



Oracle ホワイト・ペーパー  
2010年10月

## Oracle Exadata Database Machine および Oracle Exadata Storage Server の技術概要

## 免責事項

本書は、オラクルの一般的な製品の方向性を示すことが目的です。情報を提供することだけが目的であり、契約とは一切関係がありません。商品、コード、または機能を提供するものではなく、購入の判断にご利用いただくためのものではありません。オラクルの製品に関して記載されている機能の開発、リリース、および時期については、弊社の裁量により決定されます。

はじめに.....	2
Oracle Exadata 製品ファミリー .....	3
Exadata Database Machine.....	3
Exadata Storage Server .....	7
Exadata Database Machine のアーキテクチャ.....	11
データベース・サーバー・ソフトウェア.....	13
Exadata Storage Server Software .....	15
Exadata の Smart Scan 処理 .....	15
Exadata Hybrid Columnar Compression .....	19
Exadata による I/O リソース管理.....	20
Exadata を使用したサービス品質 (QoS) 管理 .....	21
結論.....	27

## はじめに

Oracle Exadata Database Machine は Oracle Database のホスティングを目的とした、配置の簡単なソリューションであり、最高レベルのデータベース・パフォーマンスを実現します。Exadata Database Machine は、データベース・サーバー、Oracle Exadata Storage Server、ストレージ・ネットワークを実現する InfiniBand ファブリック、および Oracle Database をホストするために必要なその他すべてのコンポーネントで構成された、"ワンボックスのクラウド"です。Exadata Database Machine は、オンライン・トランザクション処理 (OLTP)、データウェアハウジング (DW)、複合ワークロードの統合向けに、極めて高い I/O と SQL 処理パフォーマンスを実現します。Oracle Real Application Clusters (Oracle RAC) と Exadata ストレージを使用した超並列グリッド・アーキテクチャを通じて、あらゆる種類のデータベース・アプリケーションに卓越したパフォーマンスがもたらされます。使いやすく管理も簡単な Database Machine と Exadata ストレージは、画期的なパフォーマンスと I/O のリニアなスケーラビリティを実現するだけでなく、ミッションクリティカルな可用性と信頼性を提供します。

Exadata Storage Server は、Exadata Database Machine にとって不可欠なコンポーネントです。卓越したパフォーマンスは、この製品のいくつもの機能を通じて実現されます。Exadata ストレージはデータベース認識型のストレージ・サービス (データベース処理をデータベース・サーバーからストレージにオフロードする機能など) を提供しますが、このサービスは SQL 処理およびデータベース・アプリケーションに対して透過的に提供されます。したがって、問合せ対象表に含まれるすべてのデータではなく、アプリケーションに要求されたデータのみが返されます。Exadata Smart Flash Cache は I/O 操作を高速化することで、Oracle Database の処理パフォーマンスを大幅に向上します。このフラッシュは、インテリジェントなデータベース・オブジェクトのキャッシング機能を提供することで、物理的な I/O 操作を回避します。Database Machine 上で実行される Oracle Database は、世界初のフラッシュ対応データベースです。Exadata ストレージが提供する高度な圧縮テクノロジーである Exadata Hybrid Columnar Compression は、通常 10 倍以上のデータ圧縮レベルを実現します。Exadata の圧縮技術は、桁違いに効果的なデータ送信を実現します。また Oracle Exadata Database Machine は、世界でもっともセキュアなデータベース・マシンでもあります。Oracle Database の優れたセキュリティ機能を利用して構築された Exadata ストレージは、完全に暗号化されたデータベースに対する問合せを、オーバーヘッドをほとんど発生させることなく、1 秒当たり数百 GB という速度で実行します。これらとその他多数の製品機能の組合せが基盤となり、Exadata Database Machine の卓越したパフォーマンスを支えています。

また Exadata Database Machine は Oracle Exalogic Elastic Cloud と連携することも、単独で機能することもできます。Exalogic Elastic Cloud は、Oracle Fusion Middleware と Oracle Fusion Application を実行するために最適なプラットフォームです。Exadata と Exalogic の組合せは、ハードウェアとソフトウェアが包括的に設計されたソリューションとして、Oracle E-Business Suite、Siebel、PeopleSoft などのあらゆるエンタープライズ・アプリケーションに卓越したパフォーマンスを提供します。

## Oracle Exadata 製品ファミリー

Oracle Exadata 製品ファミリーの基盤となるのは、Oracle Exadata Database Machine (Database Machine) です。Database Machine は、最適なパフォーマンスを実現するエンタープライズ・データベースを迅速かつ簡単に配置するためのコンポーネントをすべて備えた完全統合型のデータベース・システムです。Exadata Storage Server (Exadata ストレージまたは Exadata セル) は、Database Machine に含まれる Oracle Database 用ストレージとして使用されます。また、既存の Database Machine 構成を拡張する際にも使用されます。

### Exadata Database Machine

Database Machine は配置したその日からすぐに使用できる事前構成済みシステムであるため、データベース配置プロセスの多数の統合作業、コスト、時間が削減されます。また、これらのよく知られた構成の場合、Oracle Support は提供するサービスに精通しているため、優れたサポート・エクスペリエンスが得られます。OLTP や DW、またはこれらの組合せを含むさまざまなアプリケーション向けにデータベースを配置する場合や、複数のデータベースの統合プラットフォームとして使用する場合に、共通インフラストラクチャによってデータセンターの効率化へ向けた大きな機会が生まれます。これが、真の"ワンボックスのクラウド"です。



Exadata Database Machine X2-8

Exadata Database Machine には2つのバージョンがあります。Exadata Database Machine X2-2 は、192GB のメモリを搭載した2台の12コア・データベース・サーバーと3台の Exadata Storage Server から、768GB のメモリを搭載した8台の12コア・データベース・サーバーと14台の Exadata Storage Server までの広範囲にわたって、すべて1ラックで提供されています。Exadata Database Machine X2-8 は、2TB のメモリを持つ2台の64コア・データベース・サーバーと14台の Exadata Storage Server で構成されており、1ラックで提供されています。X2-2 は Exadata Database Machine ファミリーへの入口として手ごろでありながら、1ラックでは最大レベルの拡張性を備えています。X2-8 は大容量のメモリを必要とする大規模配置や、1ラックに多数のデータベースを統合する必要がある配置に適しています。

いずれのバージョンも、データベース・ソフトウェアとして Oracle Database 11g Release 2 を使用しています。

#### Exadata Database Machine X2-2

Exadata Database Machine X2-2 では、配置するデータベースのサイズ、パフォーマンス、および I/O 要件に応じて、フル・ラック、ハーフ・ラック、クォーター・ラックという 3 つのバージョンが提供されています。それぞれのバージョンはオンラインで別のバージョンにアップグレードできるため、処理要件が拡大した場合にもスムーズなアップグレード・パスが確保されています。すべての Database Machine X2-2 に共通する仕様は、次のとおりです。

- 業界標準の Oracle Database 11g データベース・サーバーに、2 基の 6 コアの Intel® Xeon® X5670 プロセッサ (2.93GHz で動作)、96GB のメモリ、4 台の 300GB 10,000RPM SAS ディスク、2 つの 40Gb/秒の InfiniBand ポート、2 つの 10Gb/秒のイーサネット・ポート、4 つの 1Gb/秒のイーサネット・ポート、ホットスワップ対応のデュアル冗長電源が事前構成されています。また、Database Machine のオペレーティング・システムとして、データベース・サーバー上に Oracle Linux 5 Update 5 がプリインストールされています。将来的に、Solaris 11 Express (x86) がデータベース・サーバーのもう 1 つのオペレーティング・システムとして使用できるようになります。
- Exadata Storage Server に、2 ソケット、6 コアの Intel Xeon L5640 プロセッサ (2.26GHz で動作)、24GB のメモリ、384GB の Exadata Smart Flash Cache、512MB のバッテリーバックアップ式キャッシュを搭載したストレージ・コントローラに接続された 12 台の SAS ディスク、デュアル・ポートの InfiniBand 接続、組み込みの Oracle Integrated Lights Out Manager (Oracle ILOM)、ホットスワップ対応のデュアル冗長電源が事前構成されています。Exadata Storage Server では、600GB の高パフォーマンス SAS ディスクまたは 2TB の大容量 SAS ディスクのいずれかを使用できます。すべての Exadata Storage Server Software は、Exadata セル上にプリインストールされています。
- Sun Quad Data Rate (QDR) InfiniBand スイッチおよびケーブルを使用して、データベース・サーバーから Exadata Storage Server への通信と Oracle RAC ノード間の通信に必要な 40Gb/秒の InfiniBand ファブリックが構成されています。
- Database Machine をリモートから管理および監視するためのイーサネット・スイッチ。
- Database Machine をローカルで管理するためのキーボード、ビデオまたは画像表示装置、マウス (KVM) ハードウェア。
- これらのコンポーネントはすべてカスタムの 42U ラックにパッケージ化されており、システムの配電盤 (PDU) も含まれます。

それぞれのコンポーネント間の比率は、パフォーマンスを最大化し、可用性の高いシステムを実現し、すべてのデータベース・アプリケーションに最適な CPU 対 I/O 能力を提供するように設定されています。次の表に、Exadata Database Machine X2-2 のバージョンごとのハードウェア・コンポーネントを示します。

Database Machine X2-2 のコンポーネント

	Database Machine X2-2 フル・ラック	Database Machine X2-2 ハーフ・ラック	Database Machine X2-2 クォーター・ラック
データベース・サーバー	8	4	2
Exadata Storage Server	14	7	3
InfiniBand スイッチ	3	3	2

#### Exadata Database Machine X2-8

新たな最上位機種である Exadata Database Machine X2-8 には、大規模な SMP データベース・サーバーを含むグリッド・アーキテクチャを通じて、最高のスケールアップ・アーキテクチャとスケールアウト・アーキテクチャが組み合わせられています。以前は、64 コアの SMP にはそれ自体にフル・ラック機器が必要であり、それ以上のスケールアウトは困難でした。Exadata X2-8 は Sun による新しいインテルベースの超小型 64 コア・サーバーを 2 台使用して、パフォーマンスと可用性に優れたデータベース・グリッドを構築しています。各サーバーには 1TB のメモリ、インターコネク用 40Gb/秒の InfiniBand、データセンターへの接続向けの 10Gb/秒のイーサネットが搭載されています。X2-8 は、14 台の Exadata Storage Server を含む X2-2 と同じストレージ・グリッド・アーキテクチャを持つため、インテリジェントな問合せオフロード、10 倍のデータ圧縮、336TB の RAW ストレージ、そして 5.3TB の高パフォーマンス PCI フラッシュに対して 1 秒当たり最大 100 万の I/O が実現されます。Exadata X2-8 は、2368 の CPU コアと 2.6 ペタバイトの RAW ストレージを持つ 8 ラックのグリッドにまで容易に拡張できます。新しい Exadata X2-8 はあらゆるビジネス・アプリケーションに卓越したパフォーマンスをもたらすと同時に、大規模なデータベース統合を可能にします。

Exadata Database Machine X2-8 はフル・ラック構成で提供されており、Oracle Database 11g Release 2 を実行し、次のテクノロジーで構成されています。

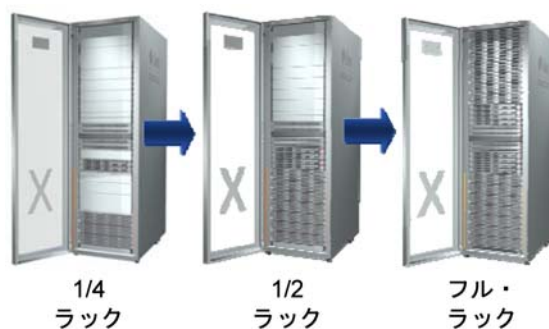
- 業界標準のデータベース・サーバー 2 台に、8 ソケット、8 コアの Intel® Xeon® X7560 プロセッサ (2.26GHz で動作)、1TB のメモリ、8 台の 300GB の 10,000RPM SAS ディスク、8 つの 40Gb/秒の InfiniBand ポート、8 つの 10Gb/秒のイーサネット・ポート、8 つの 1Gb/秒のイーサネット・ポート、ホットスワップ対応のデュアル冗長電源が事前構成されています。また、Database Machine のオペレーティング・システムとして、Unbreakable Enterprise Kernel を含んだ Oracle Linux 5 Update 5 がデータベース・サーバー上にプリインストールされています。将来的に、Solaris 11 Express (x86) がデータベース・サーバーのもう 1 つのオペレーティング・システムとして使用できるようになります。

- 14 台の Exadata Storage Server に、2 ソケット、6 コアの Intel Xeon L5640 プロセッサ (2.26GHz で動作)、24GB のメモリ、384GB の Exadata Smart Flash Cache、512MB のバッテリバックアップ式キャッシュを搭載したストレージ・コントローラに接続された 12 台の SAS ディスク、デュアル・ポートの InfiniBand 接続、組込みの Integrated Lights Out Manager、ホットスワップ対応のデュアル冗長電源が事前構成されています。すべての Exadata Storage Server Software は、Exadata セル上にプリインストールされています。
- 3 つの Sun Quad Data Rate (QDR) InfiniBand スイッチとケーブルを使用して、データベース・サーバーから Exadata Storage Server への通信と Oracle RAC ノード間の通信に必要な 40Gb/秒の InfiniBand ファブリックが構成されています。
- Database Machine をリモートから管理および監視するためのイーサネット・スイッチ。
- これらのコンポーネントはすべてカスタムの 42U ラックにパッケージ化されており、システムの配電盤 (PDU) も含まれます。

それぞれのコンポーネント間の比率は、パフォーマンスを最大化し、可用性の高いシステムを実現し、すべてのデータベース・アプリケーションに最適な CPU 対 I/O 能力を提供するように設定されています。

#### Database Machine のアップグレード性

Database Machine X2-2 の各バージョンは容量と能力を拡張できるため、処理要件が拡大した場合にもスムーズなアップグレード・パスが確保されています。オラクルのエンジニアによって、クォーター・ラックからハーフ・ラックへ、またハーフ・ラックからフル・ラックへのオンラインのフィールド・アップグレードが簡単に実行されます。



Database Machine X2-2 のアップグレード

Exadata Database Machine は極めて高性能なシステムでありながら、ほとんどすべてのサイズに拡張できるビルディング・ブロック方式を採用しています。InfiniBand ファブリックを使用し、Database Machine X2-2 の複数のフル・ラック・システムとハーフ・ラック・システムをシステム内で接続すると、大規模な単一システムのイメージ構成を形成できます。また同様に、Exadata Database Machine X2-8

の複数ラックを接続することもできます。この機能は、ラック間を InfiniBand ケーブルで接続することで実現されます。すべての InfiniBand インフラストラクチャ（スイッチやポート・ケーブル配線）は、この拡張オプションを提供するように設計されています。この Exadata Database Machine ならではの拡張機能によって、アプリケーションで必要となる最大規模のデータベースのサポートが可能になります。



接続されて単一システムを構成する 8 台の Exadata Database Machine X2-8

また、Exalogic Elastic Cloud も同じ InfiniBand ファブリックを使用して同様に Exadata Database Machine に接続されます。最大 8 つのフル・ラックの Exalogic システムと Exadata システムを、外部スイッチなしで接続できます。

### Exadata Storage Server

Exadata Storage Server では、オラクルが提供する Exadata Storage Server Software が実行されます。Exadata Storage Server のハードウェア・コンポーネント（別名 Exadata セル）は、高パフォーマンスのデータベース処理要件を満たすため綿密に選択されています。Exadata ソフトウェアは、ハードウェア・コンポーネントと Oracle Database を最大限に活用できるように最適化されています。各 Exadata セルによって、極めて高い I/O パフォーマンスと帯域幅がデータベースに提供されます。

Oracle Database の優れたセキュリティ機能を利用して構築された Exadata ストレージは、完全に暗号化されたデータベースに対する問合せを、オーバーヘッドをほとんど発生させることなく、1 秒当たり数百 GB という速度で実行します。これは、復号化処理をソフトウェアから Exadata Storage Server ハードウェアへ移行することで実現されます。この機能を可能にする拡張暗号化規格（AES）は、Exadata Storage Server で使用されている Intel 5600 プロセッサとこの Oracle ソフトウェアによってサポートされています。



Exadata Storage Server (Exadata セル)

### Exadata Smart Flash Cache

各 Exadata セルには、384GB の Exadata Smart Flash Cache が搭載されています。つまり、Database Machine X2-8 やフル・ラックの X2-2 には、5.3TB のフラッシュが搭載されていることとなります。これはほとんどすべてのデータベースを上回る数字です。このソリッド・ステート・ストレージは、Exadata ストレージのパフォーマンスを大幅に向上します。標準ディスク読取りの応答時間の 10 倍向上、標準ディスク読取りの IOPS の 100 倍向上に加え、メモリに代わる大容量を低コストで実現します。全体では、読取り/書き込み操作の平均パフォーマンスが 10 倍向上します。

Exadata Smart Flash Cache は、Exadata セル内の標準ディスクのアクティブ・データを管理しますが、その管理方法は単純な LRU (Least Recently Used) 方式ではありません。Exadata Storage Server Software と Oracle Database が連携してデータのアクセス・パターンを追跡し、データのキャッシュ内容やキャッシュ方法を把握して、キャッシュの汚染を回避します。この機能はすべて自動的に管理されるため、手動で調整する必要はありません。データベース・アプリケーションのパフォーマンスの鍵となる具体的な表や索引がある場合には、オプションでそれらを特定してキャッシュに固定できます。

### Exadata ストレージの容量、パフォーマンス、帯域幅、および IOPS

Oracle Exadata Storage Server には、600GB、15,000RPM の高パフォーマンス SAS ディスク、または 2TB、7,200RPM の大容量 SAS ディスクのいずれかが 12 台搭載されています。高パフォーマンスの SAS ディスクを使用した Exadata Storage Server は、最大 2TB の非圧縮ユーザー・データ容量と最大 1.8GB/秒の RAW データ帯域幅を実現します。大容量の SAS ディスクを使用した Exadata Storage Server は、最大 7TB の非圧縮ユーザー・データ容量と最大 1.0GB/秒の RAW データ帯域幅を実現します。圧縮形式でデータを保存すると、各セルで提供されるユーザー・データ容量とデータの帯域幅は大幅に増加します。

次の表に、Database Machine のモデルごとのストレージ容量を示します。

## Database Machine のストレージ容量

	Database Machine X2-8 および X2-2 フル・ラック	Database Machine X2-2 ハーフ・ラック	Database Machine X2-2 クォーター・ラック
<b>Exadata Smart Flash Cache</b>	5.3TB	2.6TB	1.1TB
<b>RAW ディスク容量</b>			
• 高パフォーマンス SAS	100TB	50TB	21TB
• 大容量 SAS	336TB	168TB	72TB
<b>ユーザー・データ容量</b>	最大	最大	最大
• 高パフォーマンス SAS	28TB	14TB	6TB
• 大容量 SAS (データ圧縮なし)	100TB	50TB	21TB

ユーザー・データ容量は、すべてのディスク領域をミラー化し、ディスク障害からリカバリするための領域を確保し、データベース構造用の領域（ログ、UNDO、データウェアハウジング向けの大規模な一時領域、索引など）を除外した後に、表の行を格納するために残された領域を見積もったものです。このユーザー・データ容量は圧縮されていない状態での数字を表しています。Exadata Hybrid Columnar Compression を使用した場合、容量は何倍も増加します。ただし、実際のユーザー・データはアプリケーションによって異なります。

Exadata Smart Flash Cache によって、各セルで非常に高いパフォーマンスを実現できます。Exadata Smart Flash Cache の自動キャッシュ機能により、各 Exadata セルで非圧縮データへのアクセス時に最大 3.6GB/秒の帯域幅と 75,000IOPS の提供が可能になります。データを圧縮形式で保存する場合には、ユーザー・データ容量、データの帯域幅、および IOPS は、多くの場合 10 倍またはそれ以上まで増加します。これは、Oracle Database と従来のストレージ・デバイスを使用する場合に比べ、大幅に増加することを示しています。

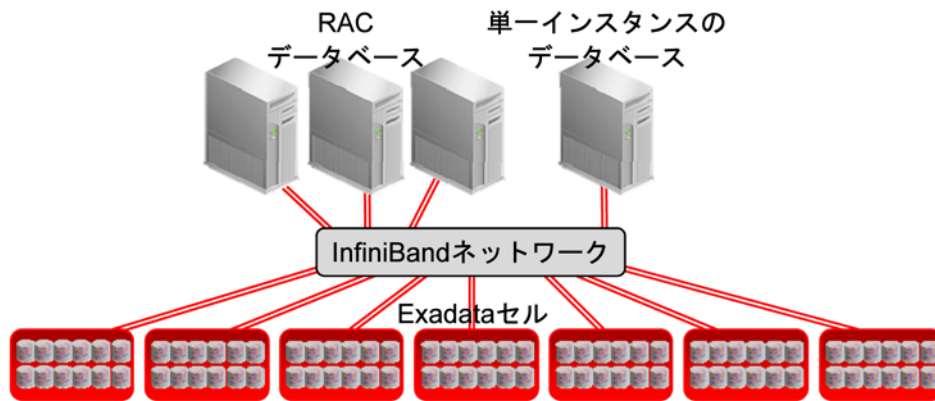
Database Machine の各モデルのパフォーマンス特性を次の表に示します。

## Database Machine の I/O パフォーマンス

	Database Machine X2-8 および X2-2 フル・ラック	Database Machine X2-2 ハーフ・ラック	Database Machine X2-2 クォーター・ラック
<b>RAW ディスク・データ帯域幅</b> • 高パフォーマンス SAS • 大容量 SAS (データ圧縮なし)	最大 25GB/秒 14GB/秒	最大 12.5GB/秒 7.0GB/秒	最大 5.4GB/秒 3.0GB/秒
フラッシュ・キャッシュ使用時の 実効データ帯域幅	最大 50GB/秒	最大 25GB/秒	最大 11GB/秒
フラッシュ・キャッシュの IOPS	最大 1,000,000	最大 500,000	最大 225,000
<b>ディスクの IOPS</b> • 高パフォーマンス SAS • 大容量 SAS	最大 50,000 25,000	最大 25,000 12,500	最大 10,800 5,400

## Exadata Database Machine のアーキテクチャ

次に、一般的なハーフ・ラック Database Machine 配置の概略図を示します。ハーフ・ラック内で、2つの Oracle Database と 1つの Real Application Clusters データベースが 3 台のデータベース・サーバーに配置され、単一インスタンスのデータベースが残りのデータベース・サーバーに配置されています（もちろん、4つのデータベース・サーバーすべてを 4 ノードの単一 RAC クラスタ向けに使用することもできます。）この場合、RAC データベースが本番データベースであり、単一インスタンスのデータベースがテストおよび開発データベースとなるでしょう。両方のデータベースはハーフ・ラックに含まれる 7つの Exadata セルを共有しますが、ソフトウェアの独立性を維持するため、それぞれ個別の Oracle ホームを使用します。この構成の全コンポーネント（データベース・サーバー、Exadata セル、InfiniBand スイッチ、およびその他のサポート・ハードウェア）は、Database Machine ラック内に搭載されています。



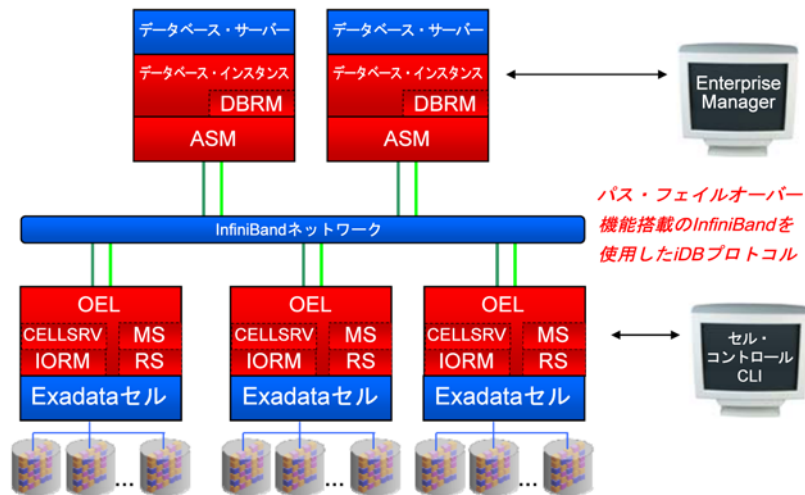
Database Machine のハーフ・ラック配置

Database Machine では、サーバーとストレージ間で最先端の InfiniBand インターコネクタが使用されます。各データベース・サーバーと Exadata セルには、高可用性を実現するデュアル・ポート QDR (Quad Data Rate) InfiniBand 接続が備わっています。各 InfiniBand リンクでは、従来のストレージやサーバー・ネットワークよりもはるかに高速の 40 ギガビットの帯域幅が提供されます。さらに、オラクルのインターコネクタ・プロトコルでは、直接データ配置 (DMA: ダイレクト・メモリ・アクセス) を使用して、余分なデータ・コピーを作成せずにデータを回線からデータベースのバッファに直接移動することで、CPU のオーバーヘッドを大幅に削減しています。InfiniBand ネットワークには、LAN ネットワークの柔軟性に加え、SAN の効率性が備わっています。オラクルでは、InfiniBand ネットワークを使用することによって、ネットワークがパフォーマンスのボトルネックにならないようにしています。この InfiniBand ネットワークでは、Oracle Database Real Application Clusters ノード用の高パフォーマンスのクラスタ・インターコネクタも提供されます。

Oracle Exadata は、どのようなパフォーマンス・レベルに対してもスケールアウトできるように設計されています。パフォーマンスとストレージ容量を増強する場合は、構成にデータベース・サーバーと Exadata セルを追加します（例：ハーフ・ラックからフル・ラックへのアップグレード）。

Exadata セルの追加数にほぼ比例して、容量と I/O パフォーマンスが増強されます。Exadata 構成内でセル間の通信は行われず、必要ありません。

Exadata ソリューションのアーキテクチャには、データベース・サーバーのコンポーネントと Exadata セル内のコンポーネントが含まれています。次の図に、クォーター・ラック構成のソフトウェア・アーキテクチャを示します。



Exadata ソフトウェアのアーキテクチャ

Exadata を使用すると、ほとんどの SQL 処理はデータベース・サーバーから Exadata サーバーにオフロードされます。Exadata では、従来のブロック処理サービスを提供しながら、データベース・インスタンスから基盤となるストレージに機能を移動できます。従来のストレージと比較した際の Exadata ストレージ独自の機能として、データベース問合せの対象となる表全体ではなく、問合せの条件を満たす行と列のみを返す機能があります。Exadata では、SQL 処理がデータ（またはディスク）のできるだけ近くにプッシュされるため、全ディスクでの並列処理が可能となっています。そのため、データベース・サーバーの CPU 消費量が減り、データベース・サーバーとストレージ・サーバー間のデータ移動時の帯域幅が大幅に削減されて、表全体ではなく問合せの結果セットが返されます。データ転送とデータベース・サーバーのワークロードを削減することによって、従来は帯域幅と CPU により制約を受けていたデータウェアハウジングの問合せに大きなメリットをもたらします。また、データ転送の削減は、大規模なバッチ処理とレポート処理操作を行うことが多いオンライン・トランザクション処理（OLTP）システムにも大きなメリットをもたらします。

Exadata は、データベースを使用するアプリケーションに対して完全に透過的に機能します。Database Machine では、従来のシステム上で実行されるものとまったく同じ Oracle Database 11g Release 2 が実行されますが、Database Machine でのほうが高速に実行されます。Exadata ストレージを使用する場合、既存の SQL 文は、非定型であってもパッケージ・アプリケーションやカスタム・アプリケーション

内であっても影響を受けることはなく、変更する必要はありません。このソリューションのオフロード処理と帯域幅の利点は、アプリケーションを一切変更することなく享受できます。さらに、Exadata では、Oracle Database の全機能が完全にサポートされています。Exadata は、Oracle Database の単一インスタンスでも Oracle Real Application Clusters 配置でも同等に機能します。Oracle Data Guard、Oracle Recovery Manager (Oracle RMAN)、Oracle GoldenGate、およびその他のデータベース・ツールなどの機能は、Exadata を使用しているかどうかに関係なく同様に管理できます。ユーザーとデータベース管理者は、Exadata ストレージ以外の従来のストレージを使用している場合と同様に作業できるため、使い慣れたツールや知識をそのまま利用できます。

従来のシステムと同じ Oracle Database とその機能が Database Machine に存在するため、Database Machine を管理する IT スタッフはこのソフトウェアに関して同様の知識を持っている必要があります。Database Machine を管理するには、Oracle Database の管理、バックアップとリカバリ、Oracle RAC と Oracle Enterprise Linux (OEL) に関する経験を持っていることが重要です。

### データベース・サーバー・ソフトウェア

Oracle Database 11g Release 2 は、Exadata ストレージの利点を活用できるように大幅に拡張されています。Exadata ソフトウェアは、データベース・サーバーと Exadata セルとの間で最適に分配されます。データベース・サーバーと Exadata Storage Server Software は、iDB (Intelligent Database プロトコル) を使用して通信を行います。iDB はデータベース・カーネルに実装され、データベース操作を Exadata 拡張操作に透過的にマッピングします。iDB では、データベースの従来のデータ・ブロック送信機能に加えて、送信アーキテクチャが実装されます。iDB は、Exadata セルへの SQL 操作の送信と、データベース・カーネルへの問合せ結果セットの返信に使用されます。Exadata セルでは、データ・ブロックではなく、SQL 問合せの条件を満たす行と列のみが返されます。既存の I/O プロトコルと同様に、iDB ではバイト範囲の読取り/書込みをディスク間で直接行うことができるため、オフロード処理を実行できない場合も、Exadata は Oracle Database に対して従来のストレージ・デバイスと同じように動作します。オフロード処理を実行できる場合は、データベース・カーネルのインテリジェンスによって、たとえば表スキャンが Exadata Storage Server で実行され、要求されたデータのみがデータベース・サーバーに返されます。

iDB は、業界標準の Reliable Datagram Sockets (RDSv3) プロトコルに基づいて構築されており、InfiniBand 上で動作します。RDS のゼロコピー実装である ZDP (Zero-loss Zero-copy Datagram Protocol) は、ブロックの不要なコピーを削減するために使用されます。データベース・サーバーと Exadata セルでは、複数のネットワーク・インタフェースを使用できます。このプロトコルは非常に高速で待機時間が短いため、I/O 操作に必要なデータ・コピー数を最小限に抑えることができます。

Oracle Automatic Storage Management (Oracle ASM) は、Exadata のファイル・システムとボリューム・マネージャとして使用されます。Oracle ASM は、ストレージ・リソースを仮想化し、Exadata の高度なボリューム管理機能とファイル・システム機能を提供します。使用可能な Exadata セルやディスク間でデータベース・ファイルが均等にストライプ化されるため、全ストレージ・ハードウェア間で I/O の負荷が均一化されます。リソースの割当て/再割当てをスムーズに実行する Oracle ASM の機能は、Exadata 環境の共有グリッド・ストレージ機能の鍵となる機能です。Oracle ASM のディスクのミラー化機能と、ホットスワップ対応の Exadata ディスクによって、個々のディスク・ドライバに障害

が発生した場合にもデータベースの動作が保証されます。データは複数のセルでミラー化されるため、1つのセルに障害が発生しても、データが失われたり、データ・アクセスが妨げられたりすることはありません。こうした超並列アーキテクチャによって、制限のないスケラビリティと高可用性が実現します。

Oracle Database 11g の Oracle Database Resource Manager (Oracle DBRM) の機能は、Exadata で使用できるように拡張されています。Oracle DBRM を使用すれば、ユーザーは、CPU、UNDO、並列度、アクティブ・セッション、およびその他の管理リソースに加え、データベース内とデータベース間の I/O 帯域幅を定義して管理できます。そのため、複数のデータベース間でストレージを共有している場合、1つのデータベースが I/O 帯域幅を独占したり、ストレージを共有しているその他のデータベースのパフォーマンスに影響を与えたりするおそれがありません。コンシューマ・グループに使用可能な I/O 帯域幅 (%) が割り当てられ、Oracle DBRM でこの割当て目標が維持されます。この機能は、関連するデータベースやコンシューマ・グループに I/O をタグ付けするデータベースによって実装されます。これにより、I/O スタック全体の I/O の優先順位を確認できる完全なビューがデータベースに提供されます。データベース内のコンシューマ・グループへの I/O の割当ては、データベース・サーバーで定義および管理されます。データベース間の I/O の割当ては、Exadata セルのソフトウェアで定義され、I/O Resource Manager (IORM) で管理されます。この Exadata セルのソフトウェアは、データベース間の I/O リソースを管理し、データベース内とデータベース間での適切な割当てを行います。全体的には、Oracle DBRM によって各データベースに指定量の I/O リソースの割当てが確保され、ユーザー定義の SLA が達成されます。

Exadata Database Machine 上でのみ提供される Oracle Database の新機能は、Oracle Database Quality of Service (QoS) Management と QoS Management Memory Guard の2つです。QoS Management を使用すると、システム管理者は Oracle Exadata Database Machine 上でホストされているアプリケーションのサービス・レベルを直接管理できます。ポリシー・ベースのアーキテクチャを使用して、QoS Management は正確な実行時パフォーマンスとリソース・メトリックを関連付け、専門システムによるデータ分析を通じてボトルネックを特定し、動的な負荷条件の下でパフォーマンス目標を達成および維持するために推奨されるリソース調整案を生成します。十分なリソースが使用可能でない場合、QoS は重大性の低い目標を犠牲にして、よりビジネス・クリティカルな目標を遵守します。QoS Management Memory Guard は Oracle Cluster Health Monitor と連携して、メモリの過剰消費による障害のリスクのあるノードを検出します。新規接続を自動的に阻止することで既存のワークロードを保護し、十分なメモリが使用可能になり次第接続を復旧します。

#### Oracle Enterprise Manager の Exadata 用プラグイン

Exadata は Oracle Enterprise Manager Grid Control と統合されているため、Exadata 環境を簡単に監視できます。Exadata プラグインを既存の Oracle Enterprise Manager システムにインストールすれば、Exadata サーバーの統計やアクティビティを監視して、システム管理者にイベントやアラートを送信できます。Oracle Enterprise Manager システムを Exadata と統合する利点は、次のとおりです。

- Oracle Exadata ストレージの監視

- ストレージ構成とパフォーマンスに関する情報の収集
- しきい値に基づいたアラートと警告の発行
- 履歴データに基づいた豊富な標準メトリックとレポートの提供

ユーザーの求める Oracle Enterprise Manager の機能はすべて、Exadata と連動して機能します。Oracle Enterprise Manager のインタフェースを使用することにより、ユーザーは、従来 Oracle Enterprise Manager とともに使用してきたその他の Oracle Database 環境と同様に、Exadata 環境を簡単に管理できます。DBA は、使い慣れた Oracle Enterprise Manager のインタフェースでレポートを表示して、Exadata システムの状態を確認したり、Exadata ストレージの構成を管理したりすることができます。

### Exadata Storage Server Software

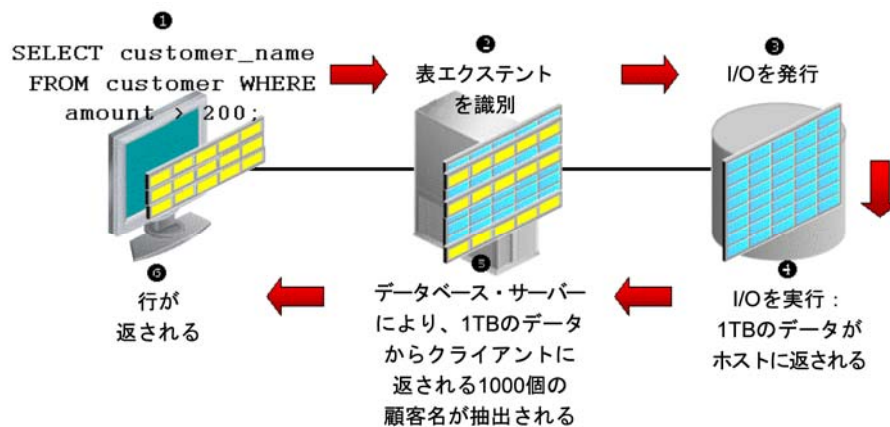
あらゆるストレージ・デバイスと同じく、Exadata Storage Server は、CPU、メモリ、バス、ディスク、NIC に加え、サーバーに通常搭載されているその他のコンポーネントで構成されたコンピュータです。また、オペレーティング・システム (OS) も実装されており、Exadata の場合は Oracle Enterprise Linux 5.5 が実装されています。Exadata セル内の Exadata Storage Server Software は、Oracle Enterprise Linux 上で実行されます。管理者は、Oracle Enterprise Linux に制限モードでアクセスして、Exadata セルを管理します。

CELLSRV (セル・サービス) は、セルで実行される Exadata ソフトウェアの主要コンポーネントで、Exadata ストレージ・サービスの大部分を提供します。CELLSRV は、データベース・サーバー上のデータベース・インスタンスと通信するマルチスレッドのソフトウェアで、iDB プロトコルに基づいてデータベースにブロックを提供します。このソフトウェアによって高度な SQL オフロード機能が提供され、SQL オフロード処理を実行できない場合は、Oracle ブロックが提供されます。また、I/O を発行する各種データベースやコンシューマ・グループに I/O 帯域幅を割り当てる Oracle DBRM の I/O リソース管理機能が実装されます。

その他にセルで実行される Oracle ソフトウェアのコンポーネントには、Management Server (MS) と Restart Server (RS) の 2 つがあります。MS は管理者向けのプライマリ・インタフェースで、Exadata セルのステータスの管理と問合せに使用されます。MS は Exadata セルのコマンドライン・インタフェース (CLI) と Oracle Enterprise Manager の Exadata プラグインと連携して機能するため、Exadata セル単体での管理と構成が可能になります。たとえば、セルからは、ストレージの構成、I/O 統計の問合せ、セルの再起動を実行する CLI コマンドが発行されます。また、分散型 CLI が提供されるため、複数のセルにコマンドを送信して、セル全体を簡単に管理できます。Restart Server (RS) は、Exadata ソフトウェアとサービスの継続的な機能を維持します。RS は、Exadata ソフトウェアの更新に使用されます。また、ストレージ・サービスの起動や実行に加え、必要に応じてサービスの再起動を行います。

### Exadata の Smart Scan 処理

iDB 認識型ではない従来のストレージでは、データベース・インテリジェンスはすべてサーバー上のデータベース・ソフトウェア内にあります。次の図に示す表スキャンの例は、このアーキテクチャで SQL 処理がどのように実行されるかを説明しています。



従来のデータベースでの I/O および SQL 処理モデル

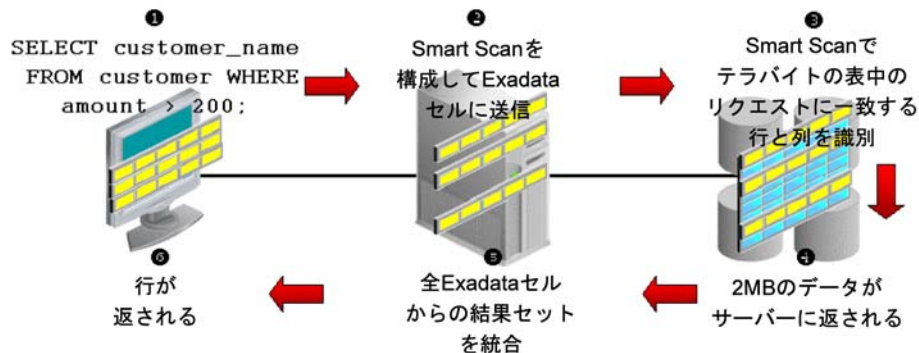
①クライアントで、対象行のみをフィルタリングして返すという条件を記述した `SELECT` 文を発行します。②データベース・カーネルで、スキャン対象の表が含まれるファイルとエクステントに、このリクエストをマッピングします。③データベース・カーネルで、ブロックを読み込む I/O を発行します。④問合せ対象の表の全ブロックが、メモリに読み込まれます。⑤SQL 処理が RAW ブロックに対して行われ、条件を満たす行が検索されます。⑥行がクライアントに返されます。

大規模な問合せの場合にはよくあることですが、読み込まれた行のほとんどは条件に一致しないものとして除外されます。それでも、表の全ブロックを読み込み、ストレージ・ネットワーク経由で転送し、メモリにコピーする必要があります。要求された SQL 操作の実行に必要な行よりもはるかに多くの行が、メモリに読み込まれます。そのため、大量のデータ転送が発生し、帯域幅が消費されるため、アプリケーションのスループットと応答時間に影響が出ます。

データベース機能をデータベース・スタックのストレージ・レイヤー内に統合すれば、問合せを始めとしたデータベース操作をはるかに効率的に実行できます。Exadata の場合は、ディスク・レベルでデータベース機能をハードウェアのできるだけ近くに実装するため、データベース操作の速度とシステムのスループットが飛躍的に向上します。

Exadata ストレージを使用すれば、データベース操作がよりいっそう効果的に処理されます。表スキャンを実行する問合せは、データベース・サーバーに返す必要のあるデータのサブセットのみを使用して、Exadata ストレージ内で処理できます。行フィルタリングや列フィルタリングに加え、（その他の機能との）一部の結合処理は、Exadata ストレージ・セル内で実行されます。そのため、関係のある必要なデータのみがデータベース・サーバーに返されます。

次の図は、表スキャンが Exadata ストレージ内でどのように処理されるかを示しています。



Smart Scan Offload Processing

①クライアントで、対象行のみをフィルタリングして返すという条件を記述した `SELECT` 文を発行します。②データベース・カーネルで、Exadata ストレージが利用できることを確認し、発行された SQL コマンドを表す iDB コマンドを作成して、Exadata ストレージに送信します。③Exadata ソフトウェアの CELLSRV コンポーネントでデータ・ブロックをスキャンして、発行された SQL コマンドの条件を満たす行と列を特定します。④条件を満たす行と要求された列のみが、メモリに読み込まれます。⑤データベース・カーネルで、複数の Exadata セルからの結果セットを 1 つに統合します。⑥行がクライアントに返されます。

Smart Scan はアプリケーションに対して透過的に動作するため、アプリケーションや SQL を変更する必要はありません。SQL `EXPLAIN PLAN` で、Exadata の Smart Scan がいつ使用されるかが示されます。返されるデータには完全な一貫性とトランザクション性があり、Oracle Database の読取り一貫性機能と動作に厳密に対応しています。Smart Scan の実行中にセルに障害が発生した場合、Smart Scan の未完了部分は別のセルに透過的にルーティングされて実行されます。Smart Scan は、Oracle Database の複雑な内部メカニズムに適切に対応しています。この内部メカニズムには、未コミット・データやロック行、連鎖行、圧縮表、言語処理、日付計算、正規表現検索、マテリアライズド・ビュー、パーティション化された表などがあります。

Oracle Database と Exadata サーバーでは、さまざまな SQL 文が連携的に実行されます。SQL 処理をデータベース・サーバーから移行することにより、サーバーの CPU サイクルが解放され、帯域幅の消費量が大幅に減少するため、その他の要求により良いサービスを提供できます。また I/O 帯域幅の競合が緩和されるため、SQL 操作が高速化され、より多くの SQL 操作を同時実行できます。それではこれから、Exadata の使用が有効であるさまざまな SQL 操作について見ていきましょう。

### Smart Scan の条件フィルタリング

Exadata では、表スキャンで条件フィルタリングを実行できます。データベース・サーバーには、表のすべての行が返されるのではなく、要求された行のみが返されます。たとえば、以下の SQL 文が発行されると、従業員の雇用日が指定された日付よりも後の日付の行のみが、Exadata からデータベース・インスタンスに送信されます。

```
SELECT * FROM employee_table WHERE hire_date > '1-Jan-2003';
```

この機能では、関連する行のみがサーバーに返されるため、データベースのパフォーマンスが大幅に向上します。こうしたパフォーマンスの向上は、より複雑な問合せにも当てはまるため、副問合せを含む複雑な問合せでも同様の利点を享受できます。

#### Smart Scan の列フィルタリング

さらに、Exadata では、表スキャンで列フィルタリング（別名：列射影）を実行できます。データベース・サーバーには、表のすべての列が返されるのではなく、要求された列のみが返されます。たとえば、以下の SQL 文が発行されると、employee\_name 列と employee\_number 列のみが、Exadata からデータベース・カーネルに返されます。

```
SELECT employee_name, employee_number FROM employee_table;
```

多数の列が含まれる表や、LOB（ラージ・オブジェクト）が含まれる列では、I/O 帯域幅を大幅に削減できます。条件フィルタリングと列フィルタリングを一緒に使用すれば、パフォーマンスが飛躍的に向上し、I/O 帯域幅の消費量を大幅に削減できます。さらに、列フィルタリングを索引で実行すると、問合せのパフォーマンスをよりいっそう向上できます。

#### Smart Scan の結合処理

Exadata では、大型の表と小型のルックアップ表との結合処理を実行します。これは、スター・スキーマ採用のデータウェアハウスではごく一般的なシナリオです。結合処理は、Bloom フィルタを使用して実装されます。このフィルタは、非常に効率的な確立論的手法に基づいて、行が目的の結果セットに含まれるかどうかを判定します。

#### 暗号化された表領域と列の Smart Scan 処理

表領域暗号化（TSE）と列暗号化（TDE）に対する Smart Scan のオフロード処理は、Exadata ストレージ内でサポートされています。これにより、企業の機密データにアクセスする際のパフォーマンスが向上します。

#### ストレージ索引

ストレージ索引は、I/O 操作を回避する Exadata ストレージの非常に強力な機能です。Exadata Storage Server Software によって、Exadata セル内にストレージ索引（データベース・オブジェクトに関するメタデータ）が作成および保存されます。ストレージ索引では、そのセルに保存されている表の列の最小値と最大値が追跡されます。I/O が実行される前に問合せで WHERE 句が指定されると、Exadata ソフトウェアによって列の値とストレージ索引内の最小値と最大値とが比較され、セル内に指定された列の値を持つ行が存在するかどうかの調査が行われます。列の値が最小/最大範囲外の場合は、問合せのスキャン I/O が回避されます。多数の I/O 操作がいくつかの検索に自動的に置換されるため、SQL 操作の多くで実行速度が大幅に向上します。操作におけるオーバーヘッドを最小限に抑えるため、Exadata Storage Server Software ではストレージ索引が透過的かつ自動的に作成され、保存されます。

### データ・マイニング・モデルのスコアリングのオフロード

データ・マイニング・モデルのスコアリング機能は、Exadata にオフロードされています。この機能によって、Database Machine へのデータウェアハウス配置がよりさらに適切なものとなり、データ分析プラットフォームとしてのパフォーマンスが向上します。データ・マイニング・スコアリング機能 (prediction\_probability など) はすべて、Exadata にオフロードされて処理されます。そのため、ウェアハウス分析が高速化されるだけでなく、データベース・サーバーの CPU 消費量に加え、データベース・サーバーと Exadata ストレージ間の I/O の負荷が削減されます。

### Exadata のその他の Smart Scan 処理

Exadata にオフロードされるその他のデータベース操作としては、データベースの増分バックアップと表領域の作成の 2 つがあります。データベースの増分バックアップの速度と効率は、Exadata によって大幅に向上しています。データベース内の変更追跡の粒度は、Exadata ストレージを使用するとよりいっそう細かくなります。Exadata では、複数のブロックからなる大規模グループ単位ではなく、個々の Oracle ブロック単位で変更が追跡されます。その結果、バックアップで消費される I/O 帯域幅が減り、バックアップの実行速度が向上します。

Exadata では、ファイル作成操作もはるかに効率的に実行されます。たとえば、Create Tablespace コマンドを発行する場合、新しい表領域の各ブロックをサーバー・メモリでフォーマットしてストレージに書き込む操作が同期実行されるのではなく、表領域の作成とブロックのフォーマットを指示する iDB コマンドが Exadata に送信されます。そのため、ホストのメモリ使用量が減り、表領域ブロックの作成とフォーマットに関連付けられている I/O がオフロードされます。これらの操作で使用される I/O 帯域幅が削減されることで、ビジネスに不可欠なその他の作業に使用できる帯域幅が増加します。

### Exadata Hybrid Columnar Compression

データを圧縮すると、大規模データベースで消費されるストレージが大幅に削減されます。Exadata は Exadata Hybrid Columnar Compression (EHCC) と呼ばれる非常に高度な圧縮機能を提供しています。Exadata Hybrid Columnar Compression は、最高水準のデータ圧縮を可能にすることで、I/O を削減し、飛躍的なコスト削減とパフォーマンス向上を実現します。平均的なストレージ削減は、EHCC の使用法によって 10~15 倍の範囲で変動します。平均の削減値を 10 倍とした場合、新たにストレージを購入する必要性が大幅に低下し、多くの場合、今後数年間の追加購入は必要なくなります。たとえば、100TB のデータベースで 10 倍のストレージ削減が達成された場合、物理的に使用されるストレージはわずか 10TB となります。これにより、90TB のストレージが新たに利用可能になるため、長期にわたってストレージ購入を先延ばしできます。

EHCC は、データベース・ブロック内にデータを編成するための新しい手法です。名前から想像できるとおり、このテクノロジーでは、行を使用した手法と列を使用した手法を組み合わせてデータを格納します。両方の長所を併せ持ったこの混成アプローチは、列形式のストレージによる圧縮メリットを実現しながら、同時に、純粋な列形式によるパフォーマンス低下を回避します。Exadata Hybrid Columnar Compression では、圧縮単位と呼ばれる論理的な構成体を使用して、一連の行が格納されます。

データがロードされると、行セットから列値が切り離され、順序付けとグループ化が行われてから圧縮されます。行セットの列データは圧縮されてから圧縮単位に格納されます。

EHCC データの Smart Scan 処理が行われ、列フィルタリング (列射影) が Exadata 内で実行されます。Exadata Hybrid Columnar Compression で圧縮されたデータに対しては、問合せを直接実行できるため、データの圧縮を解除する必要はありません。問合せ条件を満たすために必要なデータの圧縮解除は必要なく、クライアントに返される列と行のみがメモリ内で圧縮解除されます。パフォーマンスを最大化し、データベース・サーバーの負荷を軽減するため、この圧縮解除処理は Exadata セル上で実行されます。通常、Hybrid Columnar Compressed Table ではデータが 10 倍に圧縮されるため、非圧縮データに比べて I/O レートが 10 倍に向上します。

### Exadata による I/O リソース管理

従来のストレージでは、ストレージ・サブシステムから I/O 帯域幅を消費するさまざまなジョブやユーザーの作業の優先順位付けを行う機能がないため、共有ストレージ・グリッドの作成が妨げられます。これは、複数のデータベースでストレージ・サブシステムを共有する場合にも同様に当てはまります。Oracle DBRM と Exadata ストレージの I/O リソース管理機能を使用すれば、1 つの作業種類やデータベースによるディスク・リソースや帯域幅の独占を防止できるため、Exadata ストレージ使用時にユーザー定義の SLA を達成できます。Oracle DBRM では、データベース間や、さまざまなユーザーと作業種類との間で消費される I/O 帯域幅の調整と優先順位付けを行うことができます。Exadata では、データベースをストレージ環境と緊密に統合することによって、作業種類や I/O 帯域幅の消費量が認識されます。そのため、ユーザーは、Exadata システムでさまざまなワークロードを識別してワークロードの優先順位付けを行い、もっとも重要なワークロードを最優先で処理できます。

データウェアハウジングや複合ワークロード環境では、データベース内でさまざまなユーザーとタスクに正しい割合で I/O リソースを割り当てる必要があります。たとえば、I/O リソースの 70% をシステム上のインタラクティブ・ユーザーに割り当て、30% をバッチ・レポート・ジョブに割り当てる必要があるとします。これは、Oracle DBRM と Exadata ストレージの I/O リソース管理機能を使用して簡単に実行できます。

Exadata 管理者は、I/O リクエストの優先順位付け方法を指定するリソース・プランを作成できます。これは、各種作業をコンシューマ・グループと呼ばれるサービス・グループに組み入れることで実行されます。コンシューマ・グループは、ユーザー名、クライアント・プログラム名、関数、問合せ実行時間などのさまざまな属性によって定義できます。コンシューマ・グループを定義したら、ユーザーは、I/O リソースを取得するコンシューマ・グループの優先順位と、各コンシューマ・グループに割り当てられる I/O リソースの量を決定する階層を設定できます。I/O リソースの優先順位付けを行うこの階層は、データベース内操作 (データベース内で発生する操作) とデータベース間操作 (データベース間で発生する操作) の両方に同時に適用できます。

Exadata ストレージが複数のデータベース間で共有されている場合でも、各データベースに割り当てられた I/O リソースの優先順位付けを行うことができるため、1 つのデータベースにディスク・リソース

や帯域幅が独占されるのを防止して、ユーザー定義の SLA を達成できます。たとえば、2 つのデータベースで Exadata ストレージを共有するとします。ビジネス目標は、これらの各データベースが組織にとって適切な価値と重要性を持つことです。データベース A には全 I/O リソースの 33% を割り当て、データベース B には 67% を割り当てることが決定されたとします。各データベース内のさまざまなユーザーとタスクに正しい割合で I/O リソースを割り当てるために、次のようなさまざまなコンシューマ・グループが定義されます。

- データベース A に対しては、次の 2 つのコンシューマ・グループが定義されます。
  - I/O リソースの 60% をインタラクティブなマーケティング活動用に確保する
  - I/O リソースの 40% をバッチ処理のマーケティング活動用に確保する
- データベース B に対しては、次の 3 つのコンシューマ・グループが定義されます。
  - I/O リソースの 60% をインタラクティブな販売活動用に確保する
  - I/O リソースの 30% をバッチ処理の販売活動用に確保する
  - I/O リソースの 10% を大口顧客への販売活動用に確保する

これらのコンシューマ・グループに I/O リソースを割り当てる割合は、各データベースに割り当てられる全 I/O リソースに対する割合となります。

つまり、Exadata の I/O Resource Manager によって従来のストレージ・テクノロジーでは解決できなかった問題の 1 つが解決され、ストレージ・サブシステムを共有する複数のデータベースとユーザーの作業の分散と優先順位付けを行う機能を備えた、共有グリッド・ストレージ環境の構築が可能になりました。Exadata の I/O リソース管理によって、Exadata ストレージを共有する複数のデータベースで、ユーザー定義の SLA を達成できます。そのため、各データベースやユーザーに共有するディスク帯域幅が正しく割り当てられ、ビジネス目標が達成されます。

## Exadata を使用したサービス品質 (QoS) 管理

Oracle Exadata QoS Management は、システム全体に対するワークロード・リクエストを監視する、ポリシー・ベースの自動化製品です。この製品はアプリケーション間で共有されているリソースを管理し、ビジネスに必要なパフォーマンス・レベルでアプリケーションが実行されるように、システム構成を調整します。またシステムの構成や要件の変化にも柔軟に対応するため、アプリケーションのパフォーマンス・レベルの変動を防止できます。

Oracle Exadata QoS Management はターゲット・システム上の各作業リクエストのパフォーマンスを監視します。作業リクエストがデータベース・サービスを使用してデータベースへの接続をリクエストした時点から、追跡が開始されます。作業リクエストの完了にかかる時間または応答時間 (エンド・ツー・エンドの応答時間もしくはラウンドトリップ時間とも言われる) は、データ・リクエストが開始されてからデータ・リクエストが完了するまでの時間を指します。応答時間における 2 つの要素 (リソース使用にかかる時間とリソース使用の待機にかかる時間) を正確に測定することで、

QoS Management はシステム内のボトルネックを素早く特定します。さらにボトルネックを解消するためのリソース再割当てを推奨することで、サービス・レベルを維持または回復します。再割当てが必要になるとシステム管理者にアラートが通知され、QoS Management ダッシュボード上でボタンをクリックするだけで再割当てが実行されます。このアクションによって予測されるクラスタ全体のパフォーマンス影響についても、併せて詳しく表示されます。また、すべてのアクションとポリシー変更に対する監査ログとシステム・パフォーマンスの履歴グラフが維持されます。

Oracle Exadata QoS Management はシステム上のリソース管理を通じて、次の機能を提供します。

- 要件を満たすために必要なリソースが十分に使用できる場合、ワークロードが変化してもビジネスレベルのアプリケーション・パフォーマンス要件は遵守されます。
- 要件を満たすためのリソースを十分使用できない場合、Oracle Exadata QoS Management は重大性の低いパフォーマンス要件を犠牲にして、よりビジネス・クリティカルなパフォーマンス要件を遵守します。
- ロード条件が大幅に容量を上回る場合も、リソースは引き続き利用できます。

#### Oracle Exadata QoS Management による利点

アプリケーションの応答時間が容認可能な範囲を超えている場合、一般的な企業では問題解決に非常に長い時間がかかる場合があります。多くの場合、管理者が最初に抱くのは「システム設定は正しいか。パラメータの変更で問題を解決できるか。ハードウェアの追加が必要か」、といった疑問です。残念ながら、これらの疑問に正確に回答することは非常に困難であり、多くの場合は非生産的でストレスのたまる作業に何時間も費やす結果に終わります。

Oracle Exadata QoS Management によってもたらされる利点は、次のとおりです。

- Oracle Real Application Clusters リソースを管理するシステム管理者に必要な時間と専門知識を軽減できます。
- パフォーマンス障害の件数が削減されます。
- アプリケーション・パフォーマンスを制限または低下させる問題の解決にかかる時間を短縮できます。
- ワークロードが変化してもシステムの安定性が維持されます。
- アプリケーションに対して透過的にサーバーを追加または削除できます。
- サーバー障害によるシステムへの影響を軽減できます。
- 品質保証契約 (SLA) の遵守が促進されます。
- ハードウェア・リソースをより効果的に共有できます。
- メモリの過剰消費によるサーバー障害から既存のワークロードが保護されます。
- Exadata ストレージの仮想化

- Exadata には、Oracle Database、Exadata ソフトウェア、Exadata ハードウェアの長所を活かす、高度で強力なストレージ管理仮想化機能が豊富に備わっています。

#### Exadata ストレージ・ソフトウェア

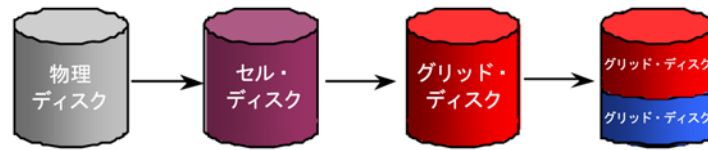
すでに説明したように、Exadata セルは、Oracle Enterprise Linux とオラクルの Exadata ソフトウェアを実行するサーバーです。最初に、セルをその他のコンピュータと同様に Exadata ストレージ処理モードで起動します。最初の 2 台のディスク・ドライブにはシステム領域と呼ばれる約 13GB の小さい LUN (Logical Unit Number) スライスがあり、Oracle Enterprise Linux オペレーティング・システム、Exadata ソフトウェア、および構成メタデータ用に確保されています。このシステム領域には、Oracle Database 11g の Automatic Diagnostic Repository (ADR) データに加え、Exadata セルに関するその他のメタデータが保存されます。システム領域 LUN は自動的に作成されるため、管理者による管理は不要です。LUN のコンテンツは、ドライブの障害から保護するために複数の物理ディスクに自動的にミラー化され、ディスクのホットスワップが可能となります。2 台のディスク・ドライブにあるその他の部分は、ユーザー・データ用に使用できます。

#### Exadata ユーザー・ストレージの仮想化

Oracle Automatic Storage Management は、Exadata セル内のストレージの管理に使用されます。Oracle ASM のボリューム管理、ストライプ化、データ保護サービスによって、ボリューム管理に最適な選択が行われます。Oracle ASM によって、ドライブやセルの障害に対するデータ保護、最高のパフォーマンス、極めて柔軟性の高い構成/再構成オプションが提供されます。

セル・ディスクは、物理ディスクからシステム領域 LUN (存在する場合) を差し引いた部分を仮想表示したもので、管理者が Exadata セル内で管理する主要ディスク・オブジェクトの 1 つです。セル・ディスクは単一の LUN で表示され、物理ディスク検出時に Exadata ソフトウェアによって自動的に作成され、管理されます。

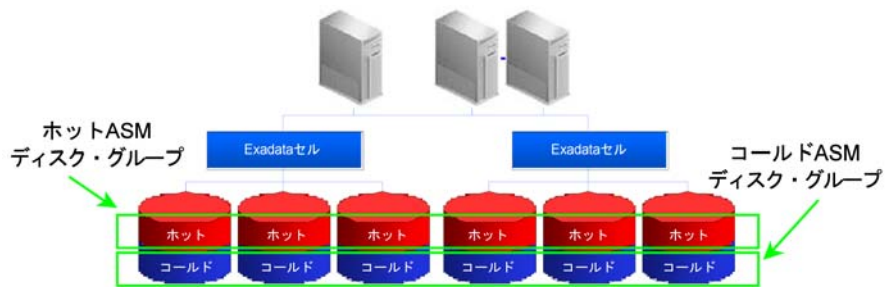
セル・ディスクをさらに仮想化して、1 つまたは複数のグリッド・ディスクにすることができます。グリッド・ディスクは Oracle ASM に ASM ディスクとして割り当てられるディスク・エンティティで、データベースの代わりにユーザー・データを管理します。もっとも簡単な例は、セル・ディスク全体を 1 つのグリッド・ディスクにすることです。ただし、セル・ディスクを分割して複数のグリッド・ディスク・スライスにすることもできます。セル・ディスク上に複数のグリッド・ディスクを配置すると、管理者は、パフォーマンスや可用性の要件が異なる複数のプールにストレージを分離できます。グリッド・ディスク・スライスは、セル・ディスクへの"ホット"、"ウォーム"、"コールド"割当てや、Exadata ディスクを共有しているデータベースの分離に使用できます。たとえば、セル・ディスクの分割は次のようにして行います。1 つのグリッド・ディスクを物理ディスクの高パフォーマンス部分に配置して三重にミラー化されるように設定し、もう 1 つのグリッド・ディスクを物理ディスクの低パフォーマンス部分に配置して、ミラー化は設定せずに、データのアーカイブやバックアップに使用します。グリッド・ディスク機能を使用して、情報ライフ・サイクル管理 (ILM) 戦略を実装することもできます。



グリッド・ディスクの仮想化

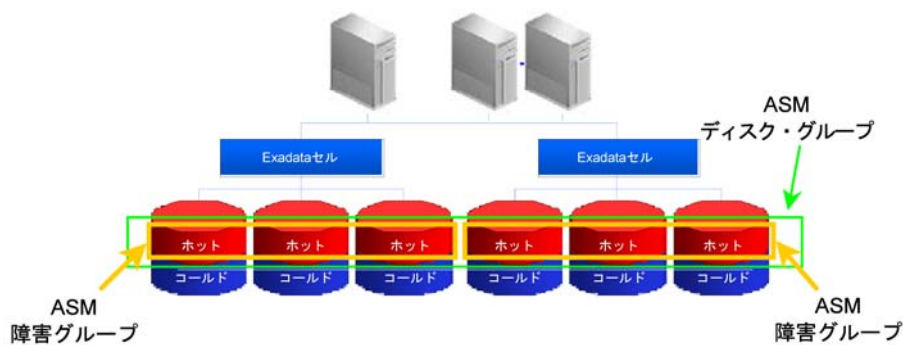
次の例では、より包括的な Exadata ストレージ・グリッドでのセル・ディスクとグリッド・ディスクの関係について説明します。

セル・ディスクとグリッド・ディスクが構成されると、Exadata 構成全体で ASM ディスク・グループが定義されます。2つの ASM ディスク・グループが定義され、1つは"ホット"グリッド・ディスクに、もう1つは"コールド"グリッド・ディスクに対応します。"ホット"グリッド・ディスクはすべて1つの ASM ディスク・グループに配置され、"コールド"グリッド・ディスクはすべて別の ASM ディスク・グループに配置されます。データがデータベースにロードされると、Oracle ASM によってデータと I/O が複数のディスク・グループに均等に分配されます。これらのディスク・グループで Oracle ASM のミラー化機能を有効にすれば、ディスク・グループにおけるディスク障害に備えてデータを保護できます。ミラー化機能は、各ディスク・グループで個別に有効/無効に設定できます。



ASM ディスク・グループとミラー化の例

さらに、Exadata セル全体の障害からデータを保護するため、ASM 障害グループが定義されます。障害グループによって、ミラー化された ASM エクステンツが別の Exadata セルに配置されます。



Oracle ASM によるミラー化と ASM 障害グループの例

Exadata で Oracle ASM を使用すると、以下を実現できます。

- セル・ディスクの構成 (LUN 作成) が、Exadata ソフトウェアによって自動的に実行されます。
- オプションとして、複数のグリッド・ディスクを物理ディスク上に共存させ、データベース・アプリケーションの要件に合わせてパフォーマンスを調整したり、Exadata で ILM 戦略を構築したりすることができます。
- Oracle ASM によって、Exadata ディスクとセル全体のデータベース・データが自動的にストライプ化されるため、I/O のロードバランシングとパフォーマンスの最適化が実現します。
- Oracle ASM の動的な追加/削除機能によって、セルやディスクの割当て、割当て解除、再割当てをスムーズに行うことができます。
- Oracle ASM のミラー化機能と Exadata セルのホットスワップ機能によって、ディスク障害が発生してもデータが透過的に保護され、アクセスが保証されます。
- Oracle ASM では二重または三重のミラー化が可能のため、データの重要度に合わせて保護レベルを調整できます。
- ASM 障害グループが Exadata で自動的に作成されるため、セルに障害が発生してもデータが透過的に保護され、アクセスが保証されます。

#### Exadata ストレージへの移行

Database Machine へデータを移行するには、いくつかの手法があります。Oracle Recovery Manager を使用して、従来のストレージからデータのバックアップを作成し、Exadata にリストアできます。また、Oracle Data Guard を使用すれば、移行が迅速化されます。そのためには、まず、Exadata ストレージをベースにしたスタンバイ・データベースを作成します。スタンバイ・データベースは Exadata ストレージを使用でき、本番データベースは従来のストレージを使用できます。高速スイッチオーバーを実行すれば、わずか数秒で、スタンバイ・データベースを本番データベースに変換できます。この場合組込みのセーフティ・ネットが提供されるため、想定外の問題が発生した場合は移行を取り消して元の状態に戻すことができます。トランスポータブル表領域や Oracle Data Pump を使用して、Exadata への移行を実行することもできます。Oracle Database 間でのデータ移行に使用できる手法であれば、いずれも Exadata に適用できます。

#### Exadata によるデータ保護の強化

Exadata は、顧客がオラクル製品に期待する高可用性 (HA) の基準に沿うように設計されています。Exadata 使用時にも、データベース機能とツールは、すべて Exadata 以外の従来のストレージ使用時と同様に機能します。そのため、ユーザーとデータベース管理者は、使い慣れたツールを使用して、Oracle Database に関する既存の知識と手順を活用できます。Exadata のアーキテクチャでは、シングル・ポイント障害がすべて排除されています。データの継続的な可用性と保護を実現するため、Exadata には、ミラー化、障害分離、ドライブやセルの障害からの保護などの使い慣れた機能が組み込まれています。Exadata ストレージ・サーバー内での高可用性を確保するその他の機能について、以降で説明します。

#### Hardware Assisted Resilient Data (HARD) 機能の Exadata への組み込み

オラクルの Hardware Assisted Resilient Data (HARD) イニシアチブは、データ破損を未然に防止する包括的なプログラムです。データ破損が起こることはごくまれですが、いったん発生すると、データベース、ひいては企業に壊滅的な影響を与えるおそれがあります。Exadata には拡張 HARD 機能が組み込まれているため、データがよりいっそう高いレベルで保護され、エンド・ツー・エンドのデータ検証が実施されます。Exadata では、格納データの広範な検証が実施されます。この検証には、チェックサム、ブロック位置、マジック・ナンバー、ヘッド・チェックとテール・チェック、配置エラーなどが含まれます。これらのデータ検証アルゴリズムを Exadata 内に実装すると、破損したデータが永続ストレージに書き込まれることを防止できます。また、HARD を従来のストレージで使用した場合にもこれらのチェックと保護は実施されるため、手動で実行する必要はありません。

#### Oracle Data Guard

Oracle Data Guard は、Oracle Database のソフトウェア機能であり、データベースを故障、障害、エラー、データ破損から保護するために、1 つまたは複数のスタンバイ・データベースを作成、保守、および監視します。Exadata でも Oracle Data Guard の機能は変わらず、本番データベースとスタンバイ・データベースの両方に使用できます。Exadata ストレージで Oracle Active Data Guard を使用すれば、問合せやレポートを本番データベースからスタンバイ・データベースに極めて高速にオフロードできるため、本番データベースでの重要な作業に影響を与えることなく障害からの保護を実現できます。

#### Oracle Flashback

Exadata には、Oracle Flashback テクノロジーを活用して、過去にさかのぼってデータの表示やリストアを行う機能が搭載されています。この Oracle Flashback 機能は、Exadata でも Exadata 以外の環境と同様に機能します。この Oracle Flashback 機能を使用すると、データベースをオンラインにしたままで、履歴データの問合せ、変更分析の実行、セルフサービスの修復を行い、論理的な破損からリカバリできます。つまり、Exadata に組み込まれた Oracle Flashback の機能を使用することで、ユーザーはスナップショットのような機能を使用して、エラーが発生する前の状態にデータベースをリストアできます。

#### Oracle Recovery Manager および Oracle Secure Backup

Exadata では、コマンドラインと Oracle Enterprise Manager ベースのツールである Oracle Recovery Manager と連携して、Oracle Database のバックアップとリカバリが効率的に実行されます。既存の RMAN スクリプトはすべて、Exadata 環境で変更することなく使用できます。Oracle RMAN はサーバーと緊密に連動するように設計されており、バックアップおよびリストアの際にブロック・レベルの破損を検出する機能が備わっています。Oracle RMAN は、ファイルの多重化やバックアップ・セットの圧縮によってバックアップ時のパフォーマンスと使用領域の最適化を行い、さらに、Oracle Secure Backup とテープ・バックアップ用のサード・パーティ製のメディア管理製品との統合を行います。

## 結論

現在の企業では、1つの共通インフラストラクチャにすべてのアプリケーションを統合配置できる統合データベース・プラットフォームに対するニーズがますます高まっています。共通インフラストラクチャは、OLTP、DW、複合ワークロードのすべてにおいて、データセンターに必要な効率性と再利用性を提供し、社内でグリッド・コンピューティングを実現します。さまざまなアプリケーションに合わせて特殊用途のカスタム・システムを構築または使用すると、無駄が多くコストもかさみます。企業のIT予算が縮小される中で、処理が必要なデータの量は日増しに増加しています。ITソフトウェアとハードウェアの総所有コスト（TCO）を調査すると、すべてのアプリケーションを配置できる高パフォーマンスの共通インフラストラクチャの選択につながります。

Exadata ベースの Database Machine を IT インフラストラクチャに組み込むことで、企業には次のような利点があります。

- データベースのパフォーマンスが向上し、時間あたりの処理量が大幅に増加します。
- 共通インフラストラクチャに配置を統合することで、変更や成長にスケーラブルかつ段階的に対応できます。
- ミッションクリティカルなデータの可用性と保護が確保されます。



Oracle Exadata Database Machine および  
Oracle Exadata Storage Server の技術概要  
2010 年 10 月 5 日  
著者 : Ronald Weiss

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

海外からのお問い合わせ窓口 :  
電話 : +1.650.506.7000  
ファクシミリ : +1.650.506.7200  
oracle.com



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

Copyright © 2010, Oracle and/or its affiliates. All rights reserved.

本文書は情報提供のみを目的として提供されており、ここに記載される内容は予告なく変更されることがあります。本文書は一切間違いがないことを保証するものではなく、さらに、口述による明示または法律による黙示を問わず、特定の目的に対する商品性もしくは適合性についての黙示的な保証を含み、いかなる他の保証や条件も提供するものではありません。オラクルは本文書に関するいかなる法的責任も明確に否認し、本文書によって直接的または間接的に確立される契約義務はないものとします。本文書はオラクルの書面による許可を前もって得ることなく、いかなる目的のためにも、電子または印刷を含むいかなる形式や手段によっても再作成または送信することはできません。

Oracle および Java は米国 Oracle Corporation およびその子会社、関連会社の登録商標です。その他の名称はそれぞれの会社の商標です。

AMD、Opteron、AMD ロゴおよび AMD Opteron ロゴは、Advanced Micro Devices の商標または登録商標です。Intel および Intel Xeon は Intel Corporation の商標または登録商標です。すべての SPARC 商標はライセンスに基づいて使用される SPARC International, Inc. の商標または登録商標です。UNIX は X/Open Company, Ltd. によってライセンス提供された登録商標です。  
0410

**SOFTWARE. HARDWARE. COMPLETE.**