

Oracle Advanced Compression

Oracle ホワイト・ペーパー
2007 年6 月

注：

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。また、情報提供を唯一の目的とするものであり、いかなる契約にも組み込むことはできません。下記の事項は、マテリアルやコード、機能の提供を確約するものではなく、また、購買を決定する際の判断材料とはなりません。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いたします。

はじめに	4
ORACLE ADVANCED COMPRESSION	4
リレーショナル・データの圧縮	5
革新的なアルゴリズム	5
表の圧縮の利点	6
最小限のパフォーマンス・オーバーヘッド	6
非構造化データの圧縮	7
SecureFiles Deduplication	7
SecureFiles Compression	8
バックアップ・データの圧縮	8
Oracle Data Pump による圧縮	9
Oracle Recovery Manager による圧縮	9
ネットワーク・トラフィックの圧縮	10
結論	10

はじめに

企業は、効率的な事業運営に必要なデータ量が急増しているという問題に直面しています。このデータ増加の傾向は、いくつかの主要な要素に起因します。その一つは、米国サーベンス・オクスリー（SOX）や法医療保険の相互運用性と説明責任に関する法律（HIPAA）など、規制に関する状況が最近変化したことにより、企業が膨大な量の情報を長期間にわたって保存する必要が出てきたことです。ブロードバンド技術の進歩により、インターネット経由のさまざまなマルチメディア・コンテンツが大量に配布されるようになったことも、データ量全体が増加する原因となっています。データの増加傾向をさらに加速させたのが Web 2.0 の登場です。コラボレーティブ・アプリケーションにより、ユーザーが作成する大量のコンテンツが普及し始めています。概算では、データ量は 2, 3 年ごとにほぼ倍増していると指摘されています。

このデータ量の急激な増加によって、IT 管理者は管理上の困難な問題に直面しています。第一の課題は、ストレージ・コストの高騰です。メガバイト当たりのストレージ・コストは過去数年で激減しているにもかかわらず、オンラインで保存する必要のあるデータ量が急増していることにより、ストレージが IT 予算最大のコスト要因の 1 つになっています。さらに、たとえデータ量が急増しても、アプリケーションのスケラビリティとパフォーマンスは、ビジネスの需要を満たし続ける必要があります。

Oracle Database 11g は、このような問題に対処するため、Advanced Compression Option を導入しました。オラクルの圧縮技術の革新により、顧客による大量データの管理コストとリソースの削減を支援します。これらの優れた新技術により、テラバイトサイズのデータベースが導入され、企業内のデータ・センターで普及しつつあります。

ORACLE ADVANCED COMPRESSION

Oracle Database 11g Advanced Compression Option には、顧客の最大限のリソース活用とコスト削減を支援する、一連の包括的な圧縮機能が導入されています。これにより、(構造化された) 通常のリレーショナル・データ、非構造化データ（ドキュメント、スプレッドシートなど）やバックアップ・データなどのあらゆるタイプのデータの圧縮が可能になり、IT 管理者は、データベース記憶域全般のフットプリントを大幅に削減できます。また、圧縮におけるもっとも明白な利点は、ストレージ・コストの削減と見なされることが多いですが、Advanced Compression Option の革新的な技術は、メモリーやネットワーク帯域幅などの IT インフラストラクチャのあらゆるコンポーネントに関するリソース要件やテクノロジー・コストを削減するように設計されています。

リレーショナル・データの圧縮

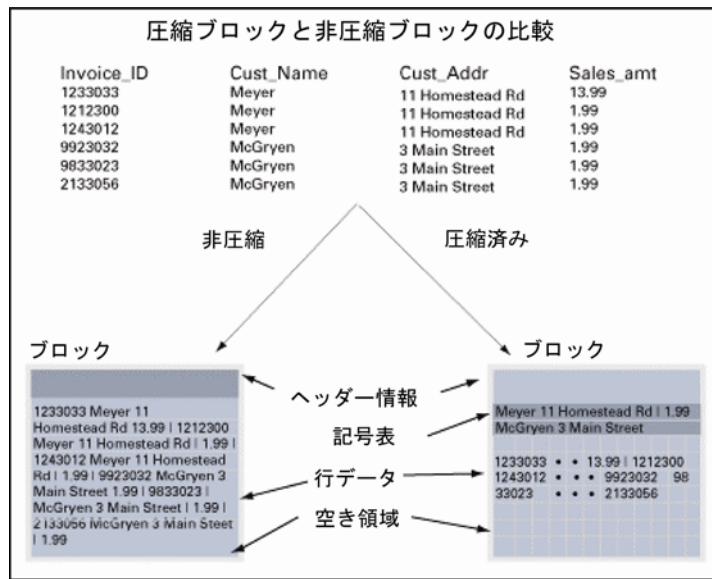
オラクルは、データベースの圧縮技術を導入したパイオニアです。Oracle9i Database は、数年前に表の圧縮を導入し、CREATE TABLE AS SELECT...などダイレクト・パス・ロード等のバルク・ロード・オペレーションの間でデータを圧縮できるようにしました。この圧縮形式は、ほとんどのデータがバッチ・プロセスの使用中にデータベースにロードされるといったデータ・ウェアハウスの環境に最適でした。Oracle Database 11g では、OLTP 表の圧縮と呼ばれる新機能が導入されており、この機能を使用すると、INSERT、UPDATE、DELETE などの従来の DML を含むあらゆるタイプのデータ操作オペレーション中にデータを圧縮できます。さらに、この新機能によって、書込み動作のオーバーヘッドが削減され、トランザクション環境や OLTP 環境に適應するために、パフォーマンスが大幅に向上します。この重要な革新により、圧縮の利点がすべてのアプリケーションのワークロードへ拡張されます。

Oracle Database 9i で導入された表の圧縮機能は、Oracle9i Database Enterprise Edition (EE) のベース機能であり、Oracle Database 11g のベース機能でもあります。ただし、新しい OLTP 表の圧縮機能は、Oracle Database 11g Enterprise Edition に加えて使用許諾が必要な Advanced Compression Option の一部です。

革新的なアルゴリズム

Oracle では、リレーショナル・データとともに動作するよう特別に設計された、独自の圧縮アルゴリズムを使用しています。このアルゴリズムは、データベース・ブロック内や複数の列間の重複値を排除することによって動作します。圧縮されたブロックには、圧縮メタデータを維持する記号表と呼ばれる構造体が含まれます。ブロックが圧縮されると、最初に重複値のコピーが記号表に 1 つ追加されることにより、重複値が排除されます。そして、各重複値が、記号表内の適切なエントリへの短い参照に置き換えられます。この革新的な設計では、圧縮されたデータを元の状態へ変換するために使用されるメタデータがブロック内に含まれるため、圧縮されたデータはデータベース・ブロック内で自己完結します。グローバルなデータベースの記号表を維持する競合他社の圧縮アルゴリズムと比較すると、オラクル独自のアプローチは、圧縮されたデータにアクセスする際に追加の I/O を導入しないことにより、大幅なパフォーマンス上の利点を提供します。

表 1.圧縮ブロックと非圧縮ブロックの比較



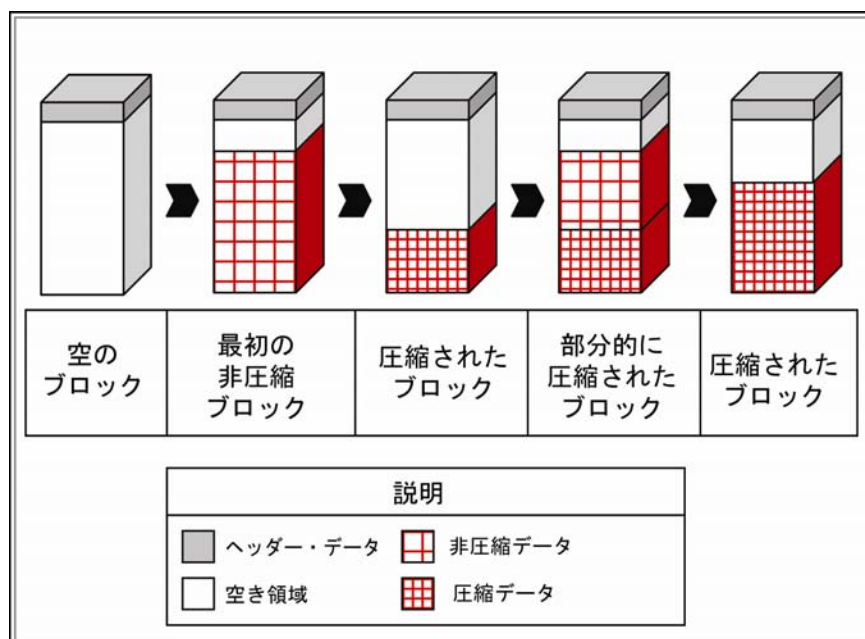
表の圧縮の利点

ある環境で得られる圧縮率は、圧縮されたデータの性質、特にデータのカーディナリティによって異なります。通常、表圧縮機能を使用することにより、顧客はストレージ領域の消費において 2~3 倍の削減が期待できます。つまり、非圧縮データ量が消費する領域量は、圧縮されたデータ量が消費する領域量の 2~3 倍になるということです。圧縮の利点は、ディスク上のストレージ節約の域のみにとどまりません。Oracle の機能の重要な利点の 1 つは、最初にブロックを解凍することなく、圧縮されたブロックを読み取れることです。したがって、圧縮データにアクセスする際にパフォーマンスが大きく低下することはありません。実際、多くの場合、Oracle ではアクセスに必要なブロックが減るため、I/O の削減によってパフォーマンスは向上します。さらに、メモリーのフットプリントを増加させることなくキャッシュ内の格納データを増やせるため、バッファ・キャッシュの効率が向上します。

最小限のパフォーマンス・オーバーヘッド

すでに説明したように、表の圧縮機能には、読取り動作に対する悪影響はありません。ただし、データの書込み中は、圧縮には書込み動作のパフォーマンス・オーバーヘッドの排除を不可避にする追加の動作の実行が必要とされます。オラクルは、そのような OLTP 表の圧縮のオーバーヘッドを最小限にするために検討を重ねました。Oracle では、書込み動作が発生するたびにデータを圧縮するのではなく、バッチ・モードでブロックを圧縮します。新しく初期化されたブロックは、ブロック内のデータが内部で制御されるしきい値に達するまで、圧縮されずに維持されます。トランザクションによってブロック内のデータがこのしきい値に達すると、ブロックのすべてのコンテンツが圧縮されます。さらに、より多くのデータがブロックに追加されて再びしきい値に達すると、ブロック全体が再圧縮されて圧縮の最高レベルに到達します。このプロセスは、圧縮を続けてもそれ以上ブロックに利点が得られないと Oracle が判断するまで繰り返されます。ブロックの圧縮をトリガーするトランザクションのみ、圧縮のオーバーヘッドは最小限で済みます。したがって、圧縮されたブロック上の OLTP トランザクションのほとんどのパフォーマンスは、非圧縮ブロックの場合のパフォーマンスと同様になります。

表 2.ブロックの圧縮プロセス



非構造化データの圧縮

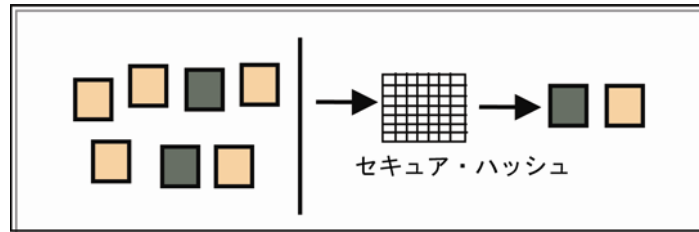
Oracle Database 11g の新機能の SecureFiles は、ドキュメント、スプレッドシート、XML ファイルなどの非構造化コンテンツを保存するための、‘両方の長所を備えた’アーキテクチャを提供します。SecureFiles は、特に Oracle データベースの利点を維持しながら、従来のファイル・システムに匹敵する高パフォーマンスをファイル・データに提供するように設計されています。SecureFiles は、ANSI 規格の LOB のスーパーセットとして設計されており、旧 LOB または BasicFiles からの移行を容易にします。SecureFiles によって、組織は、単一のセキュリティ/監視モデルや統合されたバックアップおよびリカバリ・プロセスを使用して、すべてのリレーショナル・データおよび関連するファイル・データを Oracle で管理でき、全情報にわたってシームレスな検索を実行できるようになります。Oracle Database 11g の Advanced Compression Option には、SecureFiles データのストレージのフットプリントを大幅に削減するテクノロジーが搭載されています。

SecureFiles Deduplication

アプリケーションにおいて、ファイルの完全なレプリカを保存することはきわめて一般的です。その例として、複数のユーザーが同一の添付ファイルを受信する電子メールのアプリケーションがあります。SecureFiles Deduplication は、SecureFiles データの重複したコピーを排除するインテリジェント・テクノロジーです。Oracle では、SecureFiles データのイメージを 1 つ保存し、重複したコピーをそのイメージへの参照に置き換えます。10 人のユーザーが 1MB の同一のファイルが添付されている電子メールを受信する、電子メール・アプリケーションについて検討してみます。SecureFiles Deduplication がいない場合、システムは 10 人のユーザーそれぞれに対してファイルのコピーを 1 つずつ保存するため、10MB のストレージが必要になります。この例の電子メール・アプリケーションで SecureFiles Deduplication を使用した場合、1MB の添付ファイルを一度保存するだけで済みます。これは、ストレージ要件の 90% の節約になります。ストレージの節約に加え、SecureFiles Deduplication によってアプリケーションのパフォーマンスも向上します。特に、SecureFiles イメージへの参照の

みが書き込まれるため、書き込み動作およびコピー動作の効率が向上します。さらに、重複した SecureFiles データがすでにバッファ・キャッシュにある場合、読取り動作が改善されます。

表 3SecureFiles Deduplication



SecureFiles Compression

Oracle Database 11g の Advanced Compression Option では、SecureFiles データのサイズを制御する別のメカニズムも提供します。前に説明した SecureFiles Deduplication に加え、SecureFiles Compression は業界標準の圧縮アルゴリズムを使用して、SecureFiles データのストレージ要件をさらに最小化します。ドキュメントや XML ファイルなどの一般的なファイルの圧縮では、サイズが 2~3 倍削減されます。組込みインテリジェンスを使用することにより、SecureFiles Compression は圧縮しても利点のないデータの圧縮を自動的に回避します。たとえば、SecureFile としてデータベースに挿入される前にサード・パーティのツールを用いて圧縮されたドキュメントなどです。

現在は、2 つのレベルの圧縮がサポートされています。高いレベルでは圧縮率が大きくなりますが、CPU の使用量が多くなります。通常、SecureFiles Compression の CPU オーバーヘッドは、3%から 5%です。圧縮データはデータの小さなチャンクに分類されるため、アプリケーションは圧縮された SecureFiles データ上でもランダム読取りおよびランダム書込みを実行できます。これにより、データベースへの挿入前にファイル全体を圧縮する場合と比較して、パフォーマンスが大幅に向上します。

バックアップ・データの圧縮

データベース内に保存されたデータの圧縮に加えて、Oracle Advanced Compression にはバックアップされたデータを圧縮する機能も含まれています。Oracle Recovery Manager (RMAN)と Oracle Data Pump は、Oracle Database 内に保存されたデータのバックアップにもっとも一般的に使用されるツールです。Oracle RMAN は、データベース・データのブロックごとのバックアップを行う、"物理"バックアップとしても知られています。これは、データベース、表領域、またはブロック・レベルのリカバリを行うために使用できます。Oracle Data Pump は、1 つ以上の表からフラット・ファイルへデータをオフロードすることによって"論理"バックアップを実行するために使用されます。Oracle Advanced Compression には、これらのツールのどちらかによって作成されたバックアップ・データを圧縮する機能が含まれています。

Oracle Data Pump による圧縮

Oracle Database 10g Release 2 では、Oracle Data Pump ジョブに関連するメタデータの圧縮機能が提供されています。Oracle Database 11g の圧縮機能は、データがエクスポート上で圧縮できるように拡張されています。Oracle Data Pump による圧縮はインライン・オペレーションであるため、ダンプ・ファイル・サイズが削減されることにより、ディスク領域が大幅に節約されます。オペレーティング・システムまたはファイル・システムの圧縮ユーティリティとは異なり、Oracle Data Pump 圧縮はインポート側でも完全にインラインであるため、ダンプ・ファイルをインポートする前に解凍する必要がありません。圧縮されたダンプ・ファイル・セットは、データベース管理者が手順を追加することなく、インポート動作中に自動的に解凍されます。

Oracle サンプル・データベースからの次の圧縮例では、OE および SH スキーマがエクスポートされると同時に、すべてのデータおよびメタデータが圧縮されました。ダンプ・ファイル・サイズは、74.67%削減されました。

3つのバージョンの gzip (GNU zip) ユーティリティと1つの UNIX 圧縮ユーティリティが、6.0MB のダンプ・ファイル・セットの圧縮に使用されました。ダンプ・ファイルの削減サイズは、Oracle Data Pump の圧縮に匹敵しました。ダンプ・ファイルの削減サイズは、データ型や他の要素によって異なることに注意してください。

完全な Oracle Data Pump の機能は、圧縮ファイルを使用することによって利用できません。通常ファイルで使用されるコマンドは、圧縮ファイル上でも動作します。ユーザーは、以下のオプションから、ダンプ・ファイル・セットのどの部分を圧縮するかを決定できます。

- **ALL** では、エクスポート動作全体に対して圧縮が有効化されます。
- **DATA-ONLY** では、圧縮された形式ですべてのデータがダンプ・ファイルに書き込まれます。
- **METADATA-ONLY** では、圧縮された形式ですべてのメタデータがダンプ・ファイルに書き込まれます。これはデフォルトです。
- **NONE** では、エクスポート動作全体に対して圧縮が無効化されます。

Oracle Data Pumpの詳細は、<http://otn.oracle.co.jp/products/database/utilities/index.html> を参照してください。

Oracle Recovery Manager による圧縮

エンタープライズ・データベースが増大し続けることにより、データベース管理者にとって重大な問題が生じています。データベース・バックアップを維持するためのストレージ要件およびバックアップ・プロシージャのパフォーマンスは、データベースのサイズに直接影響を受けます。オラクルのバックアップおよびリカバリ・ユーティリティの Oracle Recovery Manager (RMAN) は、Oracle Database 10g で圧縮機能を導入しました。Oracle RMAN による圧縮では、バックアップに必要とされるストレージが大幅に削減されます。Oracle RMAN が Oracle Database と密接に統合されていることにより、バックアップ・データはディスクまたはテープに書き込まれる前に圧縮され、リカバリ前に解凍する必要がありません。そのため、ストレージ・コストが大幅に削減されます。ただし、圧縮率が広範囲にわたることでバックアップ・パフォーマンスが影響を受けるため、バックアップ・ウィンドウが長くなります。

Oracle Advanced Compression には、バックアップのストレージ要件を大幅に削減しながら Oracle RMAN のパフォーマンスを向上させる、新しい Oracle RMAN の圧縮機能が導入されています。この機能は業界標準の ZLIB 圧縮アルゴリズムに基づいており、Oracle RMAN によって圧縮されたバックアップは、Oracle Database 10g によって圧縮されたバックアップよりも最大で 40% 速くなります。オラクルは、この大幅なパフォーマンスの改善を、圧縮率の減少を 20% 未満にとどめながら実現しています。迅速な Oracle RMAN の圧縮は、通常の営業時間中に行われる増分バックアップにとって最適なソリューションです。

ネットワーク・トラフィックの圧縮

Oracle Data Guard は、管理、監視、自動化のためのソフトウェア・インフラストラクチャを提供し、企業のデータを、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。Oracle Data Guard は、REDO データ（トランザクションのリカバリに必要な情報）を使用して、プライマリ・データベースおよびスタンバイ・データベースの同期を維持します。プライマリ・データベースでトランザクションが発生すると、REDO データが生成され、ローカル REDO ログ・ファイルに書き込まれます。この REDO データは、Oracle Data Guard REDO 転送サービスを使用してスタンバイ・サイトに転送されます。

ネットワークまたはスタンバイ・サーバーの停止により、REDO データがスタンバイ・サーバーに転送されるのを防ぐことができます。停止が解消されると、スタンバイ・データベースの同期化に必要なすべての REDO データを転送することで、Oracle は自動的に REDO ギャップ解消を実行します。Oracle Advanced Compression では、REDO データが REDO ギャップ解消中、ネットワークを通じて送信される際に REDO データを圧縮する機能が導入されています。この圧縮により、ネットワーク帯域幅が最大化され、ギャップ解消のスループットが増加します。ギャップ解消の速度は、圧縮によって最大で 2 倍になります。これにより、スタンバイ・データベースが迅速に同期化され、高可用性が実現します。

Oracle Data Guardの詳細は、<http://otn.oracle.co.jp/products/availability/index.html>を参照してください。

結論

データ量の急増という重要な問題に、企業は直面しています。収益に影響を与えずに、変化するビジネス状況に迅速に適応する必要があります。IT マネージャは、既存のインフラストラクチャを効率的に管理してコストを制御しながら、優れたアプリケーション・パフォーマンスを提供し続ける必要があります。

Oracle Database 11g の Advanced Compression Option は、この複雑な環境で IT 管理者を成功に導く、一連の堅牢な圧縮機能を提供します。企業は、Advanced Compression Option を利用することにより、データ・センターのすべてのコンポーネントを通して増大し続けるデータ要件を効率的に管理でき、最高レベルのアプリケーション・パフォーマンスを実現しながらコストを最小限に抑えられます。



Oracle Database 11g
Advanced Compression Option
2007年6月
著者: William Hodak
共著者: Carol Palmer, Timothy Chien, Sushil Kumar

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

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ます。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。