX function) or MOVE PARTITION (Oracle partition function).

The characteristics of these two methods are shown in Table 6-1 below, based on the results of the validation undertaken to date.

Table 6-1 ILM method characteristics

	RAID migration	Move partition
Advantages	• Minimal impact on operations when running ILM • Time required for ILM can be predicted.	• Fragmentation can be prevented using ILM. • Space requirements can be reduced.
Disadvantages	• Requires dedicated ILM database space design.	• ILM uses CPU resources. • Time is required for index rebuilding.

These characteristics can be summarized as guidelines for selecting methods.

[RAID migration]

The major advantage of RAID migration is that it does not consume OS resources (CPU), since it operates within storage. This means ILM can be run without imposing loads on operations.

> **RAID migration is extremely effective in minimizing the impact on data transfer operations using ILM.**

[MOVE PARTITION]

MOVE PARTITION is a feature of OraclePartitioning, and does not require dedicated ILM storage design for use with ILM.

It reduces the space needed, eliminates fragmentation, and compresses segments to improve SATA disk access performance.

> **MOVE PARTITION is an effective tool in implementing flexible ILM.**

7. Backup

Backing up data is an essential part of database operations. Based on the following techniques, it also helps reduce storage costs and power consumption for data backup.

- Disk cost savings, reductions in numbers of disks, and power consumption savings by using high-capacity, low-cost SATA disks for backup destination disks.
- Using Eco-mode to reduce power consumption by using disks only when required for backup. (For more information on reducing power consumption, refer to “9.1 Using Eco-mode to Reduce Power Consumption.”)

When a disk is shut down in Eco-mode, it takes approximately 1 minute to restart, resulting in potential backup delays. However, Eco-mode scheduling can be used to set a specific daily time (disk operation time) for backups, which addresses this problem by starting the disk before the backup and shutting it down once again afterwards.

If you choose to use a SATA disk instead of an FC disk as the backup destination disk, you must consider the potential impact on backup performance.

This section discusses an assessment of backup performance with FC and SATA backup disks.

7.1. OPC Backup Performance

This validation uses the snapshot high-speed copying (One Point Copy) function of AdvancedCopyManager. One Point Copy (OPC) is a function for copying operating volumes to backup volumes at times scheduled by the user.

The following sections discuss the results of an assessment of backup performance using OPC.

7.1.1. Single Volume Backup Performance

We compared the backup performance for four different patterns of OPC backups of single volumes comprising approximately 130 GB on FC and SATA disks to FC and SATA disks.

The following page shows the validation patterns and outline diagram.

Table 7-1 Validation patterns

Validation No.	Backup source	Backup destination
1	FC	FC
2	FC	SATA
3	SATA	FC
4	SATA	SATA

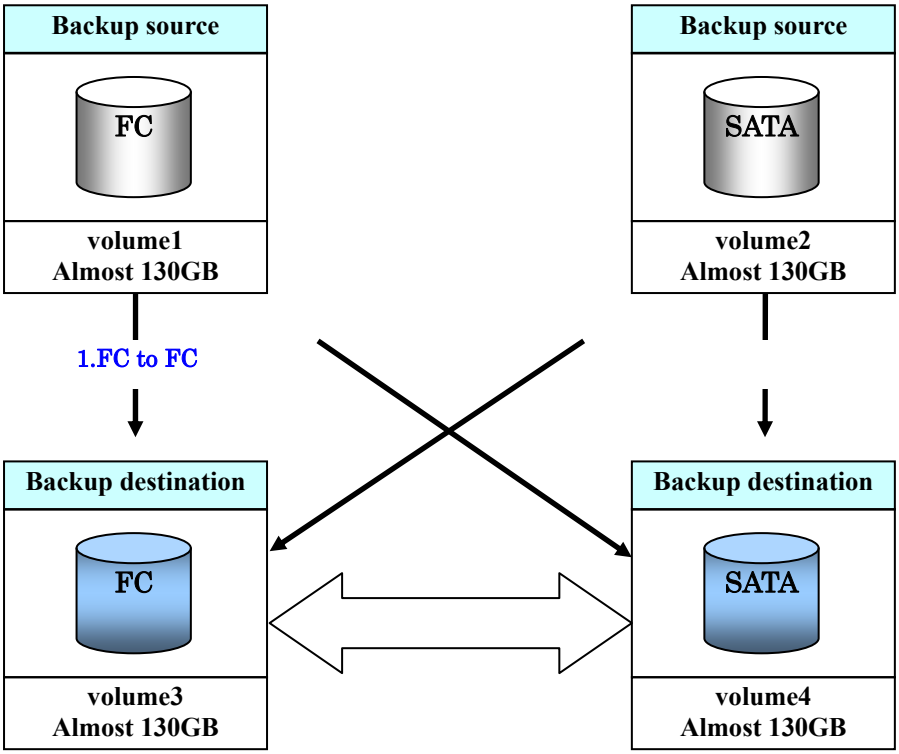


Figure 7-1 Summary of Volume Backup Validation

The validation results are shown below.

Figure 7-2 shows that backup acquisition times for the backup destination disk are virtually identical whether the backup source disk is an FC or SATA disk.

If the backup destination is a SATA disk, backup acquisition times are roughly 2.2 times the time required with an FC disk. The amount written (Figure 7-3) to the SATA disk per unit time is approximately 45% that of the FC disk.

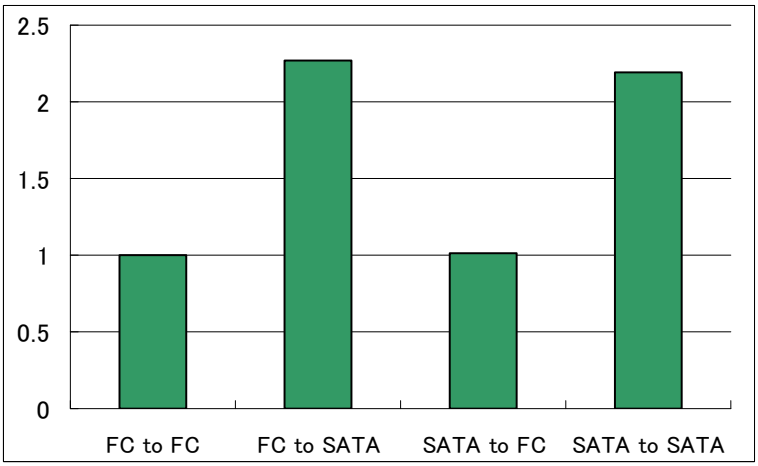


Figure 7-2 Compare to require time to get volume backup

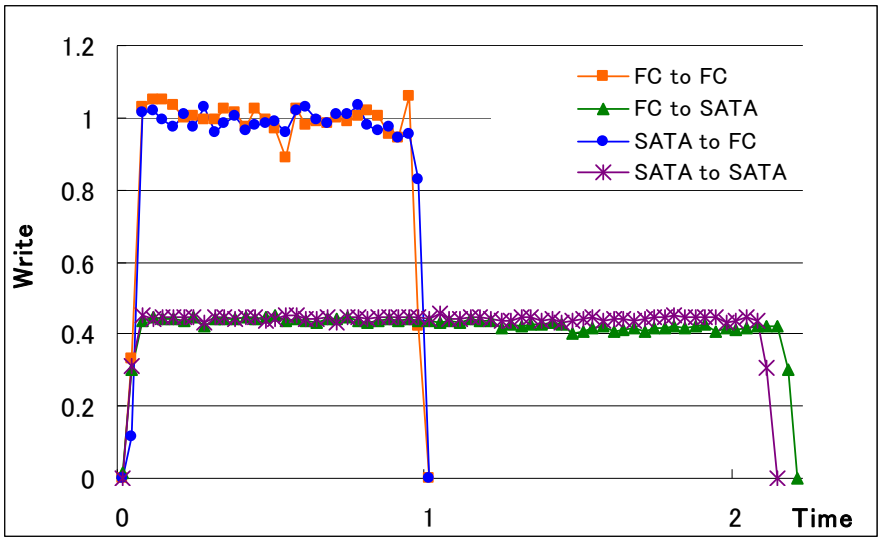


Figure 7-3 Compare to required time and amount written when volume backup

7.1.2. Performance for Whole Database Backup

We examined the effects on acquisition time attributable to differences in backup destination disk type (FC or SATA) and use of multiple backups when backing up an entire database.

The backup source used in this validation consists of one volume in the FC disk RAID group and three volumes in the SATA disk RAID group (total size of approximately 500 GB).

This backup source is used for OPC multiple (1, 2, 4) backups to a backup destination formed of 1 RAID group and four volumes.

We examined the effect of multiple writing to multiple volumes (backup destination) within the 1 RAID group on backup acquisition time.

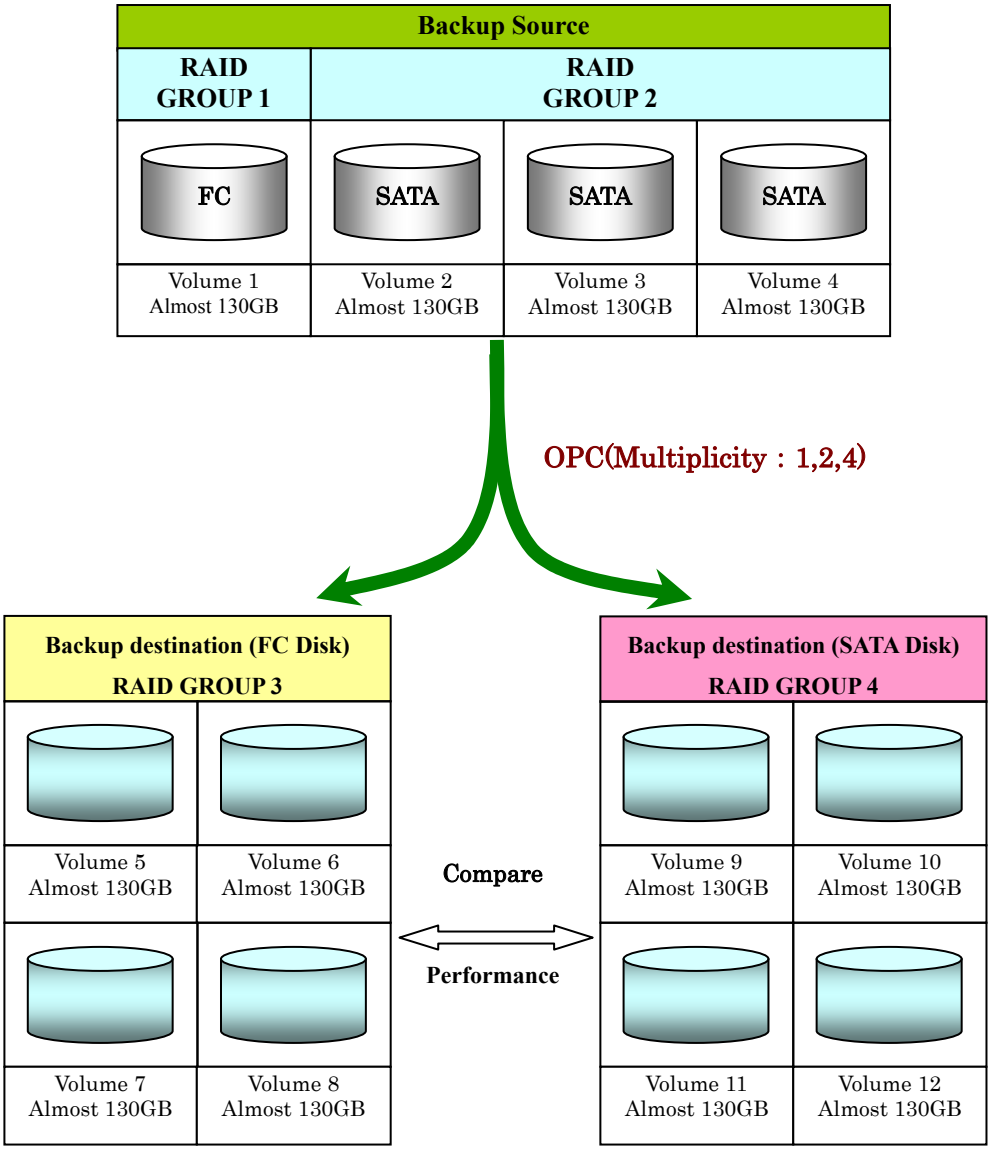


Figure 7-4 Summary of whole Database Backup validation

The outline diagrams for the multiple backups are shown below.

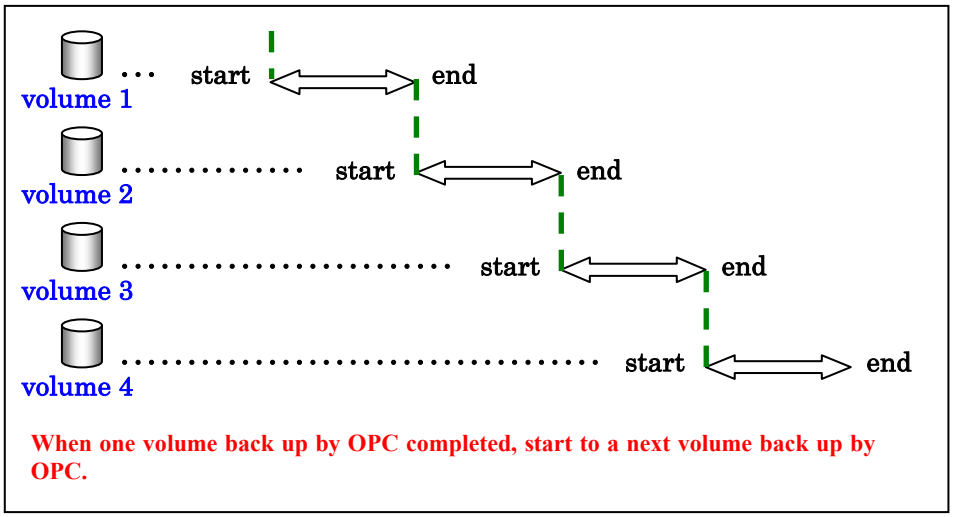


Figure 7-5 Summary of multiple backup

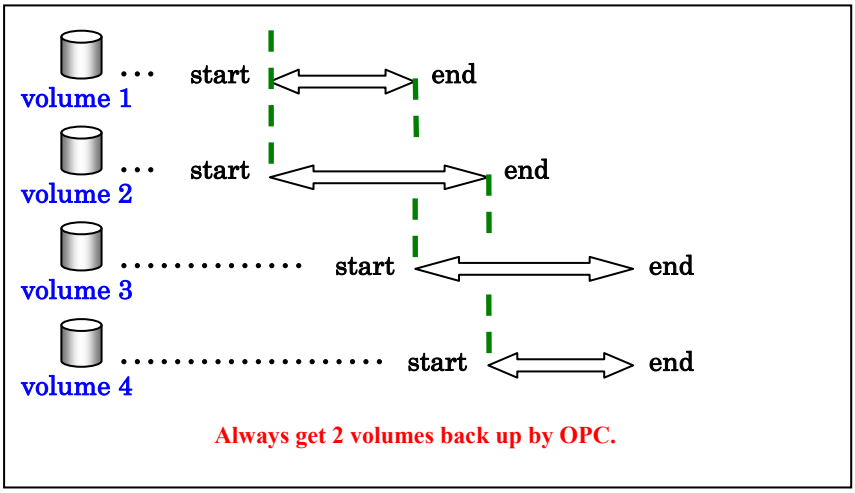


Figure 7-6 Summary of multiple backup

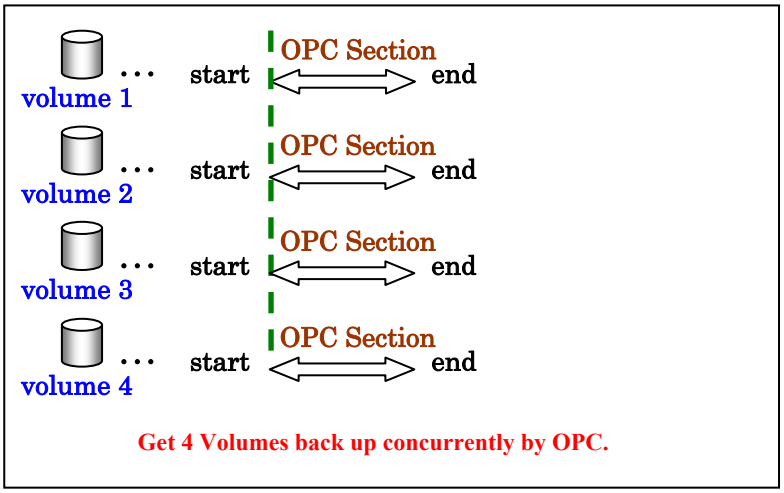


Figure 7-7 Summary of multiple backup

The validation results are shown below.

For a single backup, Figure 7-8 shows that backup acquisition times are approximately twice as long with a SATA disk as for an FC disk.

When using multiple OPC backups for a 1 RAID group, acquisition times tend to increase compared to single backup regardless of multiple processing for increased multiplicity for both FC and SATA disks. A 4-multiple backup to a SATA disk takes roughly three times longer than an FC disk 4-multiple backup.

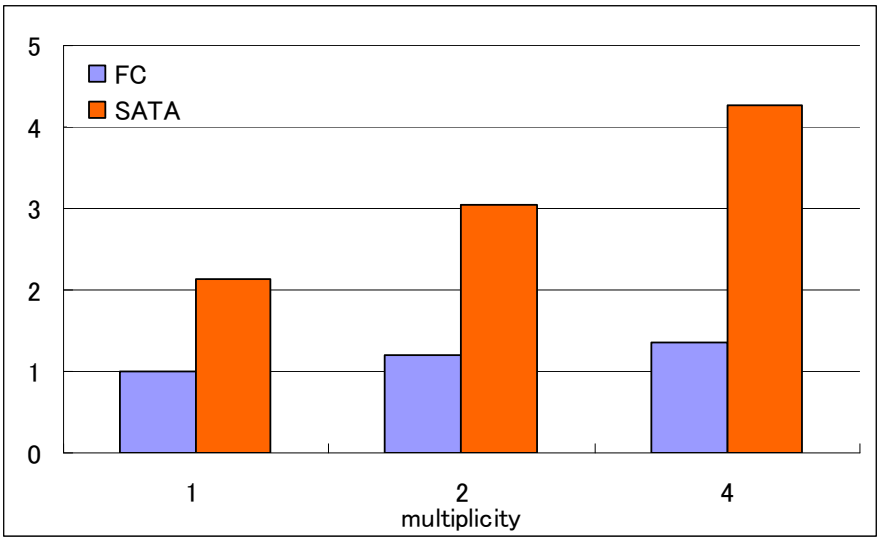


Figure 7-8 Compare to required time by each multiplicity

7.2. Summary of Backup Validation

Backup acquisition times depend on backup destination disk write speeds, regardless of backup source disk type. Without multiple processing, backup acquisition times for SATA disks are approximately twice as long as for FC disks.

For multiple backups to multiple volumes located within the same RAID group, backup acquisition times are actually longer than without multiple processing, regardless of disk type.

We recommend serial backups without multiple processing when backing up multiple volumes located within the same RAID group.

If multiple backups must be performed, take steps to ensure that multiple writes do not occur within the same RAID group by dividing and allocating the backup destination volume to multiple RAID groups.

8. Summary

This validation clarified methods for moving data in ILM, accounting for the particular characteristics of FC and SATA disks and revealing the storage configurations best suited to ILM.

MOVE PARTITION is better suited for moving data if the focus is on flexibility and ease of use. ETERNUS DX RAID migration is better in minimizing the impact of data transfer on operations. The selection between the two should be based on customer requirements.

We also determined design and operating methods for minimizing the impact on operations based on the performance characteristics for moving data. Nearline SATA disks offer lower processing performance than FC disks, making ILM storage design and size of operation I/O important factors to be considered. For systems involving numerous I/O operations, storage design must account for the access volume during and after data transfers to minimize effects on operation response for achieving ILM.

The validation also highlights important points associated with storage configurations when using Eco-mode or nearline disks to reduce power consumption for backup, a vital aspect of database systems.

Using nearline SATA disks has a major impact on backup acquisition times. This means measures like providing multiple backup destination disks are needed if acquisition time is crucial. Storage design must balance backup requirements against cost considerations.

As described above, we established effective designs and methods to reduce storage costs using ILM, while accounting for the characteristics of different disk types.

ILM involves a large number of predictable requirements, including storage design and performance characteristics for different disk types. Fujitsu ETERNUS DX offers flexible support for a wide range of storage requirements by enabling use of different disk types within a single package, providing integrated management, and combining functions to minimize the impact on operations when moving data simultaneously.

A database system incorporating ILM using Oracle Database, Fujitsu SPARC Enterprise, and ETERNUS DX can cut storage costs and power consumption while retaining high performance and reliability.

9. Appendix

9.1. Using Eco-mode to Reduce Power Consumption

The Eco-mode available with ETERNUS DX allows users to select and shut down specific disks for individual RAID groups. Eco-mode shuts down disk operations to minimize the power needed to operate the disk itself and associated power requirements, such as air-conditioning. ETERNUS DX offer significantly lower operating costs than other storage systems—benefits that become considerable with growing storage requirements. The ETERNUS DX hardware itself is designed to minimize power consumption, even without Eco-mode.

One possible application for Eco-mode is disk-to-disk backups. The backup target disk in disk-to-disk backups is normally accessed only for backups.

ETERNUS DX can define RAID groups for backup target capacity and use Eco-mode control to operate the RAID group for backups only. This helps reduce power consumption for nearline disk drives and reduces air-conditioning needs.

ETERNUS DX Eco-mode MAID power specifications

Measured for 500 GB/7,200 rpm nearline disk drive

Power when operating: 19.3 W

Power when stopped: 4.7 W

ETERNUS DX400 example

ETERNUS DX440 configuration:

275 nearline disk drives operating 16 hours per day with Eco-mode enabled

125 FC disk drives operating without Eco-mode

Additional 4 Gbit/s FC ports × 16, controllers × 2, 32 GB cache

Power consumption (normal): 8,375 kWh × 13.75 yen = 115,156.25 yen/month

Power consumption (Eco-mode): 7,037 kWh × 13.75 yen = 96,758.75 yen

Predicted energy cost savings: 18,397.5 yen/month, 220,770 yen/year

Note: Excluding system cooling costs.

9.2. Scheduled Stoppage Using Power Interlock Function

CO_2 reductions now constitute a significant issue in modern IT system operations.

SPARC Enterprise incorporates various measures to improve energy savings and reduce CO_2 emissions. Cooling uses air ducts for centralized cooling of the hottest components (CPU

and memory). Cooling efficiency is increased by splitting the package into two cooling groups. The speed of the cooling fans can be controlled to achieve power savings by reducing cooling fan speeds.

Fujitsu actively seeks to reduce energy consumption for entire systems, in addition to energy savings for individual devices such as servers and storage units. Two such measures are the Power Interlock Function and Scheduling Function. These link the power supply control for the server and peripheral devices and schedule system operations to reduce power consumption by automatically switching off power to devices at times when system operations are not required—for example, at night or during holidays. This helps reduce CO_2 emissions and energy costs.

The SPARC Enterprise M4000, SPARC Enterprise M3000, and ETERNUS DX400 used in this validation include Remote Cabinet Interface (RCI) as a power interlock function interface. The power supply to other devices can be shut off automatically in sync with the master server power supply by connecting them with RCI cables.

The SPARC Enterprise M9000, M8000, M5000, M4000, and M3000 can also be scheduled to turn on and off device power automatically.

These two power interlock and scheduling functions can be deployed to automatically turn on or off all devices connected using RCI cables at specified times.

Shutting down the system normally entails shutting down the server, then shutting down the storage unit (the reverse of the startup procedure). Without power interlocking, this entails turning off (or on) power to each device in the specified sequence and shutting down at the end of the work day, and then starting up early the next morning for the start of the work day to avoid affecting business operations.

Configuring such automatic operations with conventional servers and storage products requires a separate power supply control unit and operation management software, as well as significant labor requirements if manual intervention is required late at night or early in the morning.

Deploying SPARC Enterprise and ETERNUS DX connected via RCI eliminates the startup and operating costs associated with the additional equipment and software typically required; it also eliminates personnel costs.

Combining Oracle Database with ETERNUS DX and SPARC Enterprise and using the power interlock and scheduling functions in conjunction with ILM improves efficiency and cuts costs and CO_2 emissions.

Example of cost savings achieved using scheduled stoppages with the configuration used in this validation

This compares the cost of running the configuration used in this validation continuously for

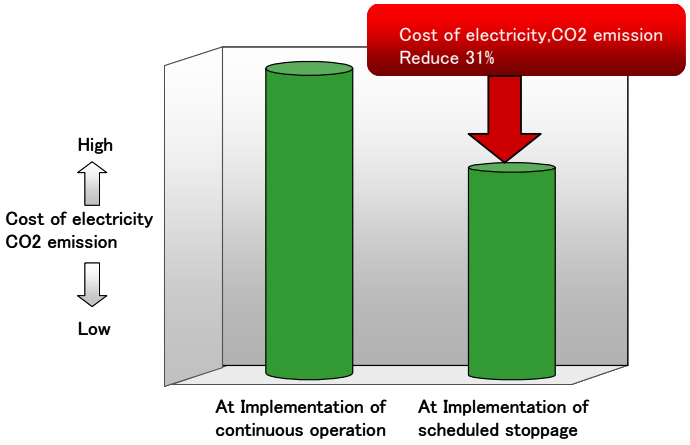
one year against the use of scheduled stoppages involving operations for 18 hours/day and shutdown periods of 6 hours/day on weekdays (240 days/year) and shutdown periods for 24 hours/day on holidays (125 days/year).

Power consumption (continuous operations): 33,857 kWh/year × 13.75 yen/h = 465,533 yen/year

Power consumption (using scheduled stoppage): 23,467 kWh/year × 13.75 yen/h = 322,671 yen/year

Predicted energy cost savings: 142,862 yen per year

Predicted CO_2 emission reductions: 4,031 kg per year



Note: Excludes equipment cooling costs.
Excludes costs associated with clients, network switches, and fibre channel switches for the setup used in this validation.

CO_2 conversion: 0.388 kg-CO_2/kWh (Fujitsu Limited calculation)

9.3. Additional Notes on ILM Based on RAID Migration

This section describes expanding or reducing space, an issue to be considered for ILM using RAID migration.

9.3.1. Support for Increased Data Volumes

Table space sizes may be inadequate and need to be expanded when online operations expand.

The following three methods are available for expanding table space size.

- Expand logical volume size using RAID migration.

- Add data files to the table space.
- Add disks to the disk group.

(1) Expanding logical volume sizes via RAID migration

ETERNUS DX provides for two methods for expanding logical volume size: RAID migration and concatenation. However, concatenation cannot be used with ILM, since volumes subjected to concatenation cannot be used with RAID migration due to their specification. Thus, logical volumes are expanded via RAID migration. The specific procedures are given below.

(i) Expand logical volume size for RAID migration from the SATA disk to FC disk.

(ii) Expand slice size on the logical volume using OS commands.

(iii) Resize (expand) disks within the disk group.

(iv) Resize (expand) the table space.

Note that the table space size should include an additional margin, since this method only allows size to be expanded for RAID migration from SATA to FC disks. If sizes remain inadequate for unexpected increases in operating volumes, use methods (2) or (3) below instead.

(2) Adding data files to the table space

The overall table space can be expanded by adding data files to the table space.

Size is expanded by this method when all of the following conditions are satisfied.

- Table space type is SMALLFILE.
- The number of disk groups is below the upper limit

The specific procedures are given below.

(i) Add a new logical volume using ETERNUSmgr.

(ii) Run grmpdautoconf to ensure the new logical volume is recognized by the OS.

(iii) Create a slice in the new logical volume.

(iv) Create a disk group.

(v) Add a data file to the table space.

Note the following points with this method.

- With RAID migration, both the original logical volume and newly added logical volume must be moved to the SATA disk.
- Adding a logical volume also increases the number of disk groups. Take care to ensure this does not lead to exceeding the limit on number of disk groups.
- Two disk groups cannot be combined into a single group. Two logical volumes will persist until deleted. To combine them into a single logical volume, move to a large space using MOVE PARTITION.

- This method cannot be used for BIGFILE type table space due to the one-data-file-per-table space limit.

(3) Adding disks to a disk group

A disk group can be expanded by adding a new disk to the disk group. The expanded free space can be used to expand the table space size.

This method can be used to expand the size if either of the two conditions below is satisfied:

- Table space type is BIGFILE.
- The number of disk groups cannot be increased (or an increase in this number is not wanted).

The specific procedures are given below.

(i) Add a new logical volume using ETERNUSmgr.

(ii) Run grmpdautoconf to ensure the new logical volume is recognized by the OS.

(iii) Create a slice in the new logical volume.

(iv) Add a slice to the disk group.

(v) Resize (expand) the data file.

Note the following points with this method.

- With RAID migration, both the original logical volume and newly added logical volume must be moved to the SATA disk.
- Operations may be affected when a slice is added to the disk group due to I/O associated with rebalancing. The disk group slice should therefore be added in a manner timed to avoid affecting operations.

[Solutions for inadequate RAID group free space]

The RAID group size can be expanded using LDE (Logical Device Expansion)[8] if there is insufficient free space within the RAID group when expanding the logical volume size or when creating a new logical volume. This applies to all three methods described above.

9.3.2. Support for Data Reduction

It may be desirable to reduce the size of logical volumes when moving logical volumes containing old data to a SATA disk if space size estimates are incorrect. Logical volume sizes cannot be reduced using storage functions; other methods must be applied.

Two methods for reducing the size of logical volumes are shown below.

- ILM using MOVE PARTITION

[8] LDE (Logical Device Expansion): Function allowing increases in memory capacity by adding new disks to the RAID group without halting operations

- ILM + MOVE PARTITION using RAID migration

(1) ILM using MOVE PARTITION

Create a new logical volume smaller than the current logical volume used with ILM, then run ILM using MOVE PARTITION with both logical volumes.

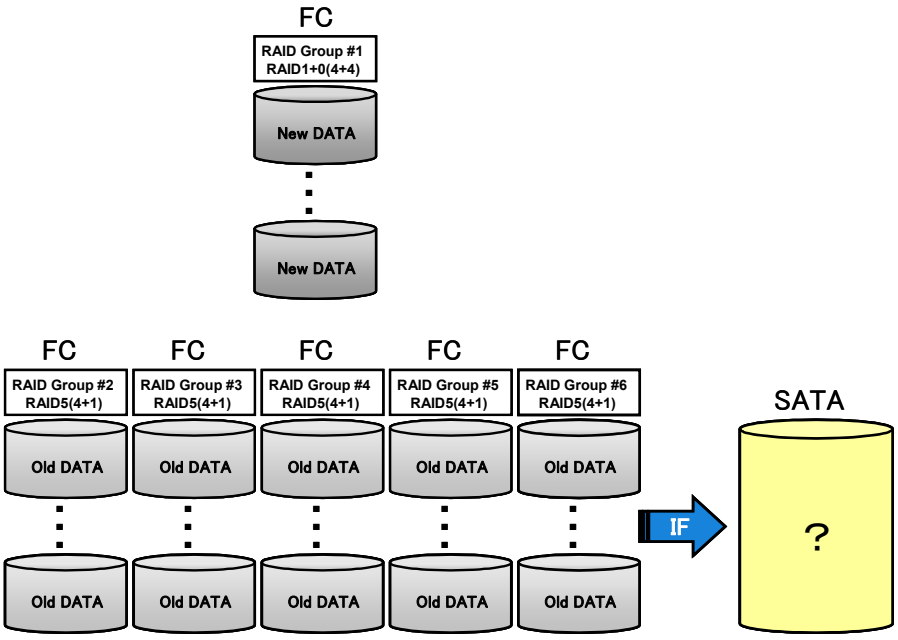
(2) ILM + MOVE PARTITION via RAID migration

Move the logical volume used with ILM to the SATA disk while minimizing the impact on operations using RAID migration. Then, use MOVE PARTITION within the SATA disk to reduce space in a manner timed to avoid affecting online operations.

9.4. Disk Cost

This section discusses ways to achieve cost reductions using SATA disks with ILM.

We examined the extent of cost savings achieved by storing old data on FC disks and on SATA disks.



The example uses a disk configuration in which all data is stored on FC disks (Table 9-1).

Table 9-1 Example disk configuration

Raid Group	Raid Level	Disk Drive	Usage
1	RAID1+0 (4+4)	FC disk 146 GB (15,000 rpm) × 8	For recent data
2	RAID5 (4+1)	FC disk 146 GB (15,000 rpm) × 5	For old data 1
3	RAID5 (4+1)	FC disk 146 GB (15,000r pm) × 5	For old data 2
4	RAID5 (4+1)	FC disk 146 GB (15,000 rpm) × 5	For old data 3
5	RAID5 (4+1)	FC disk 146 GB (15,000 rpm) × 5	For old data 4
6	RAID5 (4+1)	FC disk 146 GB (15,000 rpm) × 5	For old data 5

The capacity available for storing old data is provided by RAID groups 2, 3, 4, 5, and 6, totaling 146 GB × 20 (4 × 5 groups) = 2,920 GB. Replacing this with high-capacity, low-cost SATA disks (750 GB, 7,200 rpm) gives 750 GB × 4 = 3,000 GB for RAID5 (4+1), cutting the number of disks from 25 to 5.

In comparison with the cost of the disk drives in Figure 9-1, this configuration based on SATA disks cuts disk costs by roughly 60%.

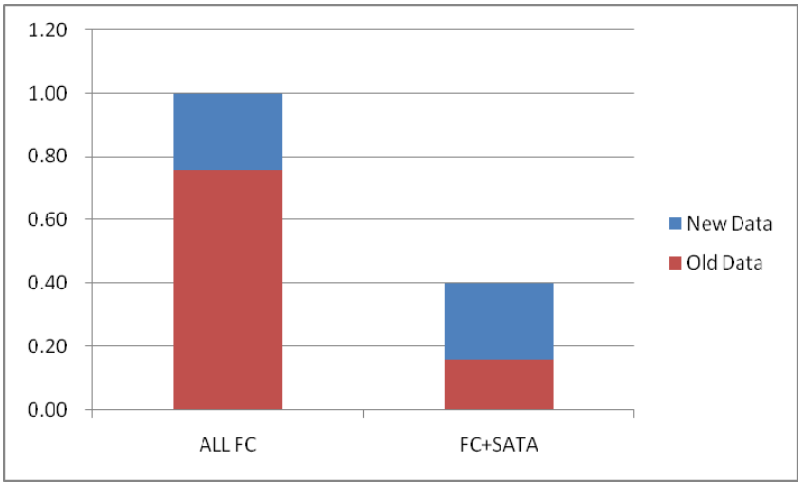


Figure 9-1 Compare to Disk drive cost

Using high-capacity, low-cost SATA disks can significantly reduce the number of disks required and storage costs.

Oracle Corporation Japan

2-5-8, Kita Aoyama, Minato-ku, Tokyo

107-0061 Japan

FUJITSU LIMITED

1-5-2, Higashi-Shimbashi, Minato-ku, Tokyo

105-7123 Japan