

データ・ウェアハウスおよび  
ビジネス・インテリジェンス  
のための Oracle Database 11g

Oracle ホワイト・ペーパー  
2007 年7 月

# データ・ウェアハウスおよびビジネス・ インテリジェンスのための Oracle Database 11g

はじめに .....	3
統合 .....	3
Oracle Warehouse Builder.....	3
主要なデータベース統合機能.....	5
パフォーマンス .....	5
スケーラビリティ .....	7
パーティション化.....	7
圧縮.....	7
Oracle Real Application Clusters .....	8
並行処理.....	8
分析 .....	8
データ・マイニング .....	9
OLAP.....	10
結論 .....	11

# データ・ウェアハウスおよびビジネス・インテリジェンスのための Oracle Database 11g

## はじめに

オラクルのデータ・ウェアハウス・ソリューションには、以下の広範なデータベース機能およびオプションが含まれます。

Oracle Warehouse Builder

Oracle Partitioning

Oracle Real Application Clusters

Oracle OLAP

Oracle Data Mining

Oracle Management Pack

これらの製品の詳細は

<http://www.oracle.co.jp/oracle11g/>のホワイト・ペーパーおよびデータ・シートを参照してください。

Oracle Database 11g は、データ・ウェアハウスおよびビジネス・インテリジェンス向けの包括的なデータベース・プラットフォームです。業界トップクラスのスケーラビリティ、パフォーマンス、緊密に統合された分析、および組み込み統合とデータ品質が、すべて 1 つのプラットフォームに結合され、信頼性の高い低コストのグリッド・インフラストラクチャで実行されます。Oracle Database 11g は、数百テラバイトの実証済みのスケーラビリティと業界有数のパフォーマンスとともに、データ・ウェアハウスおよびデータ・マートの最善の機能を提供します。また、OLAP、データ・マイニング、および統計機能をデータベースに直接組み込み、独自に統合された分析用のプラットフォームも提供します。スタンドアロンの分析エンジン機能を提供し、企業のスケーラビリティ、セキュリティ、および Oracle Database の信頼性を保証します。データの統合はデータ・ウェアハウスの中心的な要件なので、Oracle Database 11g には、オラクルのスケーラブルなデータ変換と異機種データ・アクセス機能を活用する主要な ETL ツールである Oracle Warehouse Builder が含まれます。

このホワイト・ペーパーは、データ・ウェアハウスの Oracle Database 11g 機能の概要を説明し、主要な機能とテクノロジーを取り上げます。この主要な機能とテクノロジーによって、Oracle ベースのビジネス・インテリジェンスおよびデータ・ウェアハウス・システムは、情報の統合、高速な問合せ、非常に多くのデータ量への拡張、およびデータの分析を実行できます。

## 統合

現在の情報アーキテクチャは、数年前に比べはるかに動的です。企業は、より早く多くの情報を必要としています。また、広範なユーザーおよびアプリケーションに多くの分析を実行しています。

## Oracle Warehouse Builder

このようなビジネス要件に対処するため、Oracle Database 11g には、主要なデータ統合ツールの Oracle Warehouse Builder (OWB) が含まれます。OWB の中核機能は、無償のデータベース機能として装備されています。また、特定の統合要件に対応する 3 つのオプションがあります。

- 基本製品：Oracle ユーザーが、データ・マートまたはデータ・ウェアハウスを任意のサイズまたは複雑さで効率的に構築できるように設計されています。企業対応のマルチユーザー・メタデータ・リポジトリ、データ・モデリング機能、さまざまな変換および抽出技術、'ELT'アーキテクチャのパフォーマンスおよびスケーラビリティが含まれます。
- Enterprise ETL オプション：このオプションは、特にパフォーマンスと生産性を向上させます。また、以下の機能が含まれます。
  - 高度なデータ・ロード・オプション
  - 再利用可能なコンポーネントを使用した開発者の生産性
  - 緩やかに変化するディメンション管理の組み込み
  - エンドツーエンドのデータ系統と影響分析
  - 高度な構成管理のサポート
- Data Quality オプション：製品に完全に統合され、市場の他のツールにはない固有の利点を持ち、データ・プロファイリング、データ・ルール（実質はビジネス・ルール）および情報準拠機能をサポートします。
- コネクタ：主要な動作アプリケーションの最適なアクセスを提供します。オラクルは、Oracle E-Business Suite、Peoplesoft Enterprise、Siebel (CRM)、および SAP R/3 のコネクタを提供します。

Oracle Warehouse Builder のおもな利点は、単一のツール内に統合された充実した機能です。データ・モデリング、データ準拠、およびデータ品質は、企業データ統合ツールが保持すべき中核機能です。ただし、このアーキテクチャのおもな利点は、コンポーネントの統合です。一般的なリポジトリおよびユーザー・インタフェース内のすべての機能を提供します。



図 1

Oracle Warehouse Builder (OWB) は、単一のリポジトリ上の単一ツールでこれらのすべての機能を提供して、データ統合の長年の課題を解決します。多くの統合ソリューションでは、これらの異なる機能に個別のツールを提供します。しかし、あるツールでデータ・モデリング、別のツールで ETL マッピング、さらに別のツールでデータ・プロファイリングを実行するのは、きわめて非効率적입니다。OWB は、統合プロセス全体に対して単一のメタデータ・リポジトリと UI を提供します。

Oracle Warehouse Builder は、抽出-ロード-変換 (‘ELT’) アーキテクチャを使用します。外部のデータ変換エンジンではなく、OWB が Oracle データベース内ですべての変換を実行し、データベース・プラットフォームのスケラビリティおよびパフォーマンスを活用します。

## 主要なデータベース統合機能

Oracle データベースには、広範な抽出、ロード、および変換機能があります。これらの機能は OWB で実行され、データが Oracle データ・ウェアハウスに統合される環境で使用できます。

これには、次の機能が含まれます。

- 非 Oracle システムにアクセスするデータベース Gateway
- フラット・ファイルの高速なデータ・ロードのローダー・ユーティリティ
- データ変換用の SQL 拡張子：MERGE 文
- 表関数：効率的な並行ユーザー定義変換
- Oracle データベースからの待機時間の短いログ・ベースのデータ取得の変更

## パフォーマンス

オラクルは、各タイプのデータ・ウェアハウス環境のパフォーマンスの最適化を実現します。多くの場合、非常に異なる操作を実行するさまざまなユーザーや問合せパフォーマンスのさまざまな予測および要件のため、データ・ウェアハウス・ワークロードは複雑です。

オラクルは、以下の機能を提供してデータ・ウェアハウスのパフォーマンス要件を満たします。

- 各タイプの問合せおよびワークロードに対する広範なパフォーマンス最適化技術。

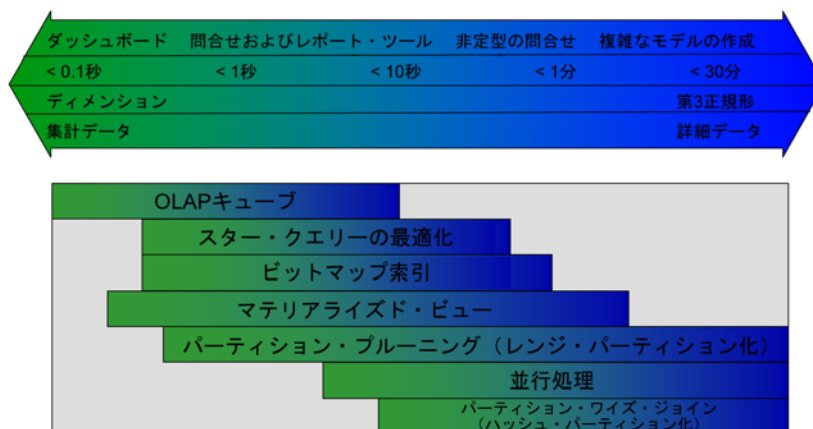


図 2

- 複雑で多様なワークロードでもデータベースのパフォーマンスを確保する高度なリソース・マネージャ。Database Resource Manager を使用して、エンド・ユーザーをリソース・コンシューマ・グループにグループ化できます。DBA は、各グループに対するリソースの割当てを正確に制御できます。各グループに対して事前に問合せを管理するポリシーおよび問合せをキューに格納するポリシーを指定したり、CPU の使用量を制御したりできます。
- データ・ロードが問合せのパフォーマンスに影響しないようにする特許権を有する一意な読取り一貫性モデル。オラクルは、マルチ・バージョンの読取り一貫性と呼ばれるテクノロジーを使用して、同時アクセスの課題を解決します。この一意なテクノロジーは、15 年以上にわたるオラクルの同時実行性モデルの基盤です。マルチ・バージョンの読取り一貫性は、要求されたデータの一貫したビューをユーザーが常に参照できることを保証します。大容量のデータ・ウェアハウス表のトリクル・フィードの更新など、他のユーザーが問合せの実行中に基本となるデータを変更する場合、問合せの開始時に存在するデータのバージョンが保存されます。問合せに返されるデータは、問合せ実行中に他の更新が発生していても、問合せが送信された時点でのコミットされたすべてのトランザクションを含むデータベースの状態を常に反映します。このテクノロジーによって、オラクルは、データ・ウェアハウス環境内のほぼリアルタイムのデータ・ロード処理で一意な位置付けを与えられています。
- データ・ウェアハウスの継続的なパフォーマンス・チューニングを自動化する Management Pack。

もっとも関心を引く Oracle Database 11g の新しいパフォーマンス機能に、OLAP ベースのマテリアライズド・ビューがあります。データ・ウェアハウス内のスター・スキーマに対する従来のパフォーマンス・チューニング方法では、サマリー表 (マテリアライズド・ビュー) を長く使用してきました。この方法を使用する場合、1 つ以上のサマリーを事前に計算し、それらのサマリーをデータ・ウェアハウスに格納すると、問合せのパフォーマンスが向上します。これらのサマリーは、スター・スキーマ表の基本セットに対する問合せを書き換えてサマリー・データにアクセスするクエリー・リライト機能がデータベースで提供されるため、アプリケーションに対して完全な透過性があります。マテリアライズド・ビューの課題の 1 つは、サマリー表として格納できる候補の数千または数百万の集計を 1 つのスター・スキーマに設定できる点です。候補となるすべての集計にマテリアライズド・ビューを作成することは、現実的ではありません。領域の利用およびメンテナンスのコストが膨大になります。しかし、このシナリオは、OLAP ベースのマテリアライズド・ビューで解決できます。OLAP ベースのマテリアライズド・ビューは、候補となるすべての集計を迅速に取得できる OLAP データ構造を活用します。このため、OLAP ベースのマテリアライズド・ビューを使用したスター・スキーマは、数千のリレーショナル・マテリアライズド・ビューにスター・スキーマのパフォーマンスを適用できます (ただし、高度に圧縮されて効率的な更新機能を提供する単一の管理しやすい OLAP キューブを使用)。汎用データ・ウェアハウス・ワークロードへのこの場合の OLAP テクノロジーは、オラクルの一意な技術です。

以前の OLAP のソリューションでは OLAP ベースの独自の API を通じてアクセスしていましたが、Oracle OLAP では、SQL ベースのツールを使用して、すべてのデータ・ウェアハウス・ユーザーに対する OLAP のパフォーマンス上の利点を提供します。

## スケーラビリティ

最大のデータ・ウェアハウスのサイズは、飛躍的に増加しています。現在、多くのビジネス・プロセスが次々と完全に自動化され、より細かいレベルで多くのデータが収集されています。このため、これらのデータ量は増加の一途をたどっています。

Oracle Database には、スケーラビリティを実現する 4 つの主要な機能（パーティション化、圧縮、クラスタリング、並行処理）があります。

## パーティション化

Oracle Partitioning は、大規模なデータベースの管理に不可欠です。"分断攻略"技術を有効にして、特に表が大きくなる場合にデータベースの大規模な表を管理します。

データベースのデータが来年に現在の 2 倍になっても、エンド・ユーザーは、実行速度が半分になるアプリケーションを許容しないでしょう。データベースのメンテナンスやバッチ処理に 2 倍の実行時間が提供されることもないでしょう。また、IT マネージャがデータ・ウェアハウスのハードウェア予算を 2 倍にすることもないでしょう。パーティション化とは、管理またはハードウェアのリソースを過度に増やすことなく、一貫したパフォーマンスを維持しながら大規模なデータセット向けのデータベースを拡張できる機能です。パーティション化では、大規模な表をより細かい単位に分割するため、データが増加しても、一定のパフォーマンス・レベルでデータ管理ウィンドウと多くのエンド・ユーザーの間合せを保守できます。

オラクルでは、Oracle8 で導入してから 10 年以上にわたってパーティション化テクノロジーの開発を続けています。Oracle Database 11g で、DBA がカスタム・パーティション化スキームを定義する機能、パーティション化された表の豊富な管理コマンド、および最適なパーティションの実装方法を管理者に提供するパーティション・アドバイザとともに、表をパーティション化する 9 つの方法を提供します。

また、パーティション化は、Oracle データベース内の ILM ("情報ライフサイクルマネジメント") 戦略を有効にします。パーティション化された単一の表は、複数のストレージ層に分散できます。古いパーティションに対応している頻繁にアクセスしない古いデータは、安価なストレージ・デバイスに格納できます。大規模なデータベースの場合、この方法によって大幅にコストを削減できます。

## 圧縮

Oracle9i Database Release 2 の導入以来、圧縮は、データ・ウェアハウスの顧客がもっとも使用する機能の 1 つです。顧客が大量のデータを格納する場合、圧縮の使用は解決策として自然な流れです。オラクルの圧縮アルゴリズムは、問合せパフォーマンスにほとんど悪影響を与えることなく、リレーショナル表に格納されたデータを圧縮する一意なメカニズムを提供します。

一般的な圧縮比率（2 : 1 から 5 : 1 の範囲）を使用しているため、この機能の普及は驚くことではありません。

Oracle Database 11g では、圧縮を強化して、頻繁な小規模の更新および挿入をサポートします。以前のリリースでは、'ダイレクトパス'処理（単一の SQL 文またはロード・ユーティリティを使用する場合に数千の行が挿入される'バルク'処理）でデータが表に挿入される場合にデータが圧縮されました。圧縮された表に小規模な更新および挿入を行う場合、新しいデータは以前のリリースで圧縮されません。この Oracle Database 11g の新しい強化機能によって、頻繁に'トリクル・フィード'または更新されるデータ・ウェアハウス表のパーティション化に利点がもたらされます。

## Oracle Real Application Clusters

Oracle Real Application Clusters（Oracle RAC）を使用すると、単一のデータベースを複数のサーバーに拡張できます。データ・ウェアハウスを使用する場合、数百の CPU への拡大を'スケールアウトする'ソリューションを提供します。Oracle RAC のアーキテクチャは、オラクルのデータ・ウェアハウスに一意な利点を提供します。新しい演算能力をデータ・ウェアハウス・システムに追加する機能に加えて、ノード障害に対する高可用性も実現します。

また、Oracle RAC は、単一のデータベース内の複数のワークロードを管理する高い柔軟性を提供します。'RAC サービス'を使用すると、異なるアプリケーションを個別のサービスに割り当てられます。このため、クラスターの別々のノードで実行できます。たとえば、データ・マート・ユーザーを特定のノード・セットに割り当て、ETL 操作を別のノード・セットに割り当てることができます。Oracle RAC のおもな利点は、これらのワークロードを動的に変更および調整する機能です。たとえば、夜間、動的かつ容易に多くのノードを ETL 処理に割り当てる一方で、これらのノードを再度割り当てて、日中に処理の間合せを実行できます。

## 並行処理

並行処理は、単一のデータベース操作の実行に複数のプロセッサを活用する機能です。オラクルのスケラブルなパラレル実行アーキテクチャは、動的な調整を行って、すべてのユーザーの要求を満たします。大規模な処理の場合、使用できるすべてのノードと CPU でデータベース操作の並行処理を実行します。小規模な処理の場合、Database Resource Manager を使用して優先順位の高い操作の適切なリソースの取得を実現する一方で、並行処理を動的に調整して全体のシステム・スループットを最大化します。

## 分析

OLAP、統計、データ・マイニングなどのテクノロジーは、データ・ウェアハウスおよびビジネス・インテリジェンスに新しく導入された技術ではありません（これらのほとんどのテクノロジーは、データ・ウェアハウスの以前から存在します）。データ・ウェアハウスのユーザーは、これらの各機能を備えた製品を何年間も購入してきました。ただし、通常、それぞれに固有のエンジン（OLAP 製品の計算エンジン、統計製品のデータ・エンジン、データ・マイニング製品のマイニング・エンジン）があります。

つまり、企業全体のビジネス・インテリジェンス環境では、少なくとも 4 つの異なるタイプの「データ・エンジン」を維持します。それぞれ固有のサーバー、固有の管理インフラストラクチャ、固有のセキュリティ管理、固有の高可用性インフラストラクチャが必要になります。各エンジンは API に組み込まれ、開発者ツールおよびエンド・ユーザー・ツールの固有のセットがあります。BI テクノロジーのスタック全体を複製する複雑さとコストは甚大です。

Oracle Database は、完全に異なる方法を提供します。オラクルは、データベース・エンジン内に OLAP、データ・マイニング、および統計を実装しました。詳細な分析を行うためにデータ・ウェアハウスから他の分析エンジンにデータを移行する代わりに、高度な分析アルゴリズムがデータの存在するデータベースに組み込まれました。また、すべての分析機能に SQL アクセスが提供されるので、SQL ベースのツールまたはアプリケーション環境で分析を実装できます。

企業のビジネス・インテリジェンス環境のバックエンド・データ・アーキテクチャを合理化する多くの利点に加えて、Oracle Database 内の分析の統合によって、スタンドアロン環境にはない多くの利点が提供されます。たとえば、スタンドアロンの OLAP サーバーは数百の CPU またはサーバー群に拡張されますか。データ・マイニング・エンジンはユーザー認証サーバーにどれだけ簡単に統合されますか。すべてのデータ・セキュリティ・ポリシーを透過的に実装できますか。統計分析の結果をデータ・ウェアハウス・データにどれだけ簡単に統合できますか。Oracle Database では、OLAP、データ・マイニング、および統計の緊密な統合によって、これらのすべての問題が簡単に解決します。

## データ・マイニング

Oracle Data Mining は、Oracle Database に組み込まれた強力なソフトウェアで、データに隠された新しい事実を発見できます。Oracle Data Mining は、企業によるベスト・カスタマへのターゲット化、不正行為の発見および防止、キー・パフォーマンス・インディケータ (KPI) に関するもっとも影響力の大きい属性の確認、さらにデータに隠された価値ある新しい情報の発見を支援します。Oracle Data Mining を使用すると、技術者は、データのパターンの発見、キー属性の特定、および新しいクラスターや関連性の確認、および価値ある事実の発見ができます。

これらのビジネス問題に対処するため、企業は、Oracle Data Mining の幅広い最新のアルゴリズムを使用してデータから新しい情報を発見できます。データ・マイニング・アルゴリズムは、データを分析してパターンおよび関係を検出する機械学習技術です。分析およびビジネス問題の種類に応じてさまざまなアルゴリズムが効果的であるため、オラクルは複数のアルゴリズムを提供します。

ほとんどのデータ・マイニング・アルゴリズムは、「監視あり学習」または「監視なし学習」のデータ・マイニング技術に大別できます。監視あり学習では、データ・アナリストがターゲット属性または従属変数（たとえば、特定の製品を購入した顧客など）を確認する必要があります。監視あり学習技術は、データを絞り込み、

他の属性とターゲット属性のパターンおよび関係を発見します（たとえば、将来の顧客が特定の製品を購入するかどうかを示す特性など）。Oracle Data Mining の監視あり学習アルゴリズムには、Naïve Bayes、Decision Tree、一般化線形モデル (Generalized Linear Models)、およびサポート・ベクター・マシン (Support Vector Machine) が含まれます。

データ・マイニング・アルゴリズムのもう 1 つの広範なカテゴリは、“監視なし学習”向けです。これらのシナリオにターゲット属性はありません。代わりに、データ・マイニング・アルゴリズムは、Apriori で定義されたビジネス目標とは関係のないデータの関連性およびクラスタを発見しようとします。これらのアルゴリズムには、拡張 k-Means クラスタリング、O-Cluster クラスタリング、相関ルール (マーケット・バスケット分析)、Nonnegative Matrix Factorization (NMF) が含まれます。

Oracle Data Mining には、データ・アナリストがデータ・マイニング・モデルを構築、評価、および適用するグラフィカル・ユーザー・インタフェースの Oracle Data Miner が含まれます。Oracle Data Miner を使用すると、データ・アナリストは、完全な柔軟性を備えたデータ・マイニング・プロセスを実行し、グラフ形式および表形式で結果を表示できます。Oracle Data Miner は、マイニング・アクティビティに関連する PL/SQL コードを生成できます。

## OLAP

Oracle OLAP は、Oracle Database に組み込まれた機能をすべて備えたオンライン分析処理 (OLAP) エンジンです。Oracle OLAP は、パフォーマンス・セクションで説明したように問合せのパフォーマンスを向上させ、豊富な分析コンテンツを追加して、データ・ウェアハウスを強化します。

Oracle OLAP の中核機能はキューブです。Oracle データベース内で管理されるこのデータ構造は、高度に最適化された多次元形式でデータを格納します。キューブは、次元データのスケラブルな圧縮保存、高速な差分更新、高速な問合せパフォーマンス、および高度な分析計算を実行または格納する機能を提供します。

Oracle OLAP を使用したオラクルの戦略は、これらの中核的な OLAP の利点をデータ・ウェアハウスに導入することです。ビジネス・インテリジェンス・ツールまたは他の SQL ベースのアプリケーションで OLAP を活用できるように、標準的な SQL を使用して Oracle OLAP の主要な機能を公開すると、これが実現します。

主要な SQL 統合機能には、一連のディメンション表およびファクト表として多次元キューブを公開する SQL ビュー、マテリアライズド・ビューとして OLAP キューブを使用できるキューブ・ベースのマテリアライズド・ビュー、SQL ベースの問合せツールと SQL 開発者によってすべてのキューブおよび属性を素早く識別できるデータ・ディクショナリの統合、および Oracle オプティマイザ内のキューブ・アクセス・パスの統合を通じて OLAP キューブにアクセスする機能が含まれます。

## 結論

Oracle Database は、データ・ウェアハウス市場のリーダーです。スケーラビリティとパフォーマンスの強固な基盤で構築され、ほぼリアルタイムのデータ・ウェアハウス向けのオラクル固有の読取り一貫性モデル、柔軟で強力な一連の表パーティション化機能、リレーショナル環境を強化する OLAP テクノロジーの利用、グリッド・アーキテクチャの他に例を見ないほどのサポートなど、革新的な機能が強化されています。Oracle Database 11g は、データ統合、OLAP、およびデータ・マイニング機能を組み込み、データ・ウェアハウスと他のビジネス・インテリジェンス・アプリケーションのもっとも包括的なデータベース・プラットフォームも提供します。さらに、他のエンタープライズ・システムと同等のデータ・ウェアハウスの主要な要件である企業対応の可用性、セキュリティ、および管理性を提供します。



データ・ウェアハウスおよびビジネス・インテリジェンスのための Oracle Database 11g  
2007 年 7 月  
著者 : George Lumpkin

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

海外からのお問合せ窓口 :  
電話 : +1.650.506.7000  
ファクシミリ : +1.650.506.7200  
[www.oracle.com](http://www.oracle.com)

Copyright © 2007, Oracle. 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。オラクル社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、本文書はオラクル社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 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。