Oracle GRID Center

富士通 ETERNUS DX / SPARC Enterprise による Oracle Database ILM ソリューション

～ DB データ長期保存におけるストレージコストと消費電力の削減 ～

Creation Date: May 3, 2009
Last Update: April 16, 2010
Version: 1.1

目次

1. はじめに

急激な経済環境の変化の中、企業の投資に対する見直しが急速に進んでいます。IT 投資に関しても例外ではなく、IT に関するコスト構造の詳細な把握とそれぞれのコストの削減が強く求められています。一方で、企業の業務が高度にシステム化され、企業内システムが扱うデータ量が増加し続けるとともに、法制度の改正やコンプライアンスに関する要件への対応のため、業務データの長期保存の必要性が高まり、システム上に保存されるデータの量は飛躍的に増大しています。このような背景からシステムの扱うデータの増加量は 2 年で 3 倍に上るとも言われており、データを格納するためのストレージに関するコストの増加をいかに抑えていくかということは、IT コスト削減の大きなテーマの 1 つとして注目されています。

富士通と日本オラクルはデータ量の増加に伴うデータ管理/保存コスト増大への 1 つの解決策として、富士通のストレージシステム ETERNUS DX とオラクルのデータベース Oracle Database 11g による Information Lifecycle Management (ILM)ソリューションを、共同検証により確立しました。本 ILM ソリューションでは、業務データへのアクセス頻度がデータ作成から時間を経るごとに低くなることを想定した業務モデルを用いて、アクセス頻度が低くなる履歴データを、よりコストの低いストレージ領域に順次移動させることにより、情報の価値や利用され方に見合ったストレージを利用することで、情報のライフサイクルにわたる管理/保存コストの最適化を図ります。

本 ILM ソリューションでは、ストレージシステムとして 1 つの ETERNUS DX 筐体内に高速の Fibre Channel (FC)ディスク、および大容量で低価格なニアライン用のシリアル ATA (SATA)ディスク両方を搭載した、階層型のストレージシステムを構築します。データベースでは、ILM の対象となる表のデータを Oracle Database 11g の Oracle Partitioning を使用して時系列にパーティション化します。これらを組み合わせて、ILM の対象となる表の各パーティションを適切なストレージ領域に配置することで全体ストレージコストの削減と最適化を実現します。

今回の検証ではお客様環境でのデータベースの ILM 運用方法確立を目的に、ストレージ内のデータの移動方法に関して、Oracle Database の MOVE PARTITION によるパーティション移動と、ETERNUS DX の特徴的な機能である RAID マイグレーション機能でのデータ移動の検証を実施しました。2 つの方法を検証することにより、業務への影響やサーバ CPU への負荷など、さまざまな状況に応じたデータベースの ILM ベストプラクティスを提供することが可能となりました。

さらに、環境問題への配慮が企業の社会的責任として求められる中、IT システムにも消費エネルギーの削減が要求されています。富士通の SPARC Enterprise、ETERNUS DX はそれぞれ省エネルギー化を意識した設計・開発がなされ、多くの省エネルギー化機能を搭載しています。本検証では ETERNUS DX の省エネルギー化機能であるエコモードを活用して、データベースのバックアップ領域として使用しているディスクの回転を、バックアップ取得時以外に停止させることにより、性能に影響を与えることなく、ストレージシステムの消費エネルギーの削減が可能であることを確認しました。これら消費エネルギーの削減は電力コスト削減につながり、IT システムのランニングコスト削減にも寄与いたします。

本ドキュメントではデータベースのデータを長期にわたって効率的に保存、管理し、ストレージコストの最適化と消費エネルギーを削減する手法として、富士通のストレージシステム ETERNUS DX と、UNIX サーバ SPARC Enterprise、オラクルの Oracle Database 11g の組み合わせによるデータベース ILM ソリューションのベストプラクティスを紹介します。

2. 検証目的

膨大な企業データに対し、その使われ方に着目すると次のような性質を持ったデータが存在することがわかります。

- 新しいデータは高い頻度でアクセスされ、オンライン業務のようにレスポンス要件の高い処理の対象となることが多い
- データが古くなるにしたがってアクセス頻度は低下して行き、それに伴って性能要件も低くなる

注文履歴表のように、注文処理の度にレコードが挿入され、それを履歴として長期間保持する必要があるようなデータがこれに該当します。このような性質のデータをアクセス頻度や処理内容とデータの保存期間に応じて、適切なコストとサービスレベルで管理する手法を Information Lifecycle Management (ILM)といいます。

ILM では、アクセスが少なくなったデータを高速・高信頼の高価なディスクから、性能はそれほど高くはなく、大容量で低価格なニアライン用のディスクへと移動させ、データの保存コストを最適化します。この ILM の手法により全てのデータを保存するために必要なストレージ容量をより低いコストで確保することができます。また、一般的にニアライン用のディスクは高速なディスクに比べて、ドライブあたりの容量が大きいため、ニアライン用のディスクを活用することで、必要なディスクドライブの数を削減でき、エネルギー消費量の削減の効果も期待することができます。

Oracle Database ではデータが挿入された日付などの基準に従って表を分割する Oracle Partitioning 機能が提供されています。通常、一つの表は一つの表領域に格納する必要がありますが、パーティション化を行うことで、パーティション単位で格納する表領域を指定することができます。データベース内の対象となる表を、データの保存ポリシーに合わせてパーティション化し、それぞれのパーティションに要求されるサービスレベルにあった表領域へ配置することでデータベースの ILM を実現することができます。

富士通のストレージシステム ETERNUS DX は高速、高信頼の FC ディスクと、大容量、低コストのニアライン用 SATA ディスクを同一筐体内で構成できるという特長があります。Oracle Database でテーブルをパーティション化し、アクセス頻度が高い、新しいデータを高速な FC ディスクへ、アクセス頻度が低くなる履歴データをニアライン用 SATA ディスクへ配置することで、一つのストレージ筐体で ILM を実現することが可能です。さらに、ETERNUS DX は論理ボリュームを別の RAID グループへ移動する RAID マイグレーション機能を搭載しており、FC ディスクで構成した RAID グループ上の論理ボリュームをニアライン SATA ディスクへ、ストレージ内で移動することが可能です。RAID マイグレーシ

ョンでのデータ移動はストレージ内で完結し、データベースサーバのリソースを使用しないため、業務影響を最小限に抑えることができます。通常 Oracle Database では、MOVE PARTITION を使用してデータを移動しますが、ETERNUS DX と組み合わせることで、RAID マイグレーションを使用してデータを移動する方法を選択することも可能です。

この度富士通と日本オラクルは、ETERNUS DX と Oracle Database を活用した ILM を実現するための設計手法と運用手順を確立することを目的に、共同検証センターである GRID Center において、以下の検証を実施しました。

- ILM における高速ディスクからニアラインディスクへのデータの移動に関する検証
- ニアラインディスクのデータに対する業務処理の性能検証
- ILM のバックアップへの影響とバックアップ性能検証

3. 検証機器紹介

本検証にて使用した検証機器について説明します。

3.1. 富士通SPARC Enterprise

「SPARC Enterprise」は、スピード経営、業務の継続性、TCO 削減、環境貢献等、企業が抱える様々な経営課題に応えるために、メインフレーム並みの高信頼性を持ちミッションクリティカル業務に最適な SPARC Enterprise M9000, M8000, M5000, M4000,M3000 と、Web フロント業務に最適な高いスループット性能を持った SPARC Enterprise T5440, T5240, T5220, T5140, T5120 をご用意しています。

3.1.1. SPARC Enterprise M4000

【SPARC Enterprise M4000 の特長】

- メインフレームの信頼性を継承し、かつ高性能なプロセッサ「SPARC64 VI／VII」を採用
- マルチコア・マルチスレッド技術により、1 台あたり最大 16 コア／32 スレッドを実現
- バス帯域幅の強化や PCI Express の採用により高性能を実現
- 徹底的なデータ保護・冗長化により、装置単体の可用性を向上
- ハードウェアパーティショニング機能や DR 機能、COD 機能により、フレキシブルなサーバ運用が可能

「SPARC Enterprise M4000」は、従来のハイエンドサーバでサポートされていた高性能・高信頼・仮想化技術等の機能を、ミッドレンジクラスに凝縮したサーバです。メインフレームの信頼性を継承し、高性能プロセッサ「SPARC64 VI／VII」の採用により、1 台あたり最大 16 コア／32 のスレッドのマルチコア、マルチスレッド環境を実現することができます。

データベースやバッチ処理等を使用したミッションクリティカル業務は、1 つのトランザクション当たりの負荷が大きく、また処理順序を考慮しなければならない場合があるという特徴があります。このような、負荷の大きいトランザクションを高速に処理実行するために、高性能プロセッサ「SPARC64 VI／VII」を開発しました。強力な命令並列処理能力や高精度の命令分岐予測等の技術に加え、新たにマルチコア、マルチスレッド技術を採用することにより、高性能を実現しています。さらに、システムバスの強化や、PCI Express を採用するなど、システム全体で高性能化が図られています。

また、ミッションクリティカル業務は、停止してしまうとビジネスに大きな影響を与えてしまいます。「SPARC Enterprise」は、メインフレームの設計思想をもとに開発しています。サーバの故障や不具合がお客様業務を止めることのないよう、エラーの発生を未然に防止する機能、トラブルが起こった際にも訂正や縮退を行い稼動し続ける機能、コンポーネント冗長化・活性交換機能など、システムダウンを最小限に抑えるテクノロジーを多数取り入れています。

さらに「SPARC Enterprise M4000」では、従来はハイエンドサーバでしか提供されていなかった仮想化技術であるハードウェアパーティショニング機能やDynamic Reconfiguration 機能を使って筐体内のリソースを分割し動的に再構成することができます。例えば、日中・夜間・月初・月末など、時間によってサーバへ負荷を与える業務は異なります。これまでは、業務負荷のピークにあわせて個別にサーバを用意していましたが、この技術を用いることにより必要な資源を必要な時に追加・削除することが可能となり、変化のある業務負荷に柔軟に対応することができます。

3.1.2. SPARC Enterprise M3000

【SPARC Enterprise M3000 の特長】

- SPARC/Solaris エントリークラス最高のプロセッサコア性能を実現
- ミッドレンジクラスの高信頼技術をエントリークラスに継承
- 省電力/省スペースなど「Green Policy Innovation」の具現化

「SPARC Enterprise M3000」は、「SPARC64 Ⅶ」を搭載し、最大で 4 コア／8 スレッドのマルチコア・マルチスレッド環境を実現しました。メモリは最大 64GB 搭載可能で、SAS は 1 ポート、PCI Express は 4 スロット標準装備するなど、わずか 2U のスペースに業務に必要な機能をすべて実装できます。エントリーモデルにおける最高クラスの性能を発揮し、データベースサーバやアプリケーションサーバなどの幅広い業務に適用可能となっています。

また「SPARC Enterprise M3000」は、「M4000」から「M9000」の高信頼性を踏襲しています。LSI レベル、ユニットレベル、システムレベルとシステム全体で信頼性を積み上げ、高い信頼性を確保しています。

さらに、「SPARC Enterprise M3000」は、富士通が独自に定める環境配慮型製品である、スーパーグリーン製品です。コンパクトな 2U(ユニット)サイズで、

PRIMEPOWER450(4U)と比較し、50%の省スペースと軽量化を実現しています。消費電力は最大 505W (100V) と 54%も削減可能です。性能の向上と合わせて、CO2 排出量を年間約 65%削減できます。しかも、一般的なサーバ設置場所の環境温度 25℃の音圧レベルで 47dB という静音設計で、他社 4 コアサーバと比較しても、消費電力値や騒音値において、最も優れたエコロジーサーバとなっています。

また、Solaris 10 OS には、Solaris コンテナと呼ばれる仮想化技術を標準で装備しています。「SPARC Enterprise M3000」でも Solaris コンテナを使用したサーバ統合により資源を集約することが可能となっており、システムの効率化を実現できます。

3.2. ストレージシステム ETERNUS DX

企業経営と IT の一体化が進行した昨今、経営情報を中心とし、情報活用、リスクマネジメントという企業経営と社会責任に対応した IT インフラの整備が欠かせません。富士通では、お客様の様々なニーズに対応したストレージシステムを提供しています。ストレージシステム ETERNUS DX は、次の 3 点を基本要件と捉え、開発しております。

1. 大量の情報を活用するためのスケーラビリティと情報の遅滞の無い提供を実現するビジネスの継続性への備え
2. 情報を正しく保存するためのデータ保全性とセキュリティの確保
3. 企業レベルの大量の情報を適切かつ柔軟に運用管理し、TCO を削減

また、ストレージシステム ETERNUS DX は、お客様のビジネス目標を達成するために業務環境の変化に応じて、業務アプリケーション、業務プロセスから、定められた権限の下、必要なデータ、必要なストレージリソースに、いつでもアクセスし続けるストレージインフラとサービスを提供しております。

3.2.1. RAIDマイグレーション

RAID マイグレーションとは、データ保証を行いながら論理ボリュームを別の RAID グループへ移行させる機能です。これによりお客様の業務に応じた RAID・論理ボリュームの再配置が可能になります。

論理ボリュームの再配置は業務運用中に実行することができ、サーバの CPU 負荷に影響を及ぼすことはありません。また、RAID レベルも RAID5 から RAID1+0 などの異なる RAID レベルへ再構築できます。

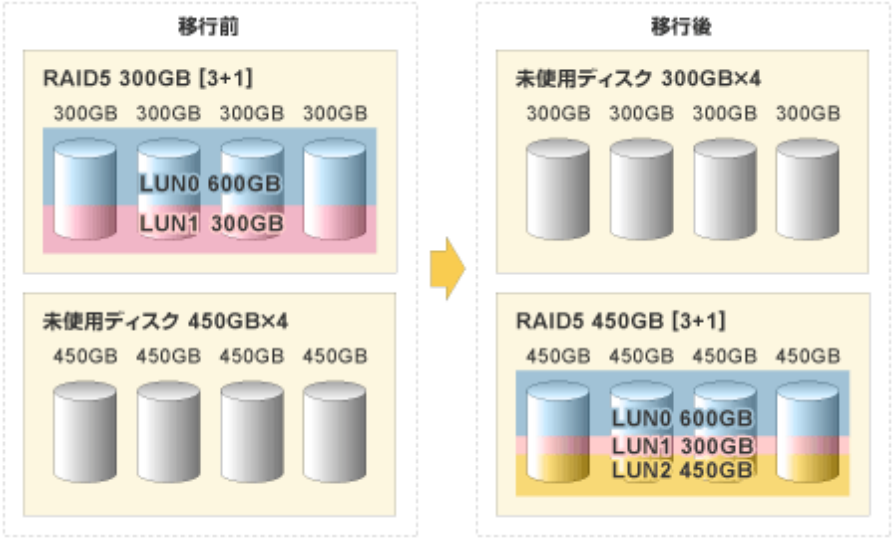


図 3-1　300GB ディスクの RAID5(3+1)構成を異なる容量の 450GB ディスクの RAID5(3+1)へ移行し、余剰スペースに論理ボリューム(LUN2)を追加した例

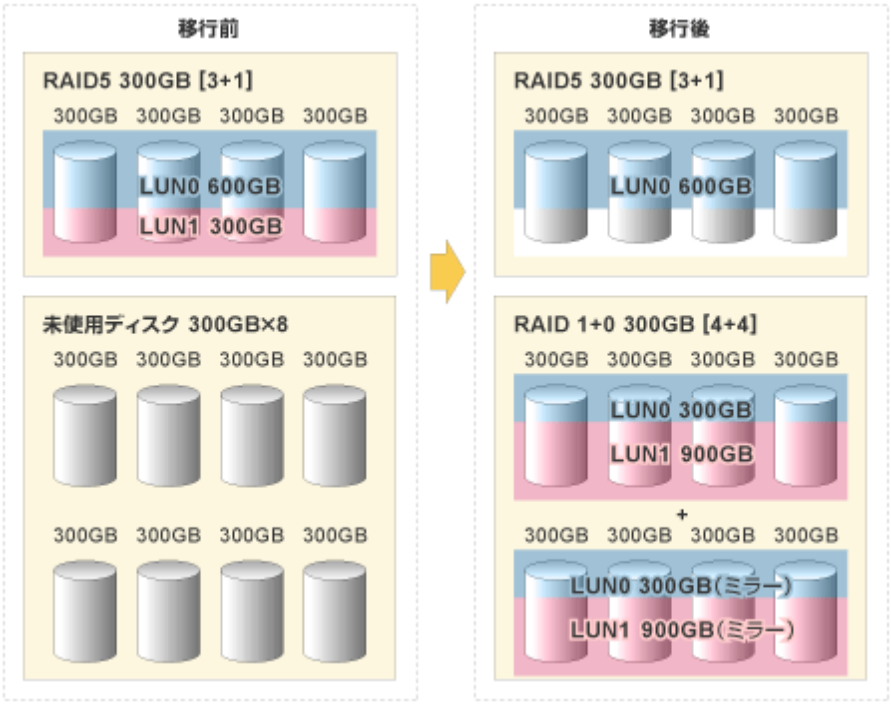


図 3-2　RAID5(3+1)構成内の一部ボリュームを異なる RAID レベル RAID1+0 へ移行した例

3.2.2. エコモード

ETERNUS DXディスクアレイは、お客様のシステム要件にあわせて、必要なときだけディスクドライブの回転を制御するエコモード（MAID技術[1]）を備えています。

エコモードとは、アクセスされる時間が限られているディスクに対し、一定期間ディスク回転を停止させ、消費電力を削減するモードです。回転停止期間のスケジューリングは、ディスクと時刻の設定により実施され、 RAID グループごとに適用でき、バックアップ等の運用にあわせた設定も可能です。

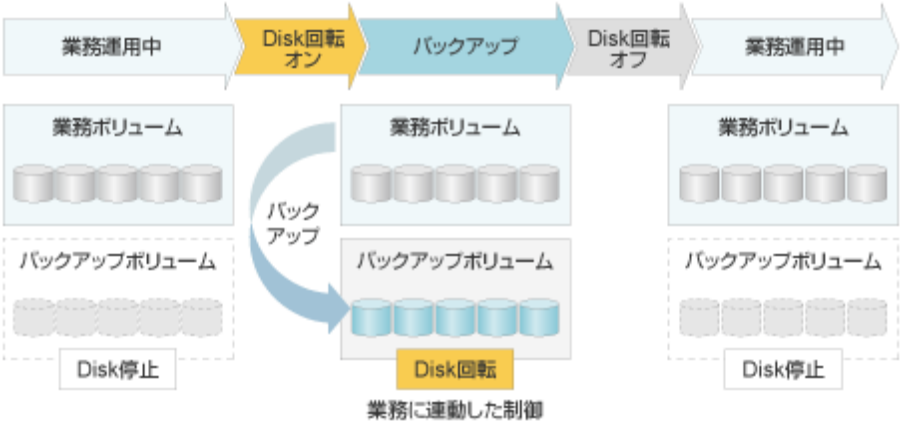


図 3-3　MAID 技術

このエコモードを適用したディスクドライブに一定時間(可変）アクセスがない場合はディスクの回転を停止させます。停止していたディスクドライブへアクセスが生じた場合は 1 分程度で使用可能となります。

下記のETERNUS DX400 seriesによる構成での活用の場合、エコモードを適用したバックアップボリューム（50 本のディスクドライブ）を 1 日 20 時間停止させることで、停止させない運用と比較すると、年間の消費電力量を約 4,720kWh、CO2 排出量を約 1,830kg削減できます。これにより、全体の電力消費量を約 15%削減し、環境負荷も低減できます。[2]

[1] MAID 技術：Massive Array of Idle Disks の略。使用頻度の低いディスクドライブのディスク回転を停止させることで、消費電力を削減するとともに、ディスクドライブの寿命を延長させる技術。

[2] 活用例において、バックアップボリュームのディスクを常時稼働させた場合と 1 日 20 時間停止させた場合の当社比。他ディスクアレイでは値が異なります。

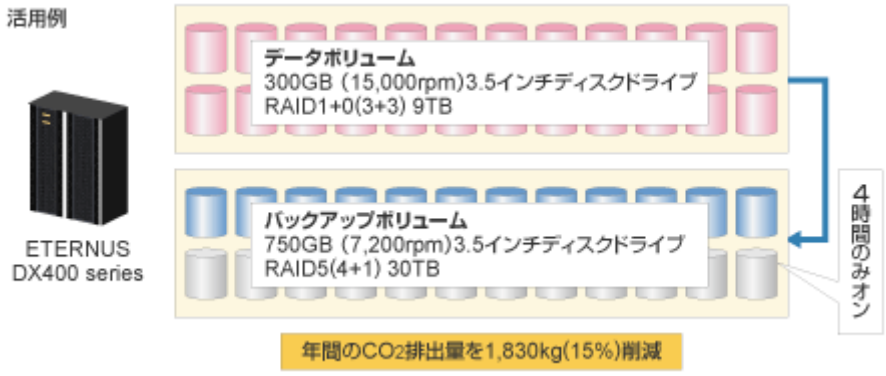


図 3-4　エコモード活用例

エコモードの詳細については、ホワイトペーパー[3]をご参照ください。

3.2.3. ETERNUS SF Storage CruiserとETERNUS SF AdvancedCopy Manager

ETERNUS SF Storage Cruiser は、ETERNUS DX ディスクアレイのディスクドライブから業務サーバのファイルシステム、接続パス、ミラーディスク、データベース等のリソースを関連付け、運用管理を行います。各リソース間の関係を簡単に把握し、ストレージシステムの増設や障害リカバリー、性能情報の取得/表示を確実に行うことが可能です。

ETERNUS SF AdvancedCopy Manager は、ETERNUS DX ディスクアレイと連携し、アドバンスト・コピー機能による高速バックアップ/リストア、レプリケーション運用を実現します。アドバンスト・コピー機能は、ある時点における業務ボリュームを短時間で同じディスクアレイ内の別ボリューム（複製ボリューム）にコピーします。コピー完了後の複製ボリュームを使ってテープ装置へバックアップすることができます。ETERNUS SF AdvancedCopy Manager のテープサーバオプションを使用すると、コピー完了とテープバックアップ開始のスケジューリング、ディスクとテープとを別々に管理するなどの煩雑さを解消した上でディスクからテープまで一貫してバックアップできます。また、複製ボリュームは業務ボリュームと切り離されるため、業務を継続しても書き換えられることはありません。

3.2.4. ETERNUSマルチパスドライバ

ETERNUS マルチパスドライバは、一つのサーバから複数パスを使用し、シングルシステムにおける連続運転およびシステム性能を向上します。

---

[3] 「富士通 ETERNUS ディスクアレイの MAID 技術による省エネルギー」
http://storage-system.fujitsu.com/jp/products/diskarray/download/pdf/MAID_whitepaper.pdf

このドライバは、サーバとディスクアレイ間の物理的なアクセスパスを多重化した構成で、パス故障時にも、ディスクアレイへのアクセスを継続させるソフトウェアです。論理ボリューム毎に運用パスと待機パスの設定が可能です。

また、ロードバランス機能はサーバとディスクアレイ間の物理的なアクセスパスを多重化し、多重パス同時使用によるパスのロードバランス(I/O の負荷分散)を行い、システム性能を向上させます。

3.2.5. ETERNUS DXディスクアレイにおけるニアライン用ディスクの特長

ETERNUS DX ディスクアレイは、筐体内にオンライン用ディスクドライブと大容量かつ、高信頼で低価格なニアライン用ディスクドライブの混在搭載が可能です。お客様のデータ活用ポリシーに従い、使用頻度の低いものをニアライン用ディスクドライブで管理したいという要求にお応えします。

ニアライン用ディスクドライブは、次のような用途に有効活用していただけます。

ディスク to ディスクのバックアップ先として

ディスク to ディスクのバックアップ運用で使用されるバックアップボリュームは、テープへバックアップするまでの一次保存用や、万一の高速リカバリー用として、複数世代管理するのが一般的です。このような大容量バックアップ用途に最適なのが 大容量かつ、高信頼で低価格なニアライン用ディスクドライブです。

ETERNUS DX ディスクアレイのアドバンスト・コピー機能を利用して高速かつコストパフォーマンスに優れるバックアップ環境を提供します。

参照用データの保存先として

一般的に参照を目的とした保存用データは、頻繁にはアクセスされないが、必要な時にスムーズに参照したいデータであると言えます。しかもこのようなデータ種は急速に増加しているのに加えて、長期間の保存も必要となりますので、コストパフォーマンスに優れるニアライン用ディスクドライブへの保存が適しています。

表 1　オンライン用ディスクとニアライン用ディスクの違い

	オンライン用ディスクドライブ （ファイバチャネルディスクドライブ）	ニアライン用ディスクドライブ （ニアライン SATA ディスクドライブ）
容量	146GB, 300GB, 450GB	500GB, 750GB, 1TB
回転速度	15,000rpm	7,200rpm
インターフェース速度	FC (4Gbit/s)	FC (4Gbit/s)
サポート RAID レベル	RAID1, RAID1+0, RAID5, RAID6	RAID5, RAID6
推奨する使用方法	・使用頻度の高いデータの保管に使用します。	・使用頻度の高くない、バックアップ用とアーカイブ用に使用します。

4. Oracle Database 11g 機能紹介

本検証にて使用した Oracle Database 11g の機能について説明します。

4.1. Oracle Partitioning

Oracle Partitioning(Oracle 8 からの機能)を利用すると、表、索引、または索引構成表を細分化し、これらのデータベース・オブジェクトにアクセスすることができます。アプリケーションの視点からは、パーティション化された表はパーティション化されていない表と変わらず、SQL を使用してパーティション化した表にアクセスする場合でも一切変更する必要がありません。

各パーティションには固有の名前があり、オプションで表圧縮を有効にしたり、パーティションを異なる表領域に保存したり、あるいは異なる ASM ディスク・グループに保存したりするなどの固有の記憶特性を持つこともできます。

Oracle Partitioning は、個別の各種パーティションに配置されるデータの移動を制御する、様々なデータ分散方法を提供しています。ILM では、値の範囲で各種パーティションにデータを分割するレンジ・パーティションが有効です。

レンジ・パーティションではキーの値の範囲に基づいて、データが分散されます。例えば、表のパーティション・キーが日付列の場合、'January 2009'のパーティションに、2009 年 1 月のデータを含むように指定すれば、そのパーティションにはキー列の値が、'01-JAN-2009'から'31-JAN-2009'の値を持つデータが含まれます。データの分散は、途切れることなく連続して行われ、レンジの下限は、先行するレンジの上限によって自動的に定義されます。

本検証ではOracle Database 11g、レンジ・パーティションを用いました。この他にもOracle Database 11gでは多くのパーティションの機能拡張があります。詳しくはホワイトペーパー「Oracle Database 11gにおけるパーティション化[4]」をご覧ください。

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

4.2. Automatic Storage Management(ASM)

ASM は、Oracle Database10g で実装された、データベースで使用されるディスクを管理する機能です。複数のディスクへのストライピングやミラーリングといった機能に加え、ダイナミックにディスクの構成を変更できるリバランス機能を備えています。これにより、業務データベースを停止することなくディスクを追加することができ、ディスク管理コスト削減や、ディスクの負荷分散による性能改善が期待できます。

4.3. Real Application Clusters(RAC)

RAC は、複数ノードでディスクとメモリ上の情報を共有し、全ノードで並列処理を行うシェアードエブリシング型のクラスタデータベースです。RAC を採用することで、以下のメリットを享受することが可能です。

1. アクティブ-アクティブ構成による、サーバリソースの有効活用
2. ノードを追加することによる、柔軟なシステム拡張
3. シェアードエブリシング構成による、障害発生時のシステム停止時間、切り替え時間の低減

5. 検証環境

本検証はデータベースサーバ 2 台、ストレージ 1 台の RAC 構成で検証を実施しました。

5.1. システム構成

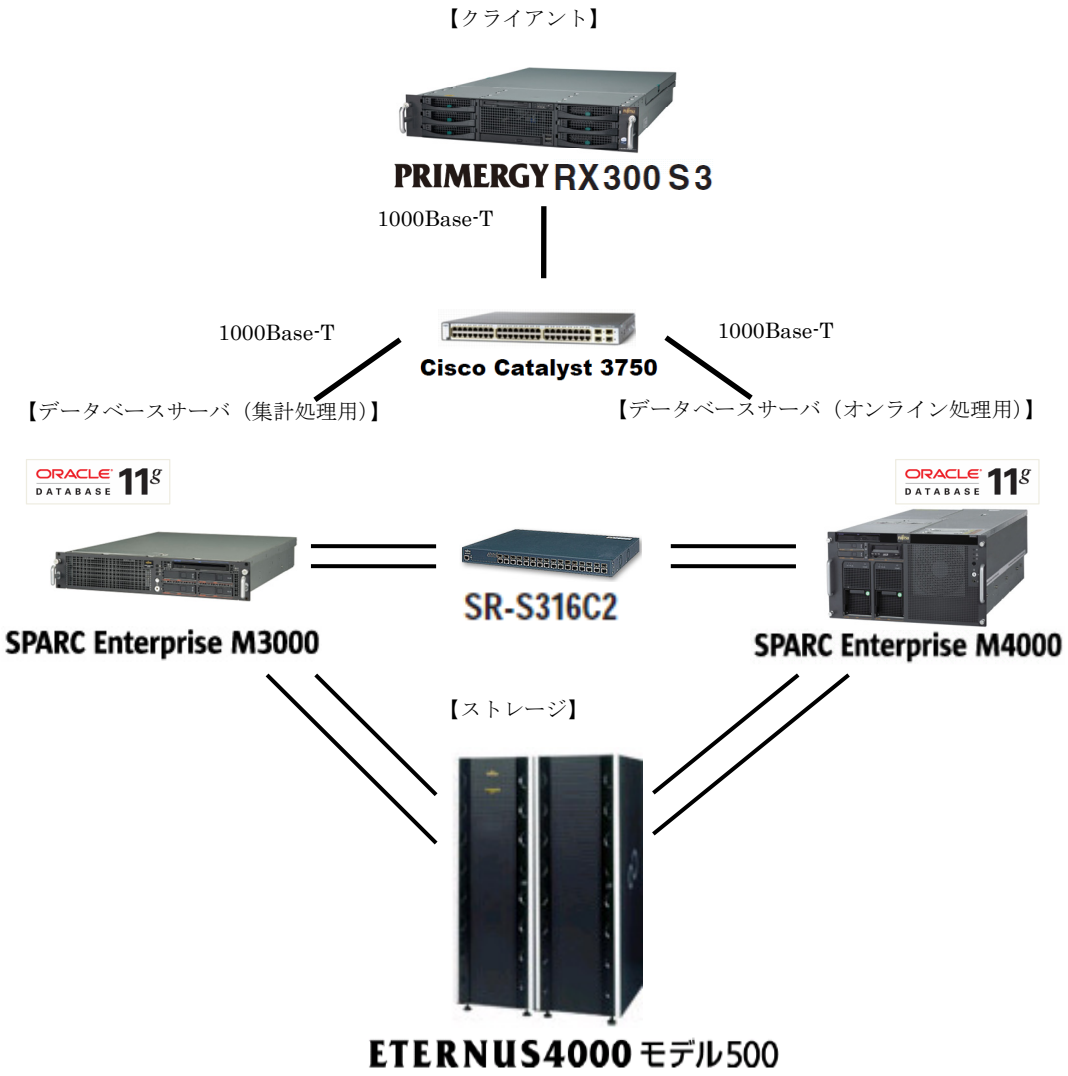


図 5-1 システム構成概略

* ETERNUS4000 モデル 500 の後継は ETERNUS DX440 です。

5.1.1. データベースサーバ（オンライン業務サーバ）

ハードウェア

モデル	富士通 SPARC Enterprise M4000
CPU	SPARC64 Ⅶ 2.40GHz/5MB キャッシュ 2CPU/8 コア/16 スレッド
メモリ	32GB
内蔵 HDD	73GB SAS Disk×2

ソフトウェア

OS	Solaris™ 10 Operating System (Generic_137137-09)
データベース	Oracle Database 11g (11.1.0.7) Enterprise Edition
ストレージ管理	ETERNUS SF AdvancedCopy Manager 13.4

5.1.2. データベースサーバ（集計処理サーバ）

ハードウェア

モデル	富士通 SPARC Enterprise M3000
CPU	SPARC64 Ⅶ 2.52GHz/5MB キャッシュ 1CPU/4 コア/8 スレッド
メモリ	12GB
内蔵 HDD	146GB　SAS Disk×2

ソフトウェア

OS	Solaris™ 10 Operating System (Generic_137137-09)
データベース	Oracle Database 11g （11.1.0.7） Enterprise Edition
ストレージ管理	ETERNUS SF AdvancedCopy Manager 13.4

5.1.3. ストレージ

モデル	富士通 ETERNUS4000 モデル 500
ディスクドライブ	ファイバチャネルディスク 146GB (15,000rpm) × 38 本 73GB (15,000rpm) × 12 本 ニアライン SATA ディスク 750GB (7,200rpm) × 10 本

5.1.4. クライアント

ハードウェア

モデル	富士通 PRIMERGY RX300S3
CPU	デュアルコア インテル Xeon 5160 プロセッサ 3.00 GHz
メモリ	3GB
内蔵 HDD	300GB (15,000rpm) SAS Disk × 3 (RAID 5)

ソフトウェア

OS	Windows Server 2003 R2
ストレージ管理	ETERNUS SF AdvancedCopy Manager 13.4
	ETERNUS SF Storage Cruiser 13.4

なお、ハードウェア、ソフトウェアの富士通での取り扱いにつきましては、富士通の貴社ご担当営業までお問い合わせ下さい。

5.2. スキーマ構成

本検証で使用したスキーマ構成は以下になります。

テーブル名	件数	概要
BUYERS	500000 件	取引先表。取引先 ID、担当者名等を格納。
DEPARTMENTS	11220 件	部門表。部門 ID,部門名等を格納。
ORDERSFACT	1 ヵ月 6200 万件×72 ヶ月	注文履歴表。1 年あたり約 80GB 2004 年から 2009 年まで計 6 年分を作成。
PRODUCTS	10000000 件	商品表。商品 ID、商品名等を格納。

テーブル名	索引名	概要
BUYERS	Idx_buyers	
DEPARTMENTS	Idx_departments	
PRODUCTS	Idx_products_prodid	
PRODUCTS	Idx_products_prodname	
ORDERSFACT	Idx_ordersfact_orderid	1 年あたり約 50GB
ORDERSFACT	Idx_ordersfact_spid	

5.3. アプリケーションモデル

本検証では、社内購買管理システムをモデルとしたアプリケーションを使用しました。オンライン処理は商品発注処理を、集計処理は売上の月次集計を想定しています。なおオンライン処理と集計処理はそれぞれ RAC を構成する別のノードで実行します。

5.3.1. オンライン処理

本検証では、Java アプリケーションを利用して以下の二つのトランザクションを多重実行しています。

1. 注文トランザクション
    1. 発注処理を行う社員 ID から部署 ID、地域 ID など必要な情報を Departments 表から取得します。
    2. 発注する商品名から商品 ID を Products 表から取得します。
    3. 注文数などを入力して insert を行います。

    このトランザクションで検索される表は、データベース・バッファキャッシュ上に読み込まれ、メモリ内で処理が終わります。

2. 注文変更トランザクション
    1. 2010 年 1~3 月の注文済みデータから発注日(timeid)と発注番号(orderid)を条件にして Idx_ordersfact_orderid を使用して注文を変更するデータを検索します。
    2. 検索したデータの発注数を変更します。

これら 2 つのトランザクションを 150 多重で実行します。1 処理は、10 トランザクションを実行し、注文トランザクションは 9 回、注文変更トランザクションは 1 回の割合で実行されます。

5.3.2. 集計処理

本検証では、集計処理を想定した以下の二つのクエリを実行しました。

1. 月別売上集計クエリ

    ある年のデータに対して、ある月の売上額の対前月比を算出します。

2. 社員別の売上集計クエリ

    ある年の 1 月の社員毎の売上額を算出し、売上額上位 100 名の名前、売

上額を返します。

6.1 Oracle標準機能によるILM及び 6.2 RAIDマイグレーションを利用したILMでは 1. の月別売上集計クエリを 2008 年のデータに対してシリアルで繰り返し実行しました。1. のクエリは 2008 年 1 月と 2 月の売上比というように二ヶ月ごとに順次実行し、12 月まで完了すると、繰り返し検索対象年(2008 年)への集計クエリを実行します。

また6.3ディスク性能差による集計処理への影響では、2009 年のデータに対する月別売上集計クエリ及び社員別の売上集計クエリのみ使用しています。

6. 検証内容／結果

MOVE PARTITION 及び RAID マイグレーションを使用した ILM 検証の内容、検証結果について報告します。

6.1. Oracle標準機能によるILM

Oracle データベースの機能を利用した ILM は、MOVE PARTITION 文を利用して実現します。MOVE PARTITION を用いることで、お客様のシステム要件や環境に合わせた ILM を実現することが可能です。

6.1.1. Oracle標準機能を利用したILMと物理設計

MOVE PARTITION を利用した ILM を実現するための効率的なデータベース物理設計について説明します。

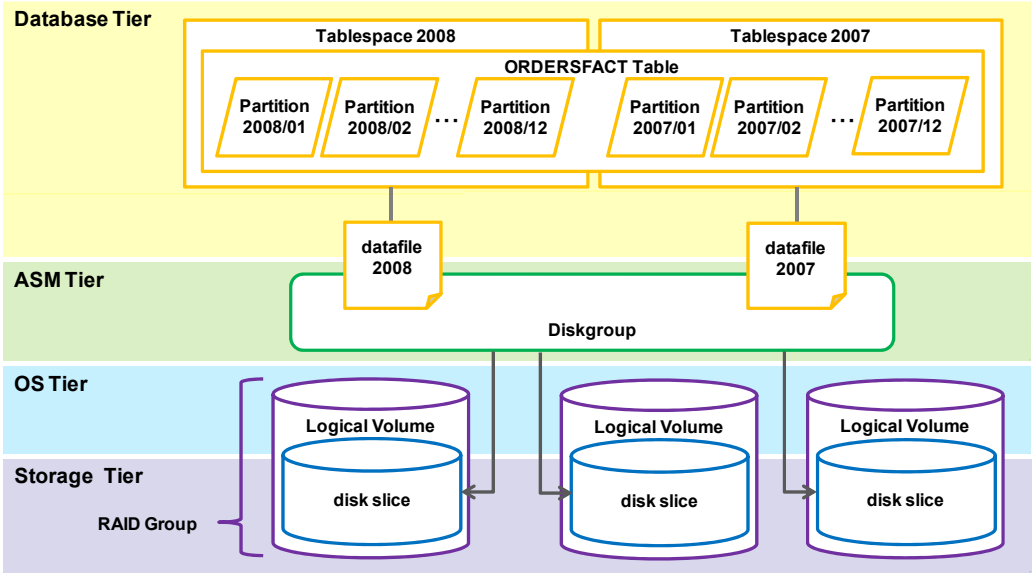


図 6-1 Oracle 標準機能を利用した ILM の物理設計

図 6-1は本検証で使用したデータ・モデルの 2007 年と 2008 年を例にしています。上記の例では、12 ヶ月分のデータを 1 つの表領域に格納し、データファイルも 1 つとしています。一方で 1 つDiskGroupには複数のデータファイルを格納します。DiskGroupは複数の論理ユニットで構成します。

MOVE PARTITION を使用するとパーティション単位で表領域を移動できるため、移動前のデータが存在する表領域がどのような構成で、どのように論理ボリュームを使用しているか等は、ILM の運用という観点では問題となりません。

ILM 導入後の運用において、時間の経過と共に古くなったパーティションを、現在のアクセス頻度に適したディスク上へ移動させるという操作が定期的に発生すると考えられます。アクセス頻度の減ったデータを定期的に低コストなディスクへ移

動することにより、高性能ディスク上のデータが増えすぎることを回避し、全体のデータが増え続けたとしても安価なディスクの追加で対処することにより、ディスク追加に伴うコストを削減することができます。

本検証で使用した ORDERSFACT 表で Oracle 標準機能による ILM を考えます。ILM による運用対象表および移動元表領域、移動先表領域は以下になります。

ILM による運用対象表：ORDERSFACT 表

移動するパーティション：P200901~P200912(約 80GB)

移動元表領域：TS_2009(FC ディスク上)

移動先表領域：TS_2009_OLD(SATA ディスク上)

データの保存期間を 5 年、保存期間を経過したデータは削除することを想定しています。そのため、2009 年のパーティション移動時、以下の表領域を削除しています。

削除表領域：TS_2004_OLD

また本検証では、MOVE PARTITION 文の実行時に UPDATE INDEXES 句を使用して索引の REDUILD を行い、以下の索引パーティションの移動も同時に行いました。(2009 年 1 月~12 月分のみ、約 50GB)

- idx_ordersfact_orderid
- idx_ordersfact_spid

なお、この運用方針は一例であり、お客様の環境やシステム要件等を考慮し、決定することを推奨します。以降、運用手順について説明します。

① まず 2009 年から 2010 年へ年が変わる前に新規のパーティションを追加し、索引のデフォルト表領域を 2010 年の表領域に変更します。（図 6-2）

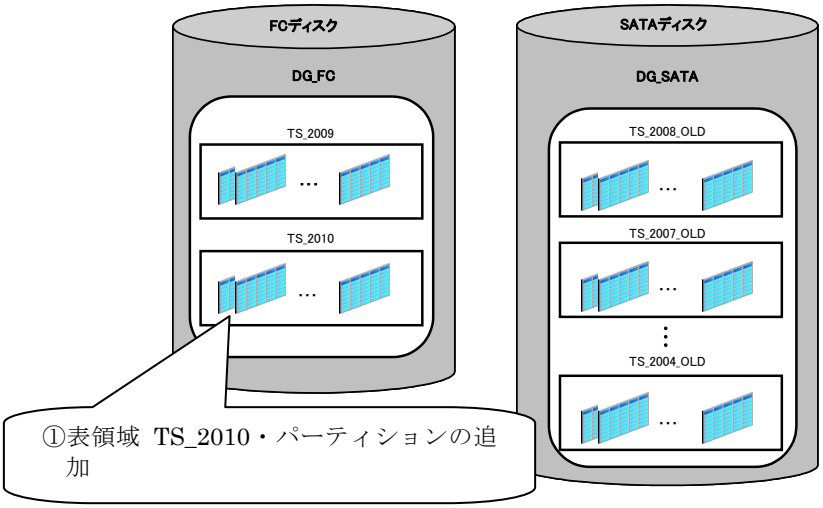


図 6-2 表領域・パーティションの追加

追加したパーティションは、統計情報が取得されていないため、実行計画が変わってしまい、性能劣化につながる恐れがあります。また追加したばかりのパーティションでは、データ件数が他のパーティションに比べて著しく少ないため、統計情報が大きく異なる可能性があります。この場合も実行計画が変わってしまい、性能劣化をもたらす可能性があります。これを防止するために今回は 2009 年のパーティションのオプティマイザ統計情報を 2010 年のパーティションにコピーする運用を行い、実行計画の変化による性能劣化を防いでいます。

② 次に、保管する必要のなくなったパーティションを削除するため、パーティション及び表領域(ここでは 2004 年のパーティション及び表領域)を削除します。(図 6-3)

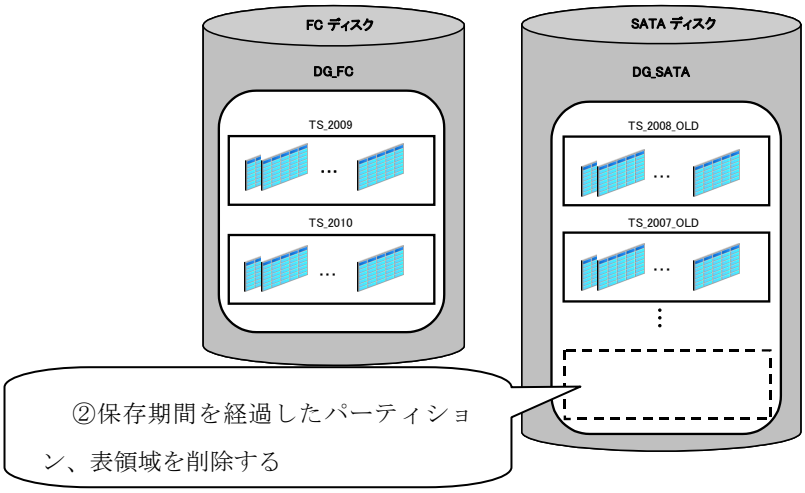


図 6-3 表領域・パーティションの削除

③ SATAディスクに移動先表領域を作成します。(図 6-4)

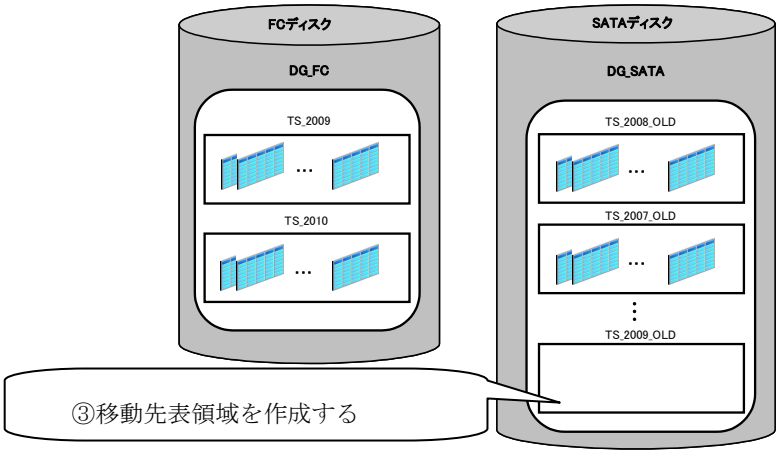


図 6-4 移動先表領域の作成

④ 移動元表領域をREAD ONLYにした後、ALTER TABLE ... MOVE PARTITION 文を実行し、パーティションを移動元表領域から移動先表領域に移動します。

また移動元表領域は削除します。(図 6-5)

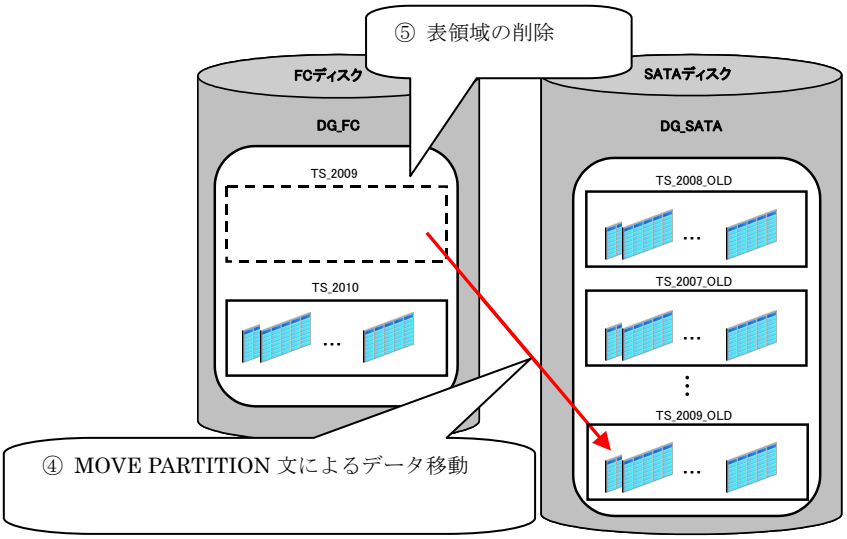


図 6-5 MOVE PARTITION によるデータの移動

6.1.2. MOVE PARTITION実行時の各リソースの使用状況

まず業務を停止した状態で、1 年分のデータ(索引込み約 130GB)を MOVE PARTITION 文によって、FC ディスクから SATA ディスクへ移動させた場合の各リソースの負荷状況を確認します。本検証では MOVE PARTITION 文をシングルプロセスで実行しているため、CPU 使用率は 1 スレッド分(約 12.5%)の使用率になっています。

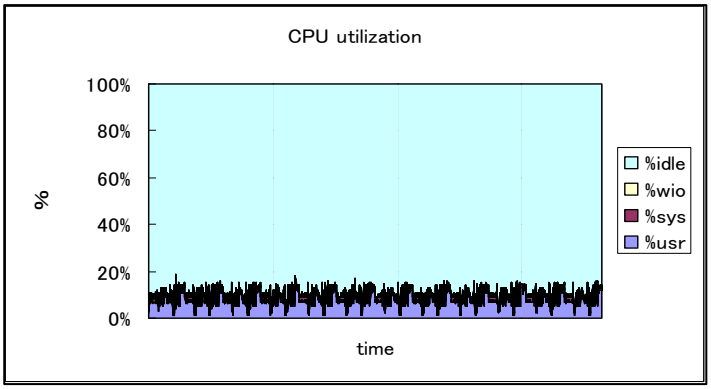


図 6-6 MOVE PARTITION 文の CPU 使用率

他に業務が行われている状態で MOVE PARTITION 文を実行すると、MOVE PARTITION 実行中は、CPU 使用率が 10%前後上昇します。そのため MOVE PARTITION 実行時には、CPU リソースが不足しないように留意する必要があります。

次にFCディスクとSATAディスクの負荷を確認します。FCディスク及びSATAデ

ィスクの負荷は図 6-7になります。

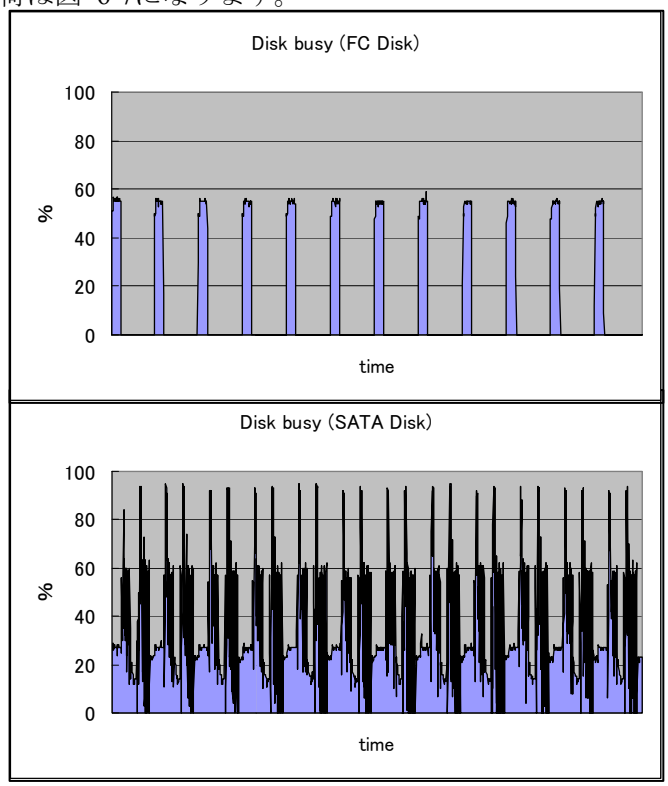


図 6-7 MOVE PARTITION 実行時のディスクビジー率

MOVE PARTITION 文を実行すると、パーティション単位で FC ディスクに格納されているデータを読み込み、SATA ディスクに書き込みを行います。そしてデータの移動が終了すると、索引の REBUILD を実行します。データは既に SATA ディスクに移動しているため、SATA ディスクからデータを読み込み、ソート処理を行った後、SATA ディスクに書き込みを行います。

本検証では月単位でパーティションを作成し、1 年分のパーティションを移動しているため、各パーティションに対応するグラフの山が 12 個あります。

また図 6-8では、パーティション単位での処理を示しています。

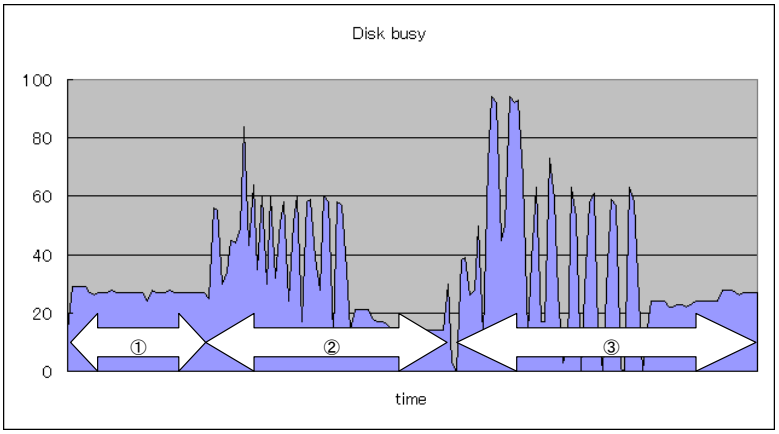


図 6-8 パーティション単位の処理

①でデータ移動、②・③で索引の REBUILD を行っています。データ移動時と比べて索引の REBUILD の方が SATA ディスクの負荷が 3 倍ほど高くなっていることがわかります。以上の結果から、SATA ディスクの負荷状況を把握し、ディスク負荷の少ない時に MOVE PARTITION 文を実行することが推奨されます。

また、グラフより索引の REBUILD に多くの時間がかかっています。つまり索引のサイズや数により、MOVE PARTITION 文にかかる時間は変化します。

6.1.3. MOVE PARTITIONが業務へ与える影響

オンライン処理、集計処理実行中に MOVE PARTITION 文を実行した場合について見ていきます。

最初にMOVE PARTITIONがオンライン処理に与える影響を確認します。オンライン処理の内容については、『5.3.1オンライン処理』を参照ください。

オンライン業務は片方のノードで実行しています。その時に各ノードでMOVE PARTITION文を実行した場合にどのような影響があるか検証します。なお検証手順については、『6.1.1Oracle標準機能を利用したILMと物理設計』を参照ください。

6.1,3,1　オンライン業務を実行していないノードでMOVE PARTITIONを実行した場合

まず、定常時(オンライン処理のみを行っている状態)のオンライン処理のスループットとレスポンスタイムを確認します。図 6-9は、定常時の平均スループット、レスポンスタイムを 1 として、相対値で表しています。

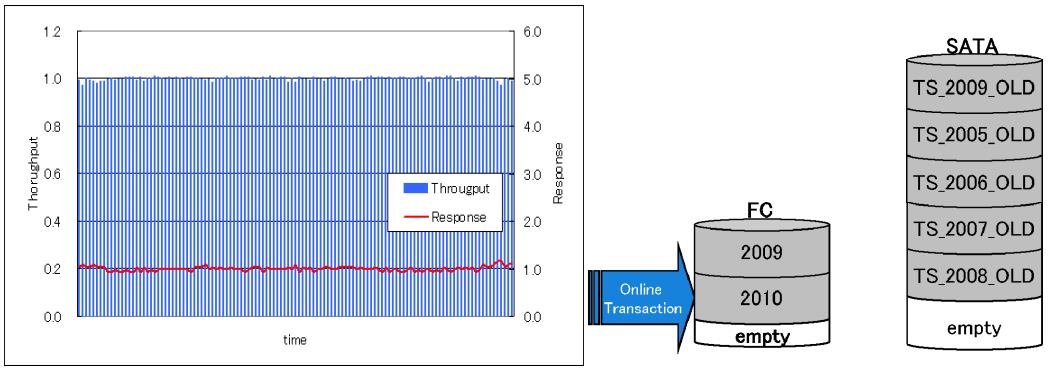


図 6-9 定常時のスループット、レスポンスタイム

図 6-10では、オンライン処理中にオンライン処理を実行していないノードで、2009 年のデータをMOVE PARTITION文でSATAディスクへの移動を実行した場合のスループット、レスポンスタイムを示しています。定常時の平均スループット、レスポンスタイムを 1 として、相対値で表しています。

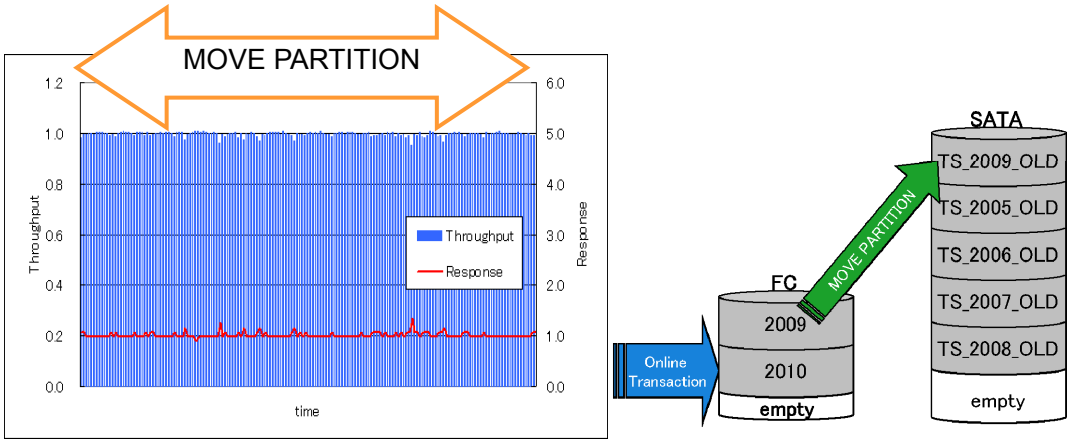


図 6-10 オンライン処理中に、他のノードで MOVE PARTITION を実行した場合のトランザクションの性能

オンライン処理実行中にオンライン処理を実行していないノードでMOVE PARTITION文を実行した場合、オンライン処理にはほとんど影響はありません。図6-11が示すように、MOVE PARTITION文の実行時間もほぼ変化はありません。

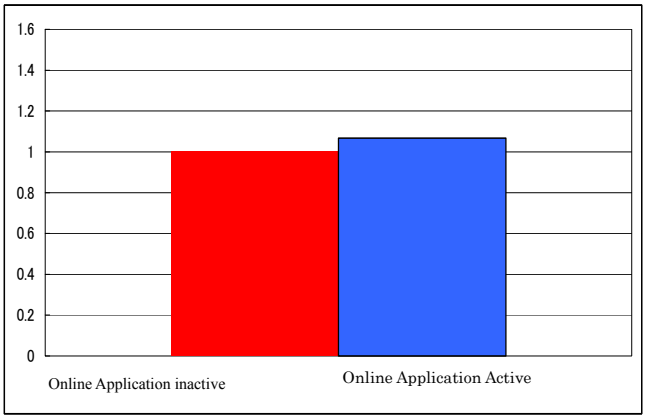


図 6-11 MOVE PARTITION 実行の時間比

以上のように、RAC 環境で業務分割を行っている場合や、負荷の少ないノードで MOVE PARTITION 文を実行することで、オンライン処理への影響を最小限にすることが可能です。

6.1.3.2 オンライン処理を実行しているノードでMOVE PARTITIONを実行した場合

この検証は、RAC 環境ではなく、シングル DB 環境で MOVE PARTITION 文を実行する場合を想定しています。

図 6-12では、オンライン処理を実行しているノードで、2009 年のデータをMOVE PARTITION文でSATAディスクへ移動した場合のスループット、レスポンスタイムを示しています。定常時の平均スループット、レスポンスタイムを 1 として、相対値で表しています。

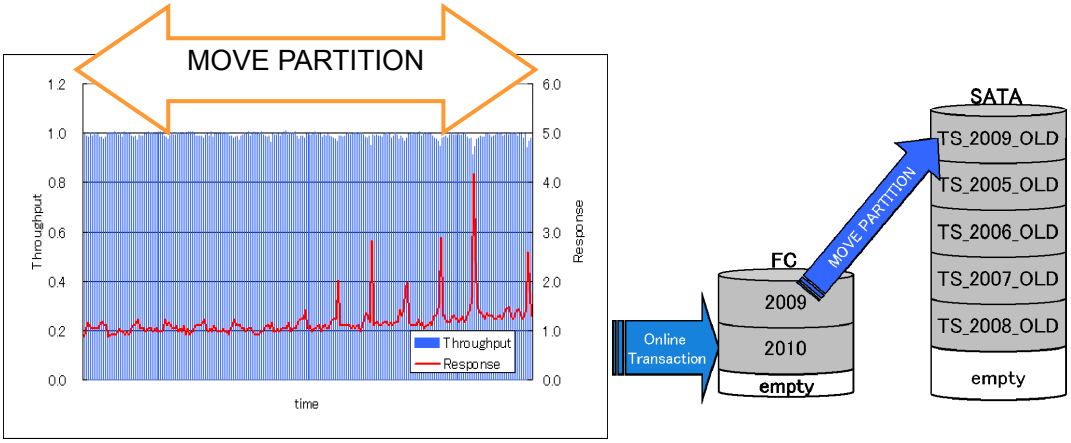


図 6-12 オンライン処理中に、同じノードで MOVE PARTITION を実行した場合のトランザクションの性能

グラフより一部レスポンスタイムが悪化しているのが確認できます。MOVE PARTITION文実行中、特にデータ移動時にFCディスクの負荷が高くなった為、注文変更トランザクションの過去の商品検索クエリに遅延が発生したと考えられます。図 6-13はオンライン処理中にMOVE PARTITION文を実行した場合のFCディスクのビジー率のグラフです。

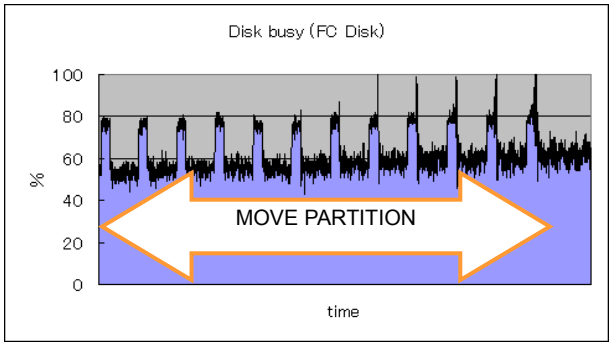


図 6-13 オンライン処理を実行しているノードのディスクビジー率

グラフより、ディスクビジー率が 100%になっている時間が確認できます。この時に注文変更トランザクションのうち、既に注文したデータを検索する SQL 文に影響が出たと考えられます。

次にCPU使用率について見ていきます。図 6-14がオンライン処理のみを行っている場合のCPU使用率とMOVE PARTITION文を同時に実行した場合のCPU使用率のグラフです。オンライン処理のみを実行している時よりも、MOVE PARTITIONとオンライン処理を同時に実行する時の方が、CPU使用率が 10%前後高くなっています。

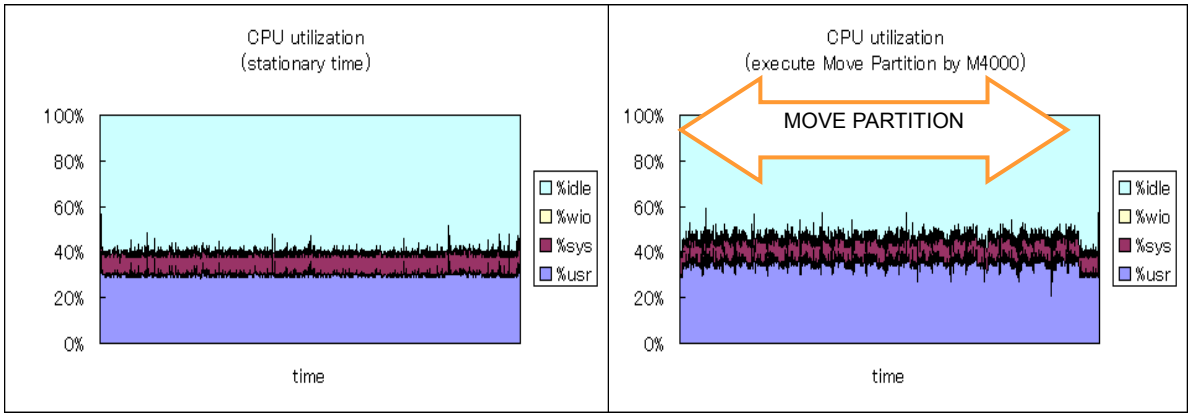


図 6-14 オンライン処理を実行しているノードの CPU 使用率の比較

以上からシングル DB 環境での運用を考える場合、CPU 使用率よりも FC ディスクおよび SATA ディスクのディスク負荷に留意する必要があります。特に FC ディスクへの負荷が高くなると、オンライン処理に影響を及ぼす可能性があります。

6.1.3.3 パーティションを異なるFCディスクに分割した場合

上記検証より、シングルDB環境でのILMを想定するとFCディスクへの負荷が高くなり業務処理に影響が及ぶことが想定されます。しかし 2010 年と 2009 年のパーティションをFCディスクの異なるRAIDグループに配置し、FCディスクの負荷を分散させることで、注文変更トランザクションへの影響を最小限に抑えることができると考えられます。図 6-15は 2009 年と 2010 年のパーティションを異なるRAIDグループに配置した場合の各トランザクションのスループット、レスポンスタイムを示しています。定常時の平均スループット、レスポンスタイムを 1 として、相対値で表しています。

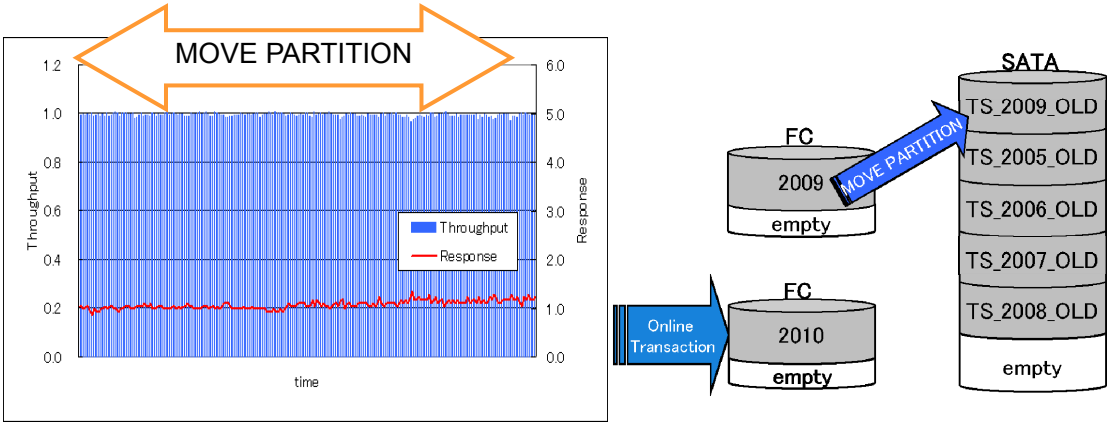


図 6-15 FC ディスクを分割した場合のトランザクション性能の相対値

図 6-15からわかるように、定常時と比較して各トランザクションの性能に大きな変化は見られません。

またそれぞれのFCディスクの負荷は図 6-16になります。

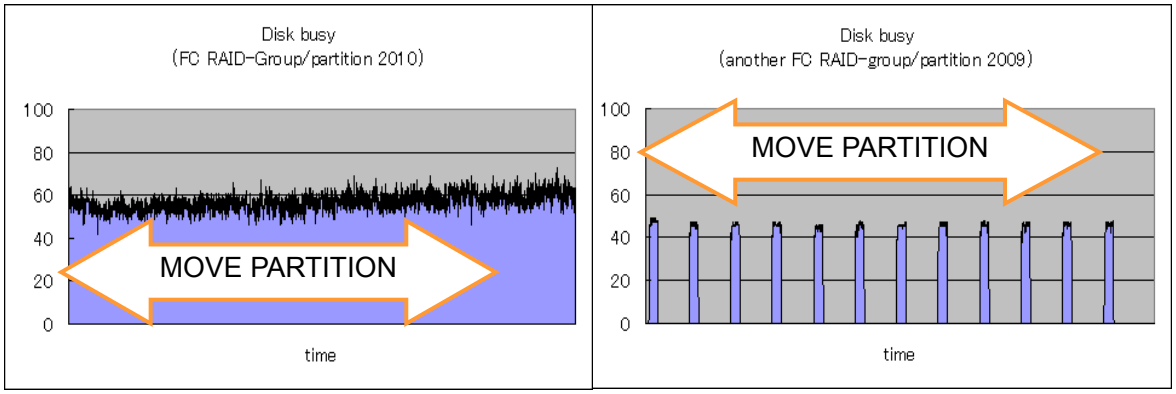


図 6-16 各 FC ディスクのディスクビジー率

FC ディスクの負荷が高い場合、オンライン処理がアクセスするパーティションと MOVE PARTITION 文により移動するパーティションをそれぞれ異なる RAID グループに配置することにより、ディスク I/O を分散させ、トランザクションへの影響を抑えることが可能になります。

6.1.3.4 集計処理への影響

集計処理を実行しているノードでMOVE PARTITION文を実行した場合の影響を見ていきます。集計処理処理の内容については、『5.3.2集計処理』を参照ください。

集計業務はオンライン処理を実行するノードとは異なるノードで実行します。なお検証手順については、『6.1.1Oracle標準機能を利用したILMと物理設計』を参照ください。

集計処理のクエリのレスポンスタイムは図 6-17になります。2008 年のデータに対するクエリレスポンスタイムの平均値を 1 として、相対値で表しています。

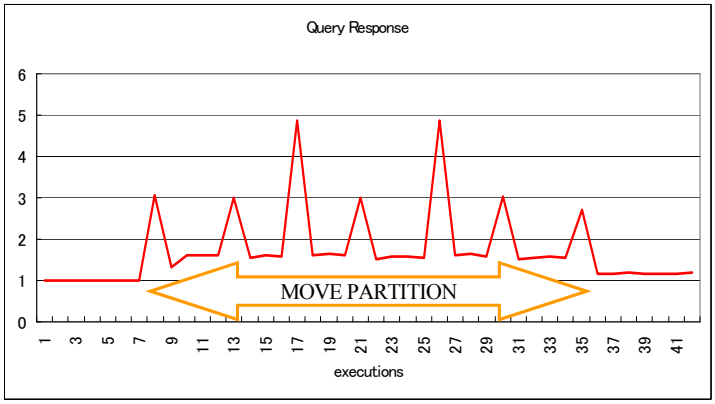


図 6-17 集計クエリのレスポンス比

MOVE PARTITION 文実行中は、レスポンスタイムが最大で約 5 倍まで劣化して

います。

次に CPU 使用率、ディスクビジー率について確認します。

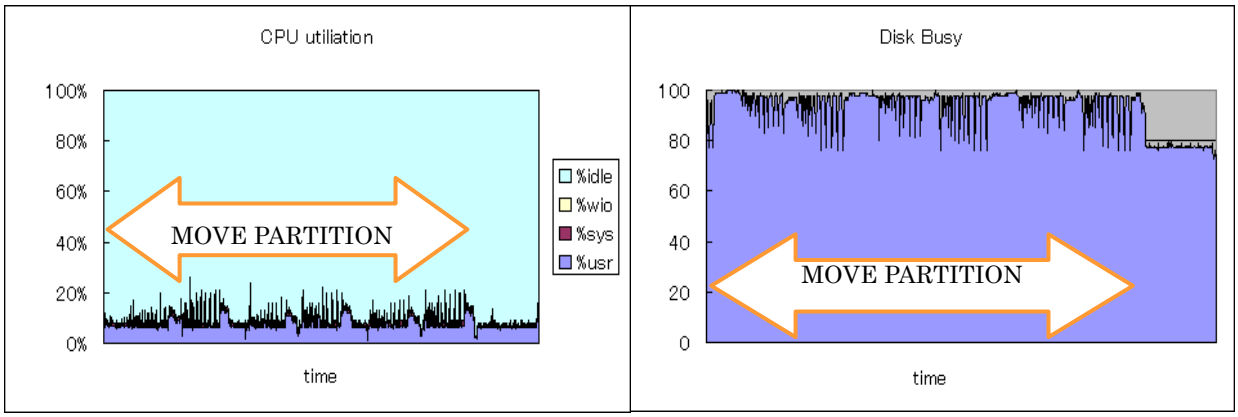


図 6-18 集計クエリ実行時の CPU 使用率、ディスクビジー率

図 6-18からMOVE PARTITION文実行中は、集計クエリのみを実行している場合と比較して、SATAディスクのビジー率が 20%前後高くなっています。またディスクビジー率が 100%に達している時間も多くあります。そのため、クエリのデータ読み込みが遅延し、トランザクションに影響が出たと考えられます。

前述した通り、索引のREBUILD時にはデータ移動時よりもディスク負荷が高くなります。そのためSATAディスクにおいて、クエリによるデータ読み込みと索引の書き込みが同時に行われると、MOVE PARTITION処理も遅延する可能性があります。図 6-19は定常時のMOVE PARTITION文の処理時間を 1 とし、集計処理を行っているノードでMOVE PARTITION文を実行した場合の処理時間を相対値で表しています。

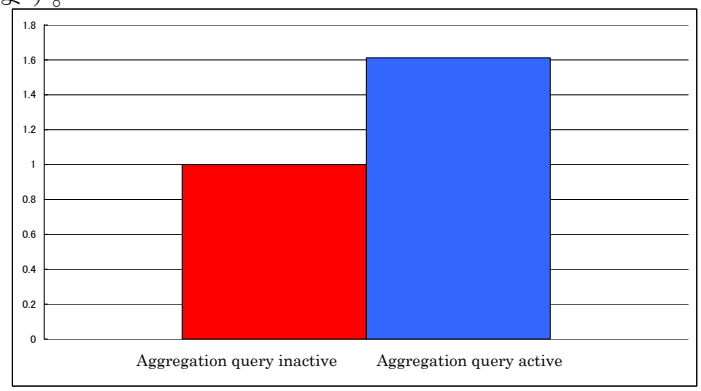


図 6-19 集計クエリ実行時の MOVE PARTITION 時間比

グラフから定常時と比較して約 1.6 倍の時間がかかっていることがわかります。

以上の結果より、MOVE PARTITION 文実行中は、移動先のデータに対するアクセスに留意する必要があります。

6.1.4. MOVE PARTITIONによるILMに関するまとめ

MOVE PARTITION 文を使用した ILM では、特にストレージの物理設計を意識する必要はありません。移動するデータの期間もお客様のシステム要件や環境に合わせて変更することができます。さらに表の断片化の解消や、パーティションの移動と同時にデータ圧縮を実行することも可能です。つまり、様々な要件にあわせた ILM が実現できます。

一方 MOVE PARTITION 文を使用した ILM の場合、以下の点に留意する必要があります。

- FC ディスクの負荷

  『6.1.3.2　オンライン処理を実行しているノードでMOVE PARTITIONを実行した場合』で示したように、オンライン処理を実行しているノードで同時にMOVE PARTITIONを実行すると、データの移動時にFCディスクの負荷が上昇します。iostatなどでディスクの負荷状況を把握した上で、MOVE PARTITION文を実行する、あるいは『6.1.3.3　パーティションを異なるFCディスクに分割した場合』で示したようにFCディスクのRAIDグループを分割することをおすすめします。

- SATA ディスクの負荷

  『6.1,3,1　オンライン業務を実行していないノードでMOVE PARTITIONを実行した場合』で示したように、MOVE PARTITION実行時に索引の移動も行う場合、索引のREBUILD時にSATAディスクの負荷が非常に高くなります。SATAディスクは多重処理に弱いという特性があるので、MOVE PARTITION実行時は、SATAディスクへのアクセスに留意する必要があります。

- CPU 使用率

  『6.1.3.2　オンライン処理を実行しているノードでMOVE PARTITIONを実行した場合』で示したように、シングルDB環境の場合、オンライン業務の行われている時間帯にMOVE PARTITION文を実行すると、CPU使用率も上昇します。sar,vmstatなどでオンライン処理のCPU使用率を監視した上で、オンライン業務に影響を与えない時間帯にMOVE PARTITIONを実行する必要があります。

これらの点を踏まえ、MOVE PARTITION を使用した ILM 計画を立案することで、より効率的なデータ管理を実現することが可能となります。

なお、『6.1.2 MOVE PARTITION実行時の各リソースの使用状況』で示したようにMOVE PARTITIONの所要時間はデータ量や索引のREBUILDをするか否かなどにより変化します。

データセグメント圧縮機能の検証に関するホワイトペーパーは、以下のURLで公開されておりますので、合わせてご参照ください。[5]

6.2. RAIDマイグレーションを利用したILM

Oracle Database の ILM では MOVE PARTITION による表領域の変更によってデータの移動を行い、アクセス頻度が少ないデータを大容量且つ安価なディスク上へ配置することで実現します。前述の MOVE PARTITION の検証結果のとおり、MOVE PARTITION では、ディスクから Oracle プロセスがデータを読み取り、移動先のディスクへデータを書き込んだ後、索引の REBUILD を行うため、データベースサーバでは CPU の負荷があることが確認できました。このため、MOVE PARTITION による ILM の運用では、少なからず業務影響が発生します。

『3.2.1RAIDマイグレーション』で紹介しましたETERNUS DXのRAIDマイグレーション機能を使用することで、RAIDグループ上に作成した論理ボリュームを他のディスクへストレージシステム内で移動させることが可能です。このため、データベースサーバのリソースを使用せずに、高速ディスクから低速ディスクへデータを移動でき、ILM操作の業務への影響を最小限に抑えることができます。

本節では、Oracle Database の ILM と、ETERNUS DX の RAID マイグレーション機能を組み合わせ、効率よく ILM を実現するための手法について、検証の結果をふまえて説明します。

6.2.1. RAIDマイグレーションを利用した効率的な運用と物理設計

RAID マイグレーションを利用した効率的な ILM の運用方法と、運用を考慮したデータベースの物理設計について説明します。

MOVE PARTITIONにてILMを実現する場合は、パーティション単位で表領域を移動できるため、移動前のデータが存在する表領域がどのような構成で、どのように論理ボリュームを使用しているか等は、ILMの運用という観点では問題とはなりません。しかし、RAIDマイグレーションは、ストレージの論理ボリューム単位でデータを移動するため、物理設計を行う際に、ILMの運用を考慮し、ILMの操作対象データの単位に、それ専用の論理ボリュームを用意するように設計するする必要があります。本検証で使用したデータ・モデルの 2007 年と 2008 年を例にすると、図 6-20のように、ILMの単位である 1 年分のデータ量を見積もって、そのデータが格納できる論理ボリュームを作成し、その論理ボリュームには、その 1

[5] 【SPARC Enterprise と Oracle Database 11g 性能検証】
http://primeserver.fujitsu.com/sparcenterprise/news/article/08/0527/
http://primeserver.fujitsu.com/sparcenterprise/documents/data/pdf/fj-gc-spe-dwh-1.2.pdf

年分のデータだけを格納するようにします。

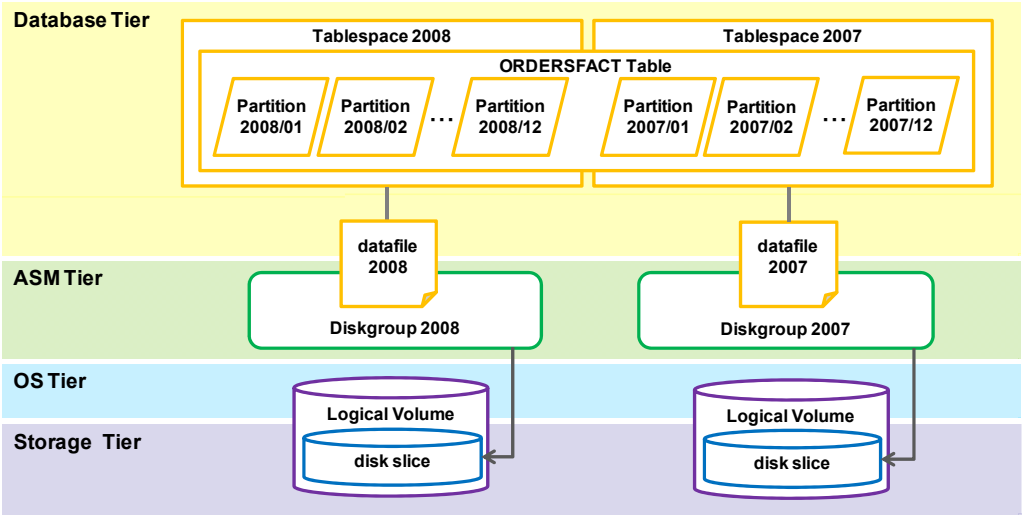


図 6-20 RAID マイグレーションを利用する場合の物理設計

上記例では、12 ヶ月分のデータを 1 つの表領域へ格納し、データファイルも 1 つとしていますが、月単位で表領域を用意し、データファイルも各月に対応したデータファイルを作成することも可能です。また、ILM を行う単位が 1 年であっても、1 ヶ月ごとに論理ボリュームとディスク・グループを用意し、12 回の RAID マイグレーションを行う運用とすることも可能です。しかし、ILM 対象のディスク・グループだけで 12 ヶ月×5 年分=60 ディスク・グループが必要となり、マスタ表やその他領域のディスク・グループを考慮すると、ASM のディスク・グループ数の上限である 63 グループに達してしまいます。後々のシステムの拡張性を考え、ILM のデータ移動単位とディスク・グループが 1 対 1 になるようにし、ディスク・グループ数を抑えるよう設計することをお奨めします。

次に運用手順について考えます。本検証のモデルでは、オンライン業務対象の最新データ 1 年分と、集計業務対象となる過去 5 年分のデータを保持する要件です。ILM のデータの移動は 1 年単位で行いますので、先に述べたように、1 年単位で論理ボリュームを用意します。最新の 2009 年の 1 年分のデータには、FC ディスクを用いて構成した RAID グループ上の論理ボリュームを使用します。過去 5 年分の古いデータは SATA ディスクを用いて構成した RAID グループ上に、1 年毎に 5 つの論理ボリュームを作成し、配置します。

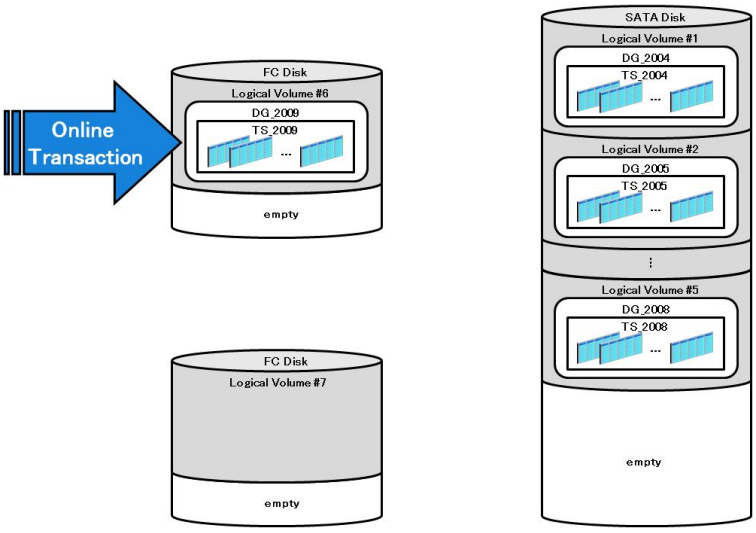


図 6-21 ILM 対象データの初期配置

これを初期配置とします。次の年である 2010 年のディスク・グループ、表領域、パーティションは、用意しておいた論理ボリューム（Logical Volume#7）へ配置するようにします。オンライン業務は徐々に 2010 年のデータに対し処理を行うようになり、2009 年のデータへはオンライン業務は行われなくなります。

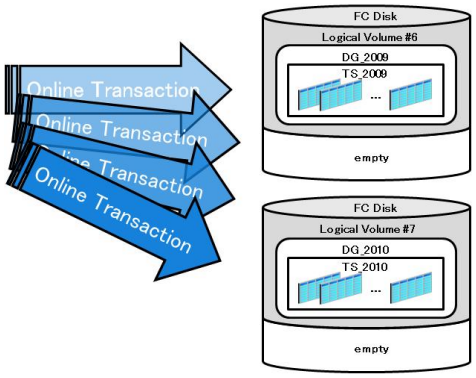
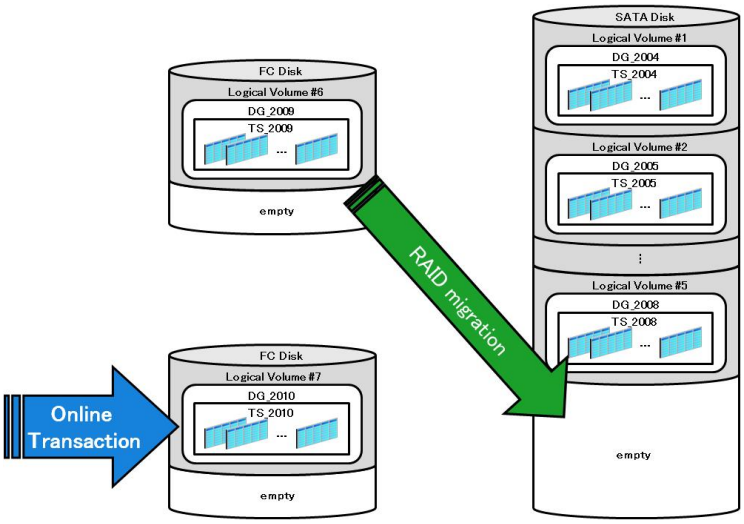


図 6-22 ディスク・グループ・表領域・パーティションの追加

オンライン業務が 2010 年のパーティションに対し行われ、2009 年のデータがオンライン業務で使用されなくなった時点で 2009 年のデータは SATA ディスクへ移動する対象となります。FC ディスク上の 2009 年のデータが格納されている論理ボリューム（Logical Volume #6）は RAID マイグレーションを使用して SATA ディスクへ移動します。



このように、最新データ用に FC ディスクからなる RAID グループを 2 つ用意しておくことで、オンライン業務と RAID マイグレーションによるディスク I/O が競合しなくなります。

次にデータの保存ポリシーを 5 年としていますので、2004 年のデータは削除の対象となります。ここで、2004 年のデータを削除する前に、論理ボリューム（Logical Volume #1）を 2009 年のデータが入っていた FC ディスクへ RAID マイグレーションを使用して移動しておきます。

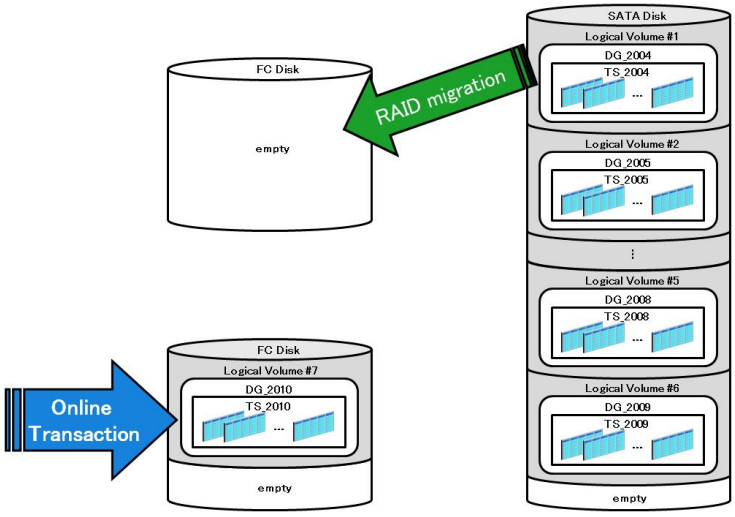


図 6-24 削除対象データの移動

論理ボリュームを移動しておく目的は、2004 年のデータを削除した後、この論理ボリュームを次の 2011 年に使用するためです。2004 年で使用していた論理ボリュームを移動しない場合は、2004 年のデータを削除した上で、ETERNUS DX 上で論理ボリュームを削除し、ETERNUS マルチパスドライバで論理ボリュームを削除したことを認識させます。また、2011 年のデータを格納するために、FC ディスクに新しく論理ボリュームを用意しなければなりませんので、その際に、論理ボリ

ュームの追加、OS/ドライバでの論理ボリュームの認識、スライス作成などが必要となります。2004 年のデータが入っている論理ボリュームを FC ディスクへ移動しておくことで、これらの作業が不要になります。また、低速な SATA ディスク上でデータを削除するより、FC ディスク上でデータを削除した方が処理が速く終了しますので、RAID マイグレーションを行った後で 2004 年のデータを削除するようにします。

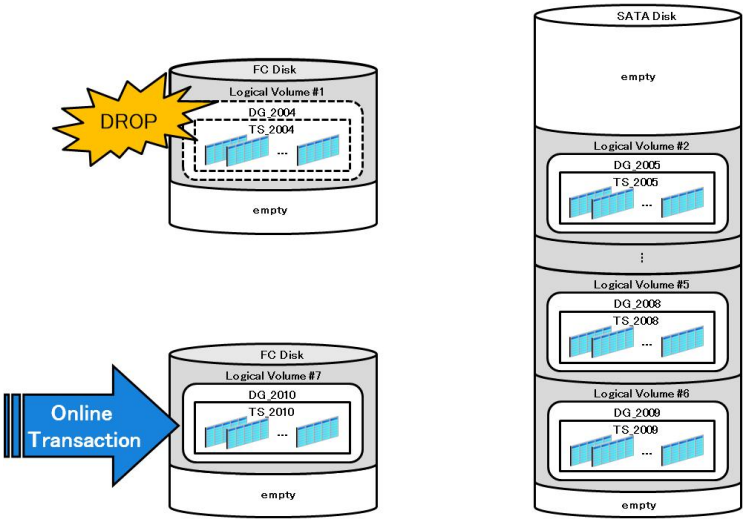


図 6-25 古いデータの削除

最終的には図 6-26のようになります。Logical Volume #1 は次の 2011 年のデータを格納するための論理ボリュームとして準備されており、これは初期配置とほぼ同等の状態といえます。2012 年以降も同様の操作でILMを運用していくことができます。

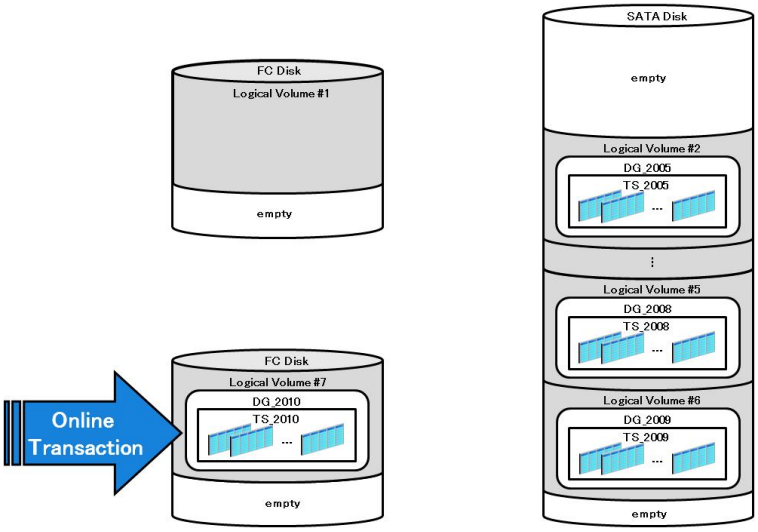


図 6-26 ILM 操作終了後の状態

本節では、RAID マイグレーションを使用した ILM を実現するための運用方法と、運用を考慮したデータベースの物理設計のポイントを説明しました。Oracle の標準機能による ILM と異なる点をまとめると以下です。

- ILM でのデータ移動を行う単位に専用の論理ボリュームを割り当てる
- 最新のデータ格納用に FC ディスクを 2 つ用意し、交互に使用する
- 保存期間を超過したデータの削除は FC ディスクへ RAID マイグレーションし、その後、削除する

なお詳しい手順は、『9.3 RAIDマイグレーションによるILMの手順例』をご参照ください。

6.2.2. RAIDマイグレーションが業務へ与える影響

RAIDマイグレーションが業務へ与える影響について確認しました。図 6-23のとおり、最新のデータに対し、オンライン業務を行っている際に、RAIDマイグレーションを実行します。

図 6-27は定常時（RAIDマイグレーションをしていない時）のオンライン業務のスループットとレスポンスタイムを示しています。平均スループット、レスポンスタイムを 1 として、相対値で表しています。

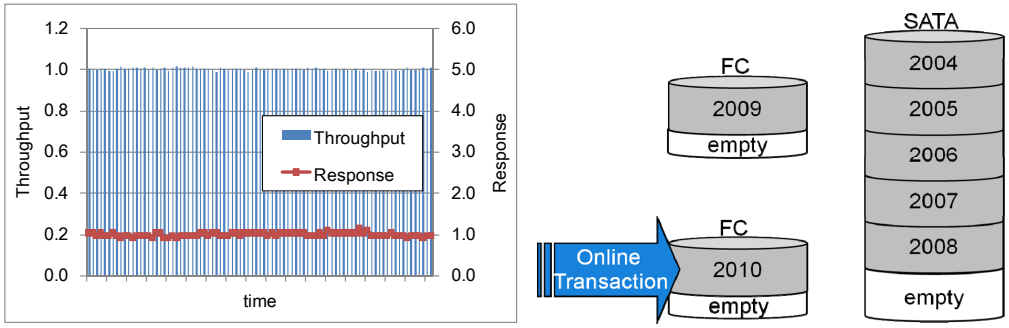


図 6-27 オンライン業務のスループットとレスポンスタイム（定常時）

図 6-28はFCディスク上の 2009 年のデータをSATAディスクへRAIDマイグレーションにより移動しているときのオンライン業務のスループットとレスポンスタイムを示しています。定常時の平均スループット、レスポンスタイムに対する相対値で表しています。

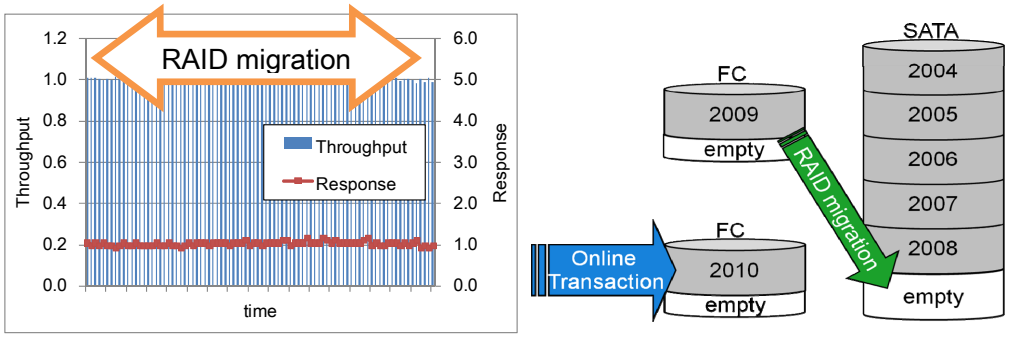


図 6-28 2009 年のデータを SATA ディスクへ RAID マイグレーションしたときの オンライン業務への影響

RAID マイグレーション中、オンライン業務はスループット、レスポンスタイムともに、ほぼ相対値 1 で遷移しており、RAID マイグレーションによるオンライン業務への影響が無いことがわかります。

図 6-29はSATAディスク上の 2004 年のデータをFCディスクへRAIDマイグレーションにより移動しているときのオンライン業務のスループットとレスポンスタイムを示しています。定常時の平均スループット、レスポンスタイムに対する相対値で表しています。

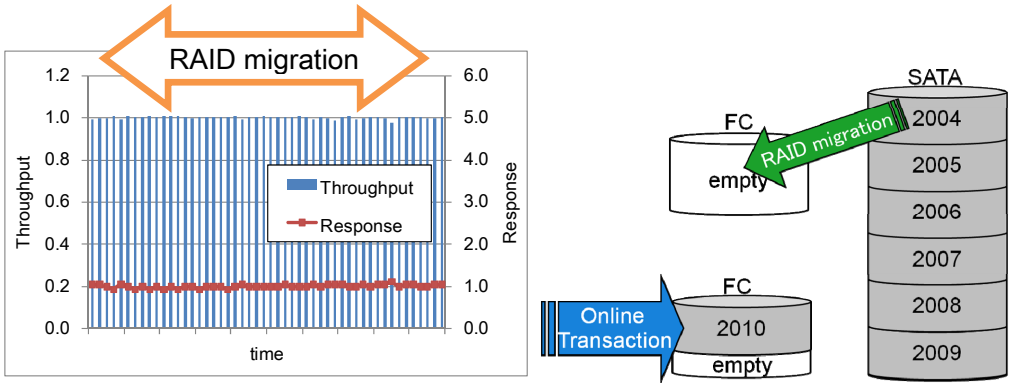


図 6-29 2004 年のデータを FC ディスクへ RAID マイグレーションしたときの オンライン業務への影響

2009 年のデータを SATA ディスクへ RAID マイグレーションしたときと同様、オンライン業務のスループット、レスポンスタイムへの影響はありません。

次にデータベースサーバの CPU 使用率、ディスクビジー率を確認します。はじめに、定常時の CPU 使用率とディスクビジー率について確認します。

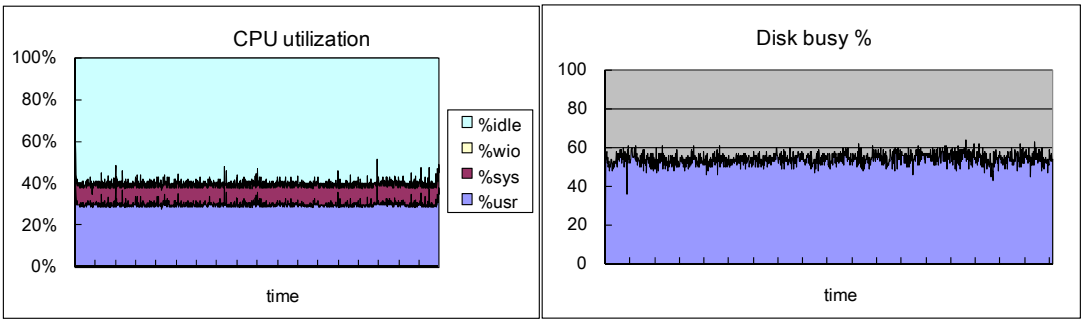


図 6-30 データベースサーバの CPU 使用率とディスクビジー率（定常時）

CPU 使用率は合計 40%前後、ディスクビジー率は 50～60%程度で安定して負荷がかかっています。

RAIDマイグレーションを実行した場合のデータベースサーバのCPU使用率とディスクビジー率を図 6-31と図 6-32に示します。どちらも、定常時とかわらない負荷であったことがわかります。

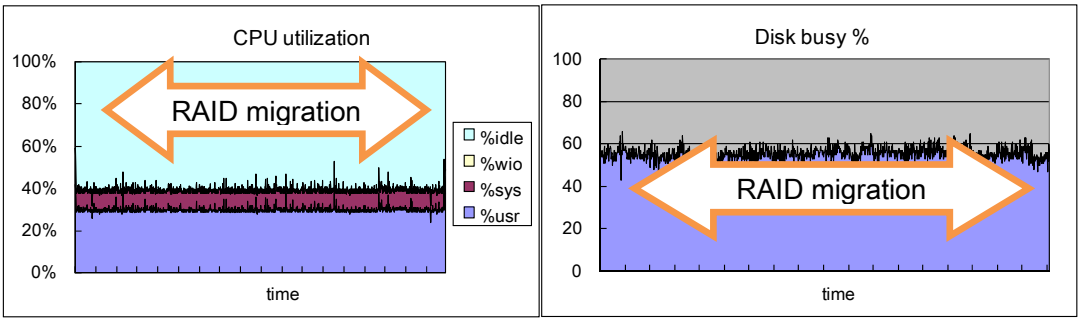


図 6-31 2009 年のデータを SATA ディスクへ RAID マイグレーションしたときの CPU 使用率とディスクビジー率

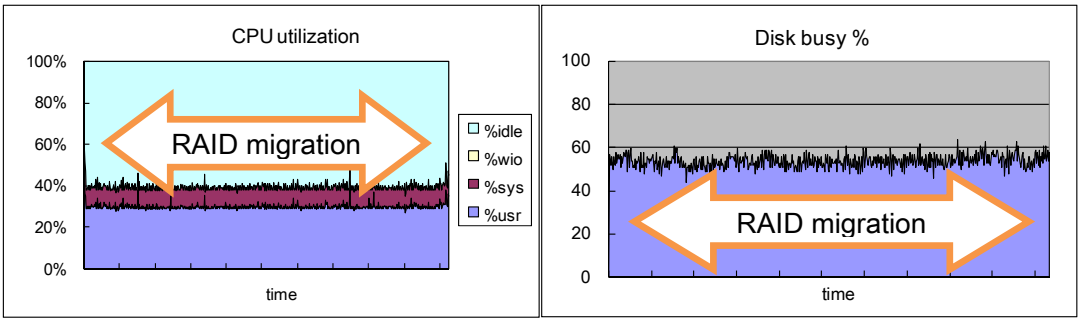


図 6-32 2004 年のデータを FC ディスクへ RAID マイグレーションしたときの CPU 使用率とディスクビジー率

RAID マイグレーションでは、データベースサーバの CPU を使用しないため、CPU 使用率が定常時とかわりません。また、オンライン業務対象のディスクと RAID マイグレーション対象のディスクを分けていますので、ディスクビジー率も定常時とかわりません。

参考①：RAID マイグレーション実行時のサーバ CPU 使用率

業務を停止した状態で、1 年分（約 130GB）のデータをRAIDマイグレーションによって、FCディスクからSATAディスクへ移動させた場合のデータベースサーバのCPU負荷状況を確認しました。図 6-33 RAIDマイグレーション中のデータベースサーバのCPU使用率よりRAIDマイグレーションではデータベースサーバのCPUを使用しないことがわかります。

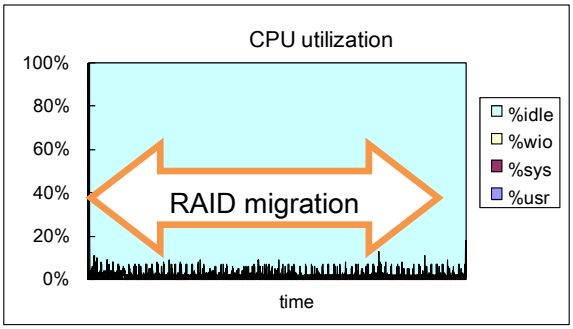


図 6-33 RAID マイグレーション中のデータベースサーバの CPU 使用率

参考②：FC ディスク側の RAID グループが一つの場合

FC ディスクを 2 組用意していない場合はオンライン業務と RAID マイグレーションで、ディスク I/O が競合するため、オンライン業務へ影響がある場合があります。今回の検証モデルでは、定常時 50～60%程度であったディスクビジー率が RAID マイグレーションを実行すると 100%に達してしまい、ディスクネックによる業務影響があることを確認しました。

以上より、本検証のように FC ディスクの RAID グループは 2 つ以上持つ構成が推奨されます。

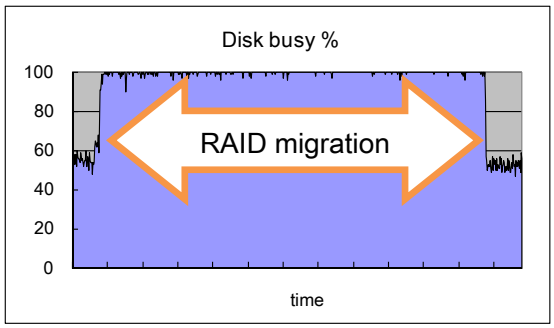


図 6-34 RAID グループが 1 つの場合の RAID マイグレーションの影響

次に集計業務への影響を確認します。はじめに、集計処理の基礎値として、クエリ単体性能を確認します。図 6-35はクエリを単体で実行したときの、CPU使用率とディスクビジー率です。

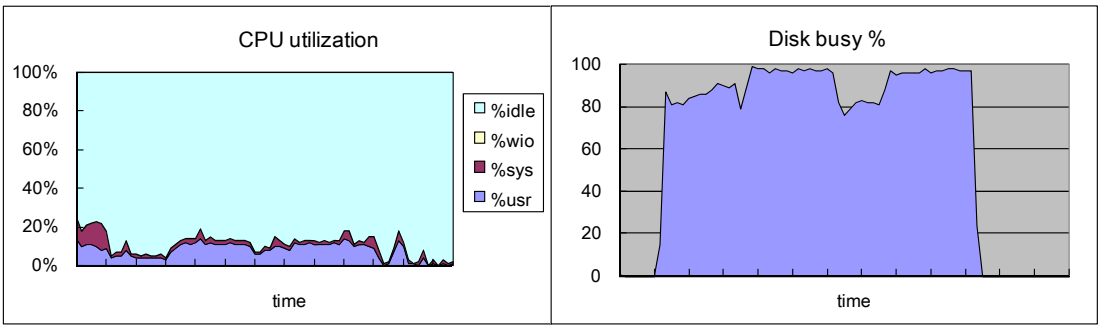


図 6-35 集計処理単体性能

集計処理を実行している SPARC Enterprise M3000 のスレッド数は 1CPU×4 コア×2 スレッドであるため合計 8 スレッドです。集計処理は 1 多重で実行しておりますので、CPU 使用率としては、10%程度でありますが、1 スレッドをほぼ使いきる位の負荷です。ディスクビジー率は、80%以上であり、ディスク負荷は非常に高いことがわかります。

この集計処理を 2008 年のデータに対し順次実行している最中にRAIDマイグレーションによって、FCディスクからSATAディスクへ 2009 年のデータを移動する場合の業務影響を確認します。図 6-36は単体実行時のクエリレスポンスを 1 とした場合の相対値でクエリレスポンスを表しています。

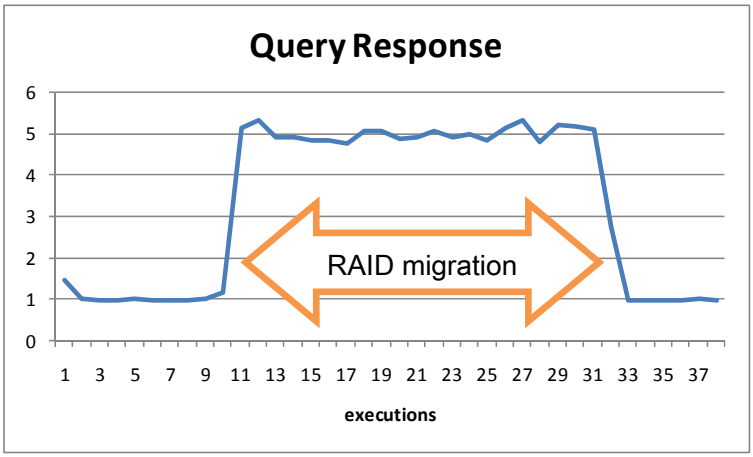


図 6-36 RAID マイグレーション中のクエリレスポンス

RAID マイグレーションを実行すると同時にクエリレスポンスが大きく遅延していることがわかります。RAID マイグレーションが終了すると、クエリレスポンスも元に戻ります。次に、CPU 使用率とディスクビジー率について確認します。

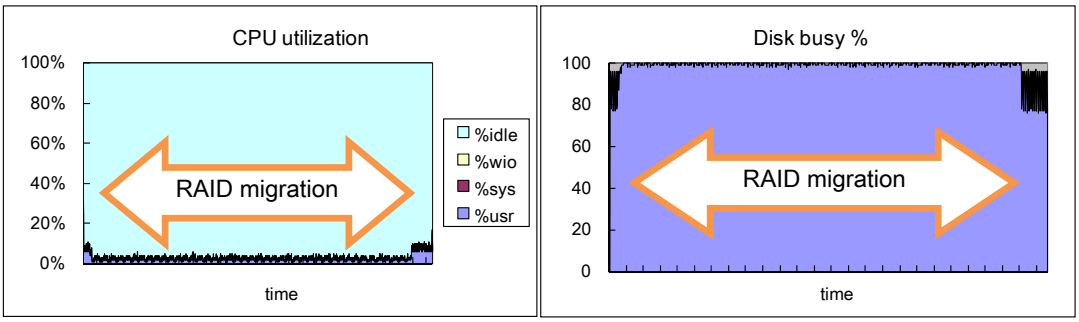


図 6-37 RAID マイグレーションの集計クエリへの影響

RAID マイグレーション開始後、すぐに、CPU 使用率が下がり、RAID マイグレーション終了後に元の CPU 使用率に戻っています。ディスクビジー率を見ると、RAID マイグレーションを開始するとすぐに、ディスクビジー率が 100%に達しています。この結果から、集計処理のディスクからの読み込みと、RAID マイグレーションによるデータからの書き込みが重なったため、ディスクネックとなり、データベース側ではディスクからの読み込みを待機したと考えられます。また、このときの RAID マイグレーションに要した時間を比較すると、以下であり、約 40%遅くなりました。

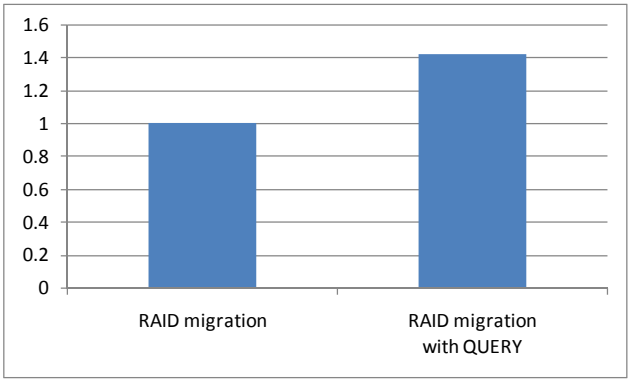


図 6-38 RAID マイグレーションに要した時間

オンライン業務についてはオンライン業務対象の RAID グループと、RAID マイグレーションの対象となる RAID グループは分けて実行していますので、業務への影響はありませんでした。集計業務については、集計業務対象の RAID グループと RAID マイグレーション対象の RAID グループを分けていないため、RAID マイグレーションの影響がありました。また、RAID マイグレーション自体の時間が長くなることも確認できました。FC ディスクから SATA ディスクへ RAID マイグレーションをする場合には、SATA ディスクにアクセスする業務が行われる時間帯を避ける、もしくは SATA ディスクにアクセスする業務を停止して、RAID マイグレーションを実行することを推奨します。

6.2.3. RAIDマイグレーションにかかる時間

RAIDマイグレーションに要する時間について説明します。今回の検証では、2009 年のデータをFCディスクからSATAディスクへ、2004 年のデータをSATAディスクからFCディスクへそれぞれRAIDマイグレーションで移動しています。図6-39はRAIDマイグレーションに要した時間を相対値で表しています。

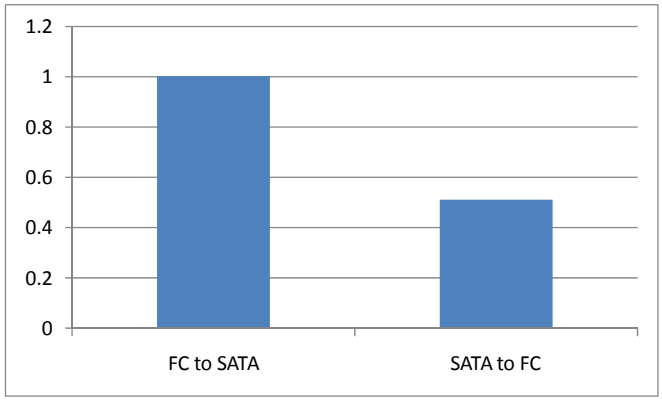


図 6-39 RAID マイグレーションに要した時間

FC ディスクへの書き出しとなる SATA から FC への RAID マイグレーションの方が SATA ディスクへの書き出しとなる FC から SATA に比べて、約半分の時間で終了しました。SATA ディスクは FC ディスクと比較して、容量に関するメリットを享受するため、性能に関する要件を抑えています。そのため、RAID マイグレーションに要する時間は、移動先のディスク種に大きく依存します。

『6.1.2 MOVE PARTITION実行時の各リソースの使用状況 』において、MOVE PARTITIONでは索引の構成や数等によって、ILM作業に要する時間が変わることがわかりました。RAIDマイグレーションの場合は、論理ボリュームをストレージ内で移動するため、論理ボリューム内がどのようになっているかに関係なく、業務とRAIDマイグレーションのディスクI/Oが重ならない限り、RAIDの構成やボリュームのサイズに依存しRAIDマイグレーションにかかる時間が決まります。

6.2.4. RAIDマイグレーションによるILMのまとめ

ETERNUS DX の RAID マイグレーション機能を使用してデータを移動することで、データ移動時の業務への影響を抑えることが可能です。RAID マイグレーションはストレージ内で処理が完結するため、データベースサーバの CPU を使用しません。ディスク I/O に関しては、FC ディスクの RAID グループを 2 組用意し、交互に使用する運用を行うことで、オンライン業務と RAID マイグレーションのディスク I/O の競合をなくすことができます。SATA へ配置したデータに対する集計処理では、集計処理の読み込みと RAID マイグレーションによる書き込みが競合

します。ILM に基づいて SATA へ配置したデータは、アクセス頻度が少ないデータです。このため、集計処理の行われる時間帯を避けて RAID マイグレーションを行うか、RAID マイグレーションを行う時間帯は SATA ディスクにアクセスする業務を停止するなど、運用にて対処してください。

以上、ILM におけるデータの移動に RAID マイグレーションを使用することの有効性が確認できましたが、使用する際には以下の点に留意してください。

- ILM の運用を前提とした専用の設計が必要
  設計手法に関しては、『6.2.1 RAIDマイグレーションを利用した効率的な運用と物理設計』を参照してください。
- ストレージの機能では領域を縮小することはできない
  領域の縮小方法に関しては『9.4.2データ縮小への対応』を参照してください。
- 領域の拡大にはLUNコンカチネーション[6]を使用することはできない
  領域の拡大方法に関しては『9.4.1データ量増加への対応』を参照してください。

6.3. ディスク性能差による集計処理への影響

ILM では、高速な FC ディスクから、SATA ディスクへアクセス頻度の少なくなったデータを移動します。それにより、集計処理のレスポンスにどの程度影響が発生するか検証した結果を以下に示します。

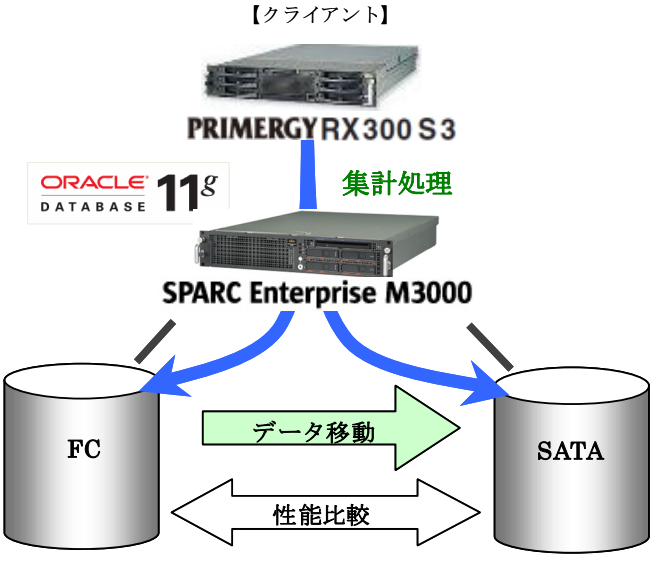


図 6-40　集計処理検証概略図

[6] LUN コンカチネーション：既存の論理ボリュームに対して、未使用領域から領域を切り出して新しい論理ボリュームを作成し、連結することで、既存の論理ボリュームの容量を増やす機能です。

図 6-41は、売上前月比（Q1）と社員別売上（Q2）の集計処理をFCディスク、SATAディスクで実施し、その処理時間を比較したグラフです。

売上前月比（Q1）の集計処理は、FC ディスクに対して、SATA ディスクの処理時間が約 1.2 倍長くかかっています。

一方、社員別売上（Q2）の集計処理に関しては、FC ディスク、SATA ディスク共に処理時間に大きな差は見られません。これは、社員別売上の処理時間が GROUP BY や ORDER BY など、CPU 集中型のクエリであり、ディスクの性能差による影響を受けにくかったためと考えられます。

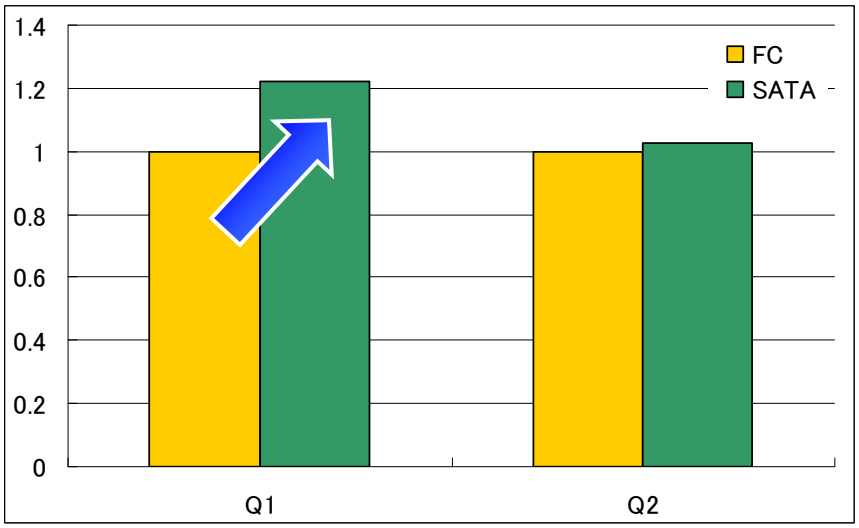


図 6-41　ディスクごとによる集計処理の処理時間の比較

6.4. ILM検証のまとめ

ILM の方式には、RAID マイグレーション（ETERNUS DX の機能）を使用する方式と MOVE PARTITION（Oracle パーティション機能）を使用する方式があります。

これら両方式の特徴について、これまでの検証結果を元に「表 6-1 ILMの方式別の特徴」に示します。

表 6-1 ILM の方式別の特徴

	RAID マイグレーション方式	move partition 方式
メリット	・ ILM 実行時、業務への影響が小さい ・ ILM の所要時間が予測できる	・ ILM の実行により、断片化の解消ができる ・ 領域サイズの縮小が可能
デメリット	・ ILM 専用のデータベース領域設計が必要	・ ILM 実行時、CPU リソースを消費する ・ 索引の REBUILD に時間がかかる

これらの特徴から、両方式を選択する際の指針を整理すると、以下のようになります。

【RAID マイグレーション方式】

RAID マイグレーションは、ストレージ内部で動作するため、OS リソース（CPU）を消費しないことが最大の特長となっています。OS リソースを消費しないため、業務への負荷をかけずに、ILM を実施することが可能となっています。

RAID マイグレーション方式は、ILM によるデータ移動の業務への影響を極力抑えたい場合に有効です。

【MOVE PARTITION 方式】

MOVE PARTITION は、OraclePartitioning 機能の一部として備わっており、ILM を実現するにあたっては、ILM 専用のストレージ設計を行う必要がありません。

また、領域サイズの縮小が可能であったり、断片化の解消やデータセグメントの圧縮により SATA ディスクへのアクセス性能を改善するという可能性も広がります。

MOVE PARTITION 方式は、ILM のフレキシビリティを求める場合に有効です

7. バックアップ

データベースの運用にバックアップは欠かせません。データベースのバックアップにおいても、以下を実施することで、ストレージコスト及び、消費電力を削減することが可能です。

- バックアップ先のディスクに、大容量で低価格な SATA ディスクを選択することによるディスクコスト削減、及びディスク本数削減と消費電力削減
- エコモードを活用して､バックアップ時のみディスクを稼働することによる消費電力削減。（削減量については「9.1 エコモードによる消費電力量削減」をご参照下さい。）

エコモードにおいて、ディスクが停止している場合、ディスクが停止状態から起動するまでに 1 分程度かかるため、その分、バックアップに遅延が発生します。しかしエコモードのスケジュール設定を利用し、日頃の定期的なバックアップの時間帯（ディスク稼働時間）をあらかじめ設定しておくことにより、バックアップ実施前にディスクを起動し、実施後に停止できるため、この問題を回避することができます。

一方､バックアップ先のディスクに FC ディスクではなく､SATA ディスクを選択する場合、バックアップ性能に影響が出ることを考慮する必要があります。

そこで、本項では、バックアップ先のディスク種別（FC ディスク、SATA ディスク）の違いによるバックアップ性能について確認します。

7.1. OP Cによるバックアップ性能

本検証は AdvancedCopy Manager の、スナップショット型高速コピー（One Point Copy）機能を使用しています。One Point Copy（以降、OPC）機能は、任意のタイミングで業務ボリュームをバックアップボリュームへコピーする機能です。

次項より、OPC によるバックアップ性能について検証した結果を示します。

7.1.1. シングルボリュームのバックアップ性能

FC ディスク、SATA ディスクにある約 130GB で構成された 1 ボリュームをそれぞれ、FC ディスク、SATA ディスクへ OPC バックアップし、全 4 パターンのバックアップ性能を比較します。

検証パターン、検証概略図は以下のとおりです。

表 7-1 検証パターン

検証 No	バックアップ元	バックアップ先
1	FC	FC
2	FC	SATA
3	SATA	FC
4	SATA	SATA

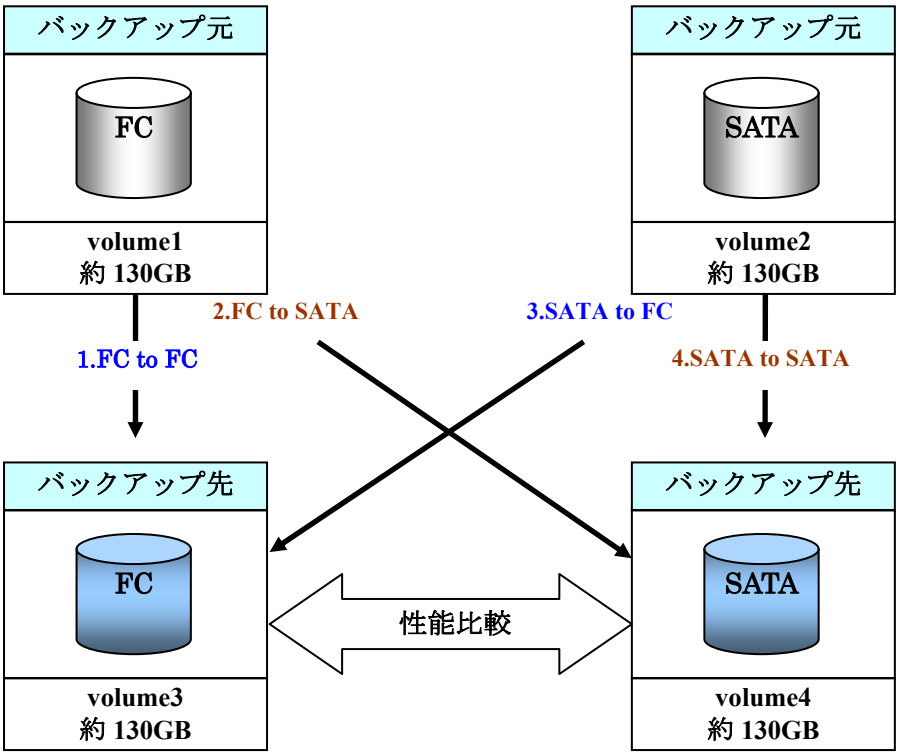


図 7-1 ボリュームバックアップ検証概略

以下に検証結果を示します。

図 7-2 より、バックアップの取得時間を比較すると、バックアップ元が FC ディスク、SATA ディスクのどちらであっても、バックアップ先のディスクに対するバックアップ取得時間に、影響はほぼ見られません。

バックアップ先が SATA ディスクである場合、FC ディスクに比べて、バックアップの取得時間が約 2.2 倍長くかかっており、この時の書き込み量（図 7-3）を確認すると、単位時間あたりの SATA ディスクへの書き込み量は FC ディスクと比べて 45%程度となっています。

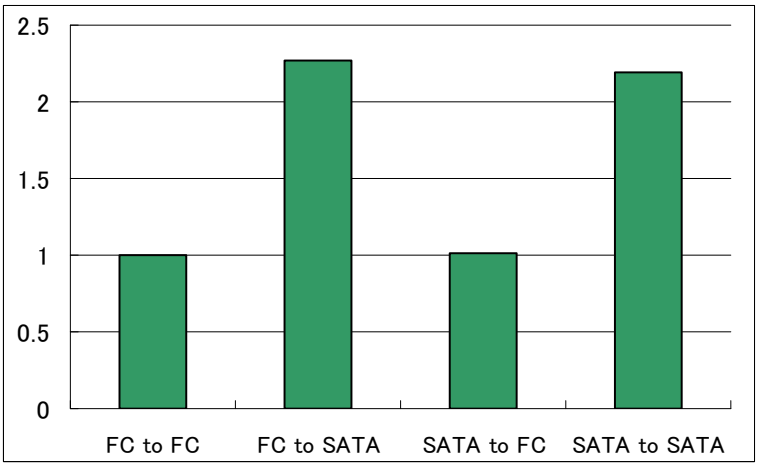


図 7-2 ボリュームバックアップの取得時間の比較

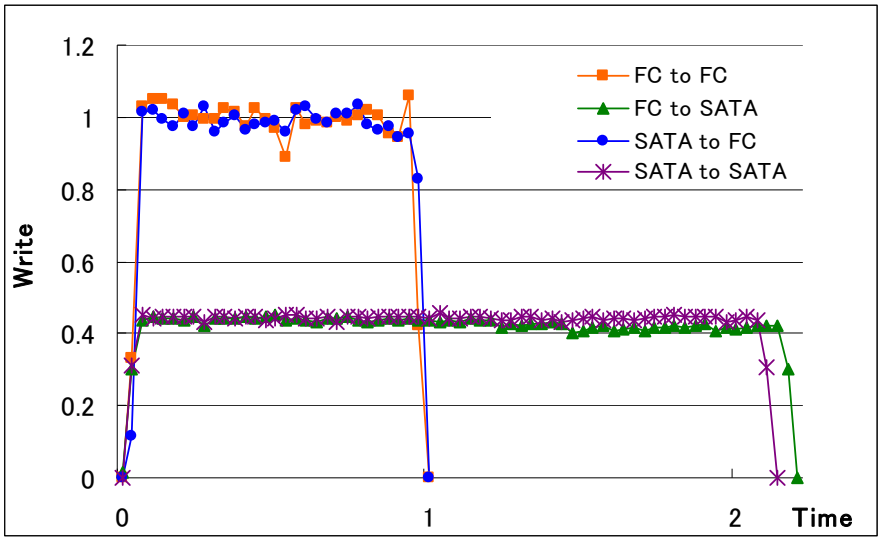


図 7-3 ボリュームバックアップ時の書き込み量と取得時間の比較

7.1.2. データベース全体のバックアップ性能

データベース全体のバックアップ時、バックアップ先のディスク種別（FC ディスク、SATA ディスク）の違いや多重バックアップによって、取得時間にどの程度影響があるかを確認します。

本検証における、バックアップ元は、FC ディスクの RAID グループ内にある 1 ボリュームと、SATA ディスクの RAID グループ内にある 3 ボリューム（合計サイズ：約 500GB）で構成されています。

このバックアップ元を、1RAID グループ、4 ボリュームで構成されたバックアップ先へ OPC による多重（1、2、4）バックアップを実行します。

ここから、1RAID グループ内の複数ボリューム（バックアップ先）に対する多

重書き込みが、バックアップ取得時間へ及ぼす影響について確認します。

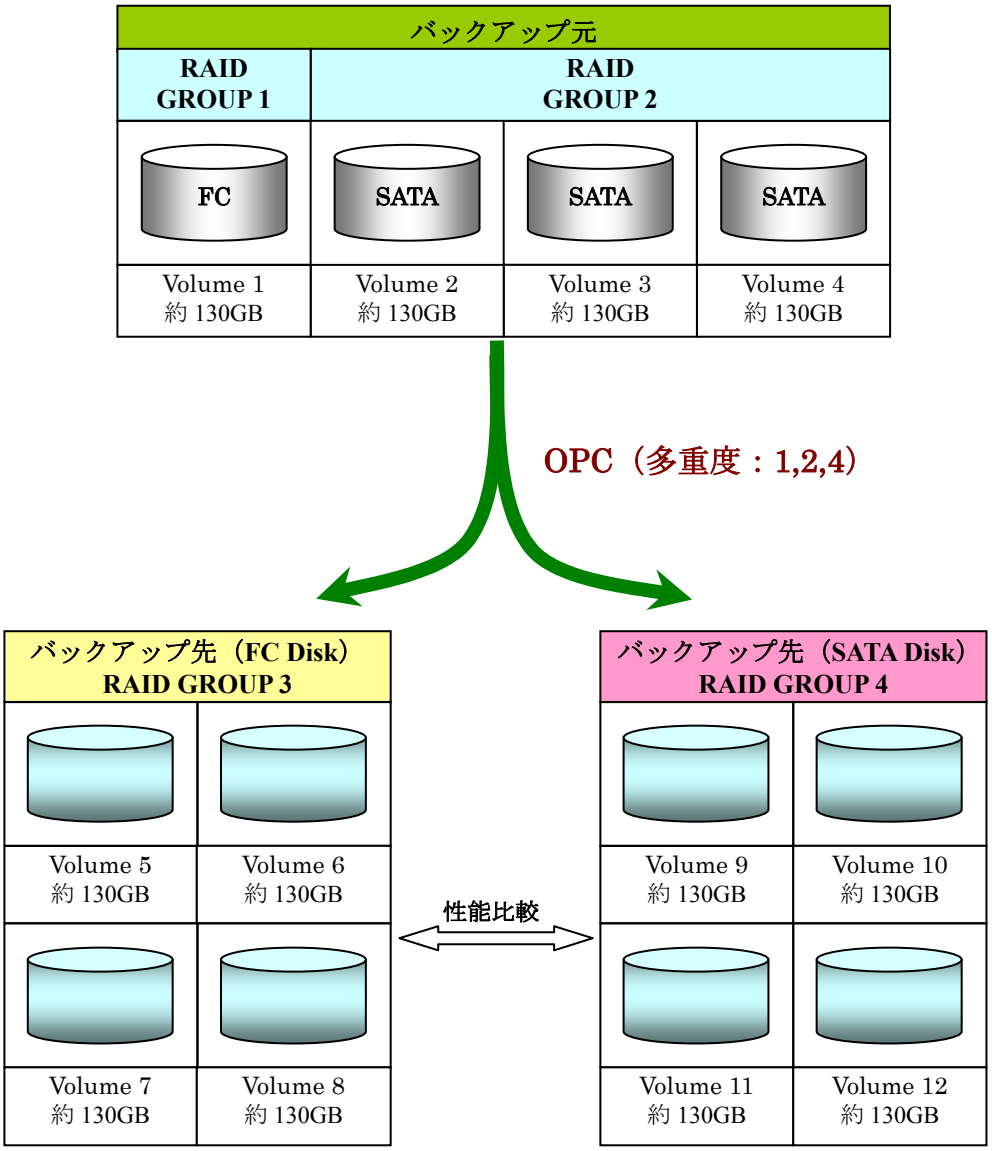


図 7-4 データベース全体バックアップ検証概略図

多重バックアップに関する概略図は以下のとおりです。

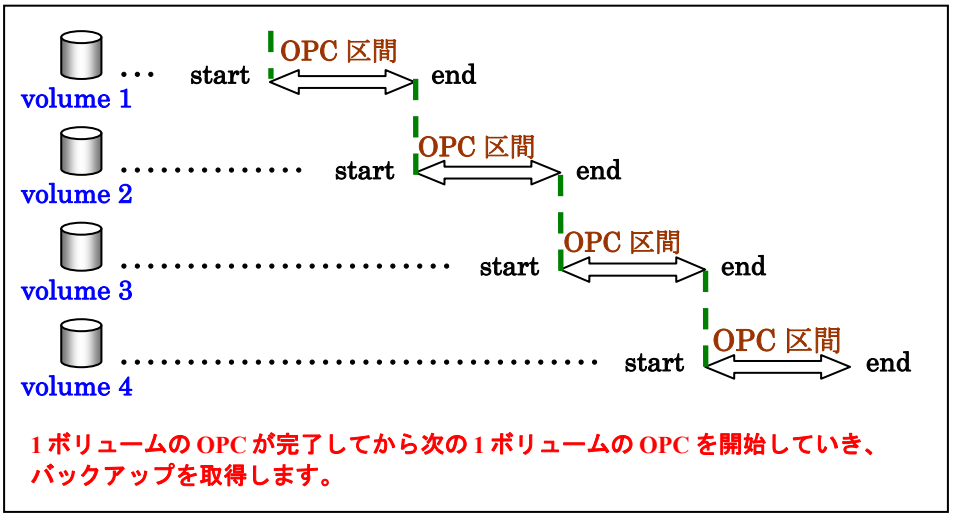


図 7-5 多重バックアップ概略図

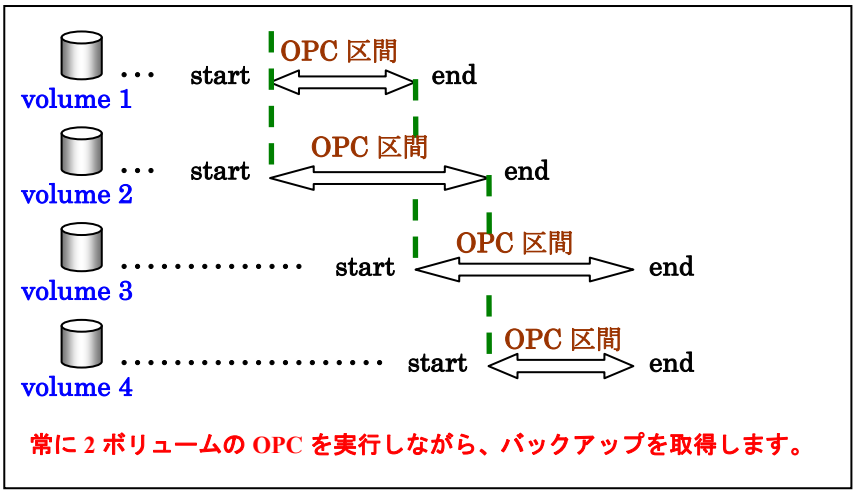


図 7-6 多重バックアップ概略図

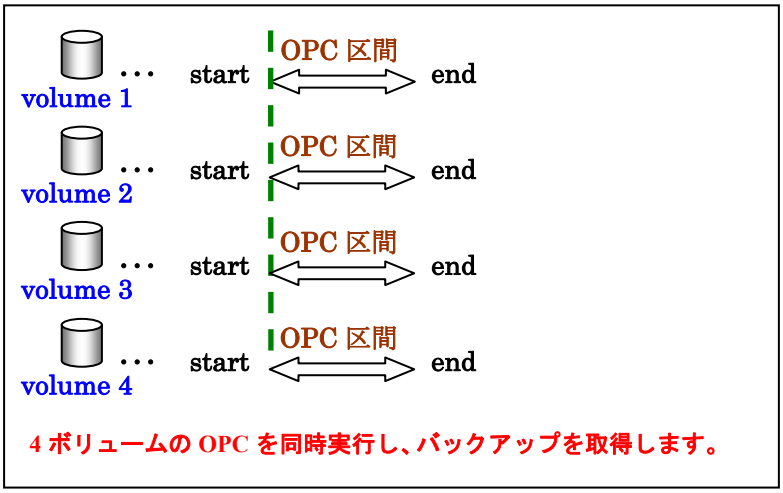


図 7-7 多重バックアップ概略図

以下に検証結果を示します。

図 7-8 を見ると、1 多重のバックアップでは、FC ディスクに対して、SATA ディスクでのバックアップ取得時間が約 2 倍かかっています。

1RAID グループに対して、多重による OPC バックアップでは、FC ディスク、SATA ディスク共に多重度が高くなるにつれて、多重処理しているにも関わらず、1 多重の時よりも取得時間が増加する傾向が見られます。特に、SATA ディスクに対するバックアップにおいては、4 多重にすると、FC ディスクでの 4 多重バックアップよりも、約 3 倍の時間がかかっています。

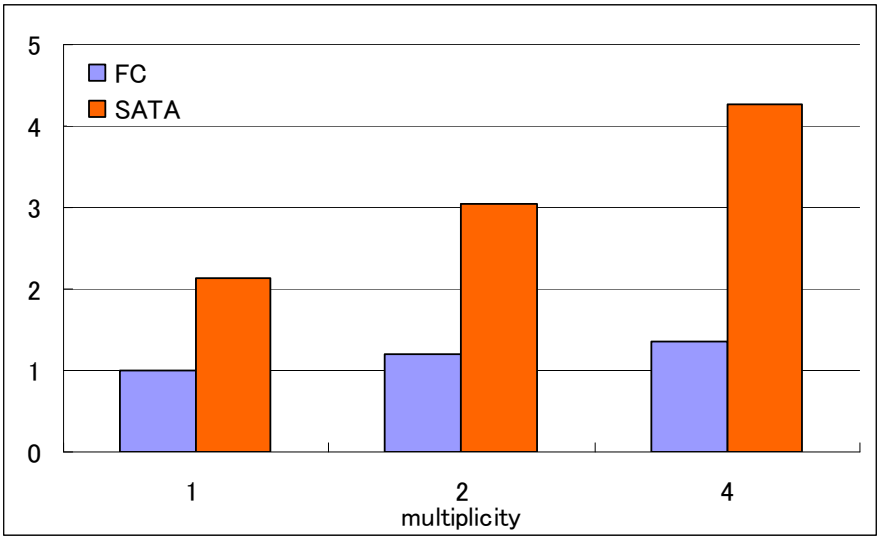


図 7-8 多重度ごとのバックアップ取得時間の比較

7.2. バックアップ検証のまとめ

バックアップの取得時間は、バックアップ元のディスク種別が FC ディスク、SATA ディスクに関わらず、バックアップ先のディスクの書き出し速度に依存します。多重処理を行わない場合、SATA ディスクに対するバックアップの取得時間は FC ディスクに比べて約 2 倍かかります。

同一の RAID グループ内に配置されている複数ボリュームに対して多重バックアップを行うと、FC ディスク、SATA ディスクに限らず、バックアップの取得時間は多重処理を行わない時よりも、かえって長くなってしまいます。

そのため、同一の RAID グループ内に配置されている複数ボリュームに対してバックアップする場合は、多重で実行せず、シリアルで実行することをお勧めします。

もし、多重バックアップを必要とする場合は、バックアップ先のボリュームを複数の RAID グループに分けて配置し、同一 RAID グループ内の複数ボリュームに対して多重書き込みが行われないように配慮する必要があります。

8. 総括

本検証により、ILM における FC ディスクと SATA ディスクの特性を考慮したデータの移動方法および、ILM に適したストレージの構成が明らかになりました。

データの移動方法としては、運用の手軽さ、柔軟さが重視される場合は MOVE PARTITION が適しています。それに対してデータ移動時の業務影響を極力抑えたいという場合には ETERNUS DX の RAID マイグレーションが有効です。お客様の要件にあわせて使い分けることができます。

データを移動させる処理の性能特性から業務影響を抑える設計の手法、運用方法も明確になりました。ニアライン SATA ディスクは FC ディスクに比べ処理性能が劣るため、ILM におけるストレージ設計は、業務における I/O の多寡がポイントとなります。業務上 I/O が多いシステムにおいては、ILM を実現する際に、業務レスポンスに影響が少なくなるよう、データ移動中及び移動後のアクセス量を意識したストレージ設計が重要となります。

また、データベースシステムに欠かせないバックアップ運用において、消費電力削減を意識したエコモード活用及びニアラインディスク活用時のストレージ構成における留意点を明確にしました。

ニアライン SATA ディスクを使用することによるバックアップ取得時間への影響は大きく、取得時間が重要になる場合は、バックアップ先ディスクを複数用意する等の対策が必要です。ストレージ設計時におけるバックアップ要件と、コスト削減効果のバランスを適切にする事が重要です。

以上のように、ILM を用いたストレージコストの削減を実施するにあたり、ディスク種別の各特性を意識した上での効果的な設計手法や運用方法を確立できました。

ILM を実現するにあたり、ストレージ設計やディスク種別における性能特性等、想定される要件が多く存在します。富士通 ETERNUS DX は単一筐体内に豊富なディスク種別を格納し一元管理させ、同時にデータ移動時の業務影響を抑える機能を併せ持つことで、多様なストレージ要件に柔軟な対応を取ることが可能です。

Oracle Database と富士通の SPARC Enterprise、ETERNUS DX による ILM を導入したデータベースシステムでは、高い性能と信頼性はそのままに、ストレージコストの削減と、消費電力の削減を実現できます。

9. Appendix

9.1. エコモードによる消費電力量削減

ETERNUS DX ディスクアレイの持つエコモードは、RAID グループごとに選択されたディスクの回転を止めることができます。エコモードを使用することによって、ディスクの回転を止めるため、エネルギーの節約と空調冷却の軽減に効果があります。ETERNUS DX ディスクアレイのランニングコストは他のストレージシステムに比べてかなり少なく、格納データが増えるほどその効果は顕著になります。ETERNUS DX ディスクアレイは、ハードウェア自体が（エコモードを使用しない場合でも）消費電力を低く抑えるよう設計されています。

エコモードの利用方法の１つとして、ディスク to ディスクバックアップへの適用が考えられます。通常、ディスク to ディスクバックアップにおけるバックアップターゲットのディスクは、バックアップ時以外はアクセスされることがありません。

ETERNUS DX ディスクアレイでは、バックアップターゲット容量用の RAID グループを定義し、エコモードの制御によって RAID グループをバックアップの間だけ回転させることができます。これによりニアラインディスクドライブの消費電力を抑え、空調冷却を軽減することができます。

ETERNUS DX ディスクアレイ エコモード - MAID 電力仕様
500GB/7,200 rpm ニアラインディスクドライブで測定
回転時の電力： 19.3W
停止時の電力： 4.7W

ETERNUS DX400 series の例

ETERNUS DX440 の構成：
ニアラインディスクドライブ 275 本をエコモード設定で 16 時間/日稼動
FC ディスクドライブ 125 本をエコモード設定なしで稼動
その他：4Gbit/s FC ポート x16、コントローラー x2、32GB キャッシュ
消費電力（通常）：8,375kWh x 13.75 円 = 115,156.25 円/月
消費電力（エコモード使用時）：7,037kWh x 13.75 円 = 96,758.75 円
電力コスト削減予想：月 18,397.5 円、年間 220,770 円
注： 装置冷却コストを除く。

9.2. 電源連動機能を使用した計画停止

今日の IT システムを考える上で CO2 削減は非常に重要です。

「SPARC Enterprise」では、省電力化を進め CO2 を削減するために数々の工夫をしています。例えば冷却の取り組みにおいては、エアダクトを採用し、CPU、メモリなど、発熱の大きい部分を集中的に冷却します。また、筐体内を 2 つの冷却グループに分割して冷却することで、冷却効率を向上させています。また冷却ファンの回転数は多段階で制御することができ、冷却効率の向上とあわせて、冷却ファンの回転数を低減し省電力を実現しています。

さらに富士通では、サーバやストレージ等の機器単体での省電力化の取り組みのみならず、システムとしての省電力化にも積極的に取り組んでいます。その一つが「電源連動機能」と「スケジュール機能」です。これは、サーバと周辺機器の電源制御を連動させ、かつシステムの稼動をスケジューリングすることにより、夜間や休日などシステムが稼動する必要のない期間に自動的に機器の電源を切断し、電力消費量を低減するものです。これにより、CO2 排出削減、電力コスト削減に寄与することができます。

今回の検証に使用した「SPARC Enterprise M4000」、「SPARC Enterprise M3000」、「ETERNUS4000」及び後継モデル「ETERNUS DX400 series」は、電源連動機能のインターフェースとして Remote Cabinet Interface（以降 RCI）が標準装備されています。RCI ケーブルで各機器を接続することにより、マスタとなるサーバの電源切断に同期して、他機器の電源を自動的に切断することができます。

また、「SPARC Enterprise M9000、M8000、M5000、M4000、M3000」は、スケジューリングにより自動的に電源を切断するだけでなく、停止状態から自動的に電源を投入することができます。

電源連動とスケジューリング、この 2 つの機能を利用することで、RCI で接続された全機器の電源を指定時間に自動的に切断／投入することができます。

通常、システムを停止する場合はサーバをシャットダウンしてから、ストレージをシャットダウンしなければなりません（起動時は、逆の順番です)。「電源連動」が無い場合は、機器一台ずつ決められた順序で電源切断／投入作業を実施しなければなりません。しかも、業務に支障がないよう、就業後に停止し、業務の始まる早朝に起動しなければなりません。

一般的なサーバとストレージ製品の組み合わせで、このような自動運用を実現

する場合は、別途、電源制御装置および運用管理ソフトウェア等が必要となり、深夜・早朝に人手で実施する場合には、多大な工数が必要となります。

RCI で接続された「SPARC Enterprise」と「ETERNUS DX」であれば、このような追加機器・ソフトウェアのイニシャルコスト・ランニングコストもしくは、人件費は必要ありません。

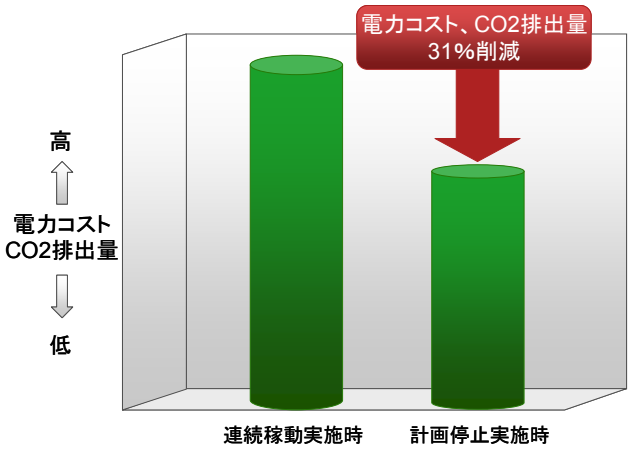
「Oracle Database」、「ETERNUS DX」そして「SPARC Enterprise」の組み合わせで、ILM とあわせて電源連動機能、スケジュール機能を用いることにより、さらなる効率化によるコスト削減と CO2 排出量削減を達成できます。

本検証構成で計画停止を実施した場合の削減例

本検証構成で 1 年間連続稼動した場合と、平日（240 日/年）は 18 時間稼動し、6 時間の停止、休日（125 日/年）は 24 時間の停止を実施した場合を比較。

消費電力（連続稼動実施時）：33,857kWh/年 x 13.75 円/h = 465,533 円/年
消費電力（計画停止実施時）：23,467kWh/年 x 13.75 円/h =322,671 円/年

電力コスト削減予想：年間 142,862 円
CO2 排出量削減予想：年間 4,031kg



注： 装置冷却コストは含みません。
本検証環境のクライアント、ネットワークスイッチ、ファイバチャネルスイッチのコストは含みません。
CO2 原単位：0.388kg-CO2/kWh（富士通（株）集計値）

9.3. RAIDマイグレーションによるILMの手順例

本検証モデルを例に、RAID マイグレーションによる ILM の手順を説明します。

① 新しいディスク・グループの作成

2010 年用のディスク・グループ DG_2010 を作成します。ASM インスタンスへ接続し、以下を実行します。

```
SQL> create diskgroup DG_2010 external redundancy
     disk '/dev/FJSVmplb/rdsk/mplb4s0';
```

② 新しい表領域の作成

表領域 TS_2010 を作成します。データファイルは①で作成した DG_2010 上へ配置します。データベース・インスタンスへデータベース管理者として接続し、以下を実行します。

```
SQL> create bigfile tablespace TS_2010
     datafile '+DG_2010/rac/ts_2010.dbf' size 129G;
```

作成した表領域へのクオータを業務ユーザーへ割り当てます。

```
SQL> alter user ilmuser quota unlimited on TS_2010;
```

③ 新しいパーティションの作成

2010 年のパーティション P201001 から P201012 を作成します。ここからは業務ユーザーで実行します。

はじめに、索引のデフォルト表領域を②で作成した表領域へ変更します。この手順を怠ると、パーティション追加時、自動的に追加されるローカル索引が索引のデフォルト表領域へ格納されます。

```
SQL> ALTER INDEX IDX_ORDERSFACT_ORDERID
     MODIFY DEFAULT ATTRIBUTES TABLESPACE TS_2010;
SQL> ALTER INDEX IDX_ORDERSFACT_SPID
     MODIFY DEFAULT ATTRIBUTES TABLESPACE TS_2010;
```

続いてパーティションを追加します。

```
SQL> alter table ORDERSFACT add partition P201001
     values less than (20100200) tablespace TS_2010;
SQL> alter table ORDERSFACT add partition P201002
     values less than (20100300) tablespace TS_2010;
                          ⋮
SQL> alter table ORDERSFACT add partition P201011
     values less than (20101200) tablespace TS_2010;
SQL> alter table ORDERSFACT add partition P201012
     values less than (20110100) tablespace TS_2010;
```

④ オプティマイザ統計情報の設定

追加したばかりのパーティションでは、オプティマイザ統計情報が収集されていません。この場合、SQL 実行時に動的サンプリング機能によって、オプティマイザ統計情報が収集されるか、デフォルトの値が使用され、実行計画が生成されます。これによって、実行計画が変更されてしまい、性能劣化につながることがあります。これを防止するために、オプティマイザ統計情報を収集する必要がありますが、パーティション追加直後では、データが入っていないため、0 件のデータでオプティマイザ統計情報が収集されます。このため、他のパーティションと比べて、オプティマイザ統計情報が大きく異なり、この場合も、実行計画が変更される可能性があります。これを防止するために、今回は、2009 年のパーティションのオプティマイザ統計情報を 2010 年のパーティションにコピーする運用としました。

```
SQL> begin
  DBMS_STATS.COPY_TABLE_STATS (
    ownname      =>  'ILMUSER',
    tabname      =>   'ORDERSFACT',
    srcpartname  =>   'P200901',
    dstpartname  =>   'P201001',
    scale_factor =>    1,
    force        =>    FALSE);
  end;
  /
```

上記のコマンドをコピー元（srcpartname）とコピー先（dstpartname）を変更し、12 ヶ月分実行します。

⑤ 古くなったデータの移動

2009 年のデータが入っている論理ボリュームを SATA ディスク側 RAID グループへ ETERNUS DX の機能を使用して RAID マイグレーションを実行します。

⑥ 削除対象データの移動

削除対象の 2004 年のデータが入っている論理ボリュームを FC ディスク側 RAID グループへ ETERNUS DX の機能を使用して RAID マイグレーションを実行します。

⑦ パーティションの削除

2004 年のパーティション P200401 から P200412 を削除します。

```
SQL> alter table ORDERSFACT drop partition P200401;
SQL> alter table ORDERSFACT drop partition P200402;

                              ⋮
SQL> alter table ORDERSFACT drop partition P200411;
SQL> alter table ORDERSFACT drop partition P200412;
```

⑧ 表領域の削除

2004 年の表領域 TS_2004 を削除します。データベース管理者として接続し、以下を実行します。

```
SQL> drop tablespace TS_2004 including contents and datafiles;
```

⑨ ディスク・グループの削除

2004 年のディスク・グループ DG_2004 を削除します。今回は RAC 構成ですので、削除コマンドを実行するノード以外では、削除対象のディスク・グループを dismount しておく必要があります。ディスク・グループの削除を実行しないノードのインスタンスへデータベース管理者として接続し、以下を実行します。

```
SQL> alter diskgroup DG_2004 dismount;
```

ディスク・グループを削除します。ASM インスタンスへ接続し、以下を実行します。

```
SQL> drop diskgroup DG_2004 including contents;
```

9.4. RAIDマイグレーションによるILMの補足

RAID マイグレーションによる ILM を行う場合、留意しておく必要のある領域の増減について説明します。

9.4.1. データ量増加への対応

運用中のオンライン業務の増加に伴い、表領域サイズが足りなくなることがあります。その場合、表領域サイズを拡張する必要があります。

表領域サイズの拡張方法としては、以下の 3 つの方法が考えられます。

- RAID マイグレーションによる論理ボリュームのサイズ拡張
- 表領域へのデータファイルの追加
- ディスク・グループへのディスクの追加

(1) RAID マイグレーションによる論理ボリュームのサイズ拡張

ETERNUS DX で論理ボリュームのサイズを拡張する方法には、RAID マイグレーションとコンカチネーションの 2 つの方法があります。ただし、コンカチネーションを行ったボリュームは RAID マイグレーションできないという仕様上の制限があるため、ILM のシナリオにおいては、コンカチネーションを使用することができません。従って、サイズ拡張は、RAID マイグレーションで実施します。具体的な手順は以下のとおりです。

① SATA ディスクから FC ディスクへの RAID マイグレーション時に、論理ボリュームのサイズ拡張を行います。
② OS コマンドで論理ボリューム上のスライスをサイズ拡張します。
③ ディスク・グループ内のディスクをリサイズ（拡張）します。
④ 表領域をリサイズ（拡張）します。

なお、この方法ではSATAディスクからFCディスクへのRAIDマイグレーション時にしかサイズ拡張できないため、ある程度、余裕を持って表領域サイズを見積もることをお勧めいたします。それでも、予想外の業務増大によりサイズが足りなくなった場合には、後述の「(2) 表領域へのデータファイルの追加」または「(3) ディスク・グループへのディスクの追加」で対処してください。

(2) 表領域へのデータファイルの追加

表領域にデータファイルを追加することで、表領域全体のサイズを拡張することができます。

以下の条件すべてに該当する場合には、この方法でサイズを拡張します。

- 表領域タイプが SMALLFILE の場合
- ディスク・グループ数の上限に余裕がある場合

具体的な手順は以下のとおりです。

① ETERNUSmgr で新規の論理ボリュームを追加します。
② grmpdautoconf を実行し、新規の論理ボリュームを OS に認識させます。
③ 新規の論理ボリュームにスライスを作成します。
④ ディスク・グループを作成します。
⑤ 表領域へデータファイルを追加します。

なお、この方法の場合、以下の点に注意する必要があります。

- RAID マイグレーションの際、もともとあった論理ボリュームと追加した論理ボリュームの両方を SATA ディスクへ移動させる必要があります。
- 論理ボリュームを追加するとディスク・グループも増えるので、ディスク・グループ数の上限に達しないよう注意してください。
- 2 つのディスク・グループを 1 つにまとめることはできないので、古くなって削除するまでこの状態（論理ボリュームが 2 つ）で残ります。1 つの論理ボリュームに、まとめたい場合は、MOVE PARTITION で 1 つの大きな領域へ移動させてください。
- 表領域タイプが BIGFILE の場合、1 表領域に対してデータファイル数が 1 個という制限があるため、この方法を使用することはできません。

(3) ディスク・グループへのディスクの追加

ディスク・グループに新規ディスクを追加することで、ディスク・グループのサイズを拡張することができます。拡張してできた空き領域を利用して、表領域のサイズを拡張します。

以下のいずれかの条件に該当する場合には、この方法でサイズを拡張します。

- 表領域タイプが BIGFILE の場合
- ディスク・グループ数を増やせない（増やしたくない）場合

具体的な手順は以下のとおりです。

① ETERNUSmgr で新規論理ボリュームを追加します。
② grmpdautoconf を実行し、新規論理ボリュームを OS に認識させます。
③ 新規論理ボリュームにスライスを作成します。
④ ディスク・グループにスライスを追加します。
⑤ データファイルをリサイズ（拡張）します。

なお、この方法の場合、以下の点に注意する必要があります。

- RAID マイグレーションの際、もともとあった論理ボリュームと追加した論理ボリュームの両方を SATA ディスクへ移動させる必要があります。
- ディスク・グループへのスライス追加の際にリバランスによる I/O が発生するため、業務への影響が考えられます。そのため、ディスク・グループへのスライス追加は、業務にあまり影響がないタイミングで実施してください。

【RAID グループの空き領域が足りない場合の対処】

上記(1)～(3)のいずれの方法にも該当しますが、論理ボリュームのサイズ拡張や新規論理ボリュームの作成を行う際にRAIDグループ内の空き領域が足りない場合には、LDE（Logical Device Expansion）[7]を使用してRAIDグループのサイズ拡張を行うことができます。

9.4.2. データ縮小への対応

領域を大きく見積もりすぎた場合、古いデータの入った論理ボリュームを SATA 側へ移動する際に、論理ボリュームサイズを縮小したいという要件が考えられます。しかし、論理ボリュームサイズの縮小は、ストレージの機能では実現できないため、別の方法でサイズ縮小を実現する必要があります。

論理ボリュームサイズの縮小方法としては、以下の２つの方法が考えられます。

- MOVE PARTITION による ILM
- RAID マイグレーションによる ILM＋MOVE PARTITION

(1) MOVE PARTITION による ILM

ILM 対象の論理ボリュームの現行サイズより小さなサイズの論理ボリュームを新規作成し、これらの論理ボリューム同士で、MOVE PARTITION を使用して ILM を実施します。

(2) RAID マイグレーションによる ILM＋MOVE PARTITION

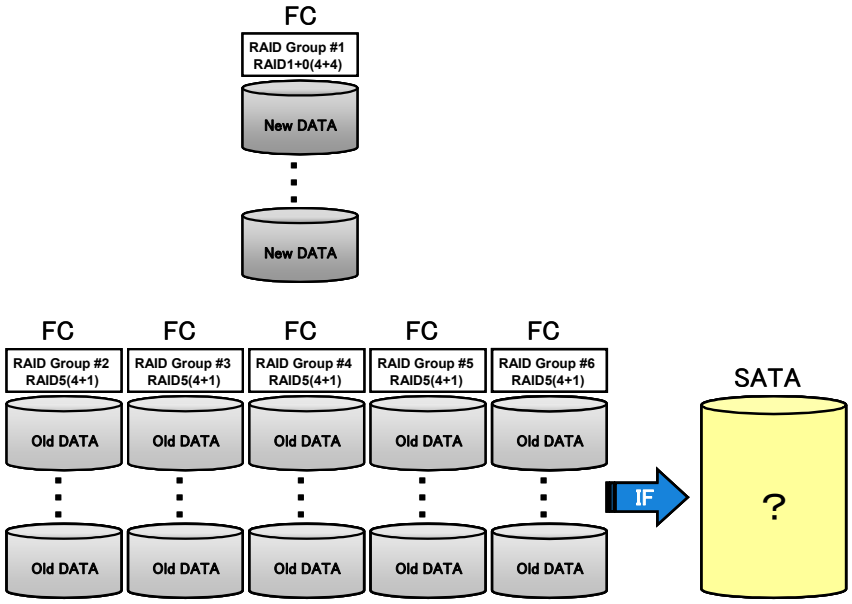
RAID マイグレーションによって、オンライン業務への影響を抑えつつ、ILM 対象の論理ボリュームを SATA ディスクへ移動し、その後、オンライン業務に影響がないタイミングでSATAディスク内でMOVE PARTITIONを実行することにより、領域サイズを縮小します。

[7] LDE(Logical Device Expansion)： 業務を中断することなく RAID グループに新しいディスクを増設し記憶容量の増加を可能にする機能。

9.5. ディスクのコストについて

ILM における、SATA ディスクを使用したコスト削減について説明します。

以下のように、古いデータを FC ディスクへ格納する場合と SATA ディスクへ格納する場合で、どの程度コスト削減につながるかを確認します。



例として、表 9-1のように全てのデータをFCディスクへ保存するディスク構成であったとします。

表 9-1 ディスク構成例

Raid Group	Raid Level	Disk Drive	Usage
1	RAID1+0（4+4）	FC ディスク 146GB (15,000rpm)×8 本	最新データ格納用
2	RAID5（4+1）	FC ディスク 146GB (15,000rpm)×5 本	古いデータ格納用 1
3	RAID5（4+1）	FC ディスク 146GB (15,000rpm)×5 本	古いデータ格納用 2
4	RAID5（4+1）	FC ディスク 146GB (15,000rpm)×5 本	古いデータ格納用 3
5	RAID5（4+1）	FC ディスク 146GB (15,000rpm)×5 本	古いデータ格納用 4
6	RAID5（4+1）	FC ディスク 146GB (15,000rpm)×5 本	古いデータ格納用 5

古いデータの格納に使用できる容量は RAID グループ 2、3、4、5、6 を合わせて、146GB×20 本（4 本×5 グループ）＝ 2920GB です。これを大容量で低価格な SATA ディスク（750GB、7,200rpm）に置き換えると、RAID5（4+1）で 750GB×4 本=3000GB となり、使用するディスクの本数は 25 本から 5 本に削減することができます。

また、図 9-1 より、ディスクドライブにかかる費用を相対値による比較でみると、SATA ディスクを使用したディスク構成の場合、今回の例では、ディスク費用が約 60%削減されることになります。

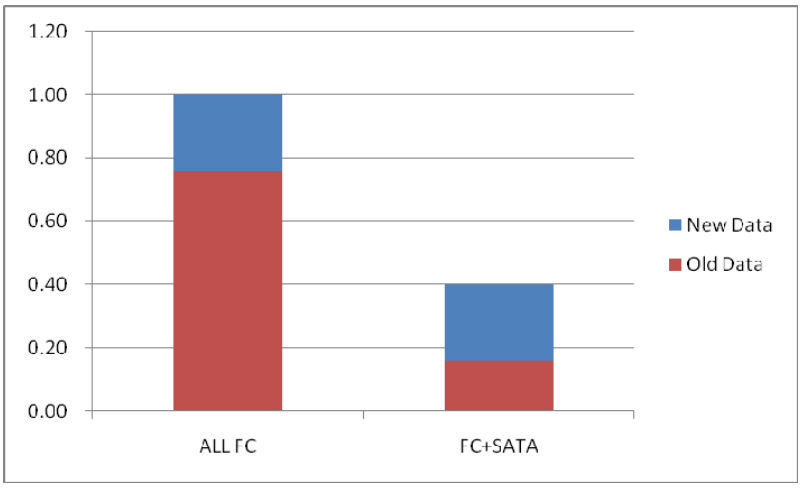


図 9-1 ディスクドライブにかかる費用の比較

以上のことから、大容量で低価格な SATA ディスクを有効に使用することで、ディスク本数や、ストレージにかかる費用を大幅に削減することが可能です。

日本オラクル株式会社
〒107-0061
東京都港区北青山 2-5-8
オラクル青山センター

富士通株式会社
〒105-7123
東京都港区東新橋1-5-2
汐留シティセンター