



ビジネス・インテリジェンスを実現する
パワフルな統合分析環境
IBM Red Brick™ Warehouse

目次

- 1 一般向けデータ・ウェアハウジング
- 2 シングル・ステップのロード処理
- 3 優れたロード性能
- 4 大容量のデータ・ストレージ
- 5 クエリーへの迅速な応答
- 6 データ・ウェアハウジングで成功を収めたRed Brick Warehouse

ハイライト

リレーショナル・データベース管理システムとオンライン分析処理

一般向けデータ・ウェアハウジング

業界エキスパートは、IT利用企業の92%が業務用にデータ・ウェアハウスを配備すると予測しています。データ・ウェアハウジングの人気の高まるにつれ、そのコア・テクノロジーはますます高度になっています。

データ・ウェアハウス/データマート・アプリケーションの最も重要なコンポーネントは、大量の情報を格納し、幅広いビジネス上の問いにスピーディーに信頼性の高い答えを出すために使用されるリレーショナル・データベース管理システム(RDBMS)です。

データ・ウェアハウスは通常、いろいろな見地から得られたデータの抽出と表示を行うためのオンライン分析処理(OLAP)に特化したRDBMSによって、有益なビジネス・インテリジェンスを提供します。OLAPクエリーをデータ・ウェアハウスについて実行すると、例えば、ある製品によって本年度の特定の月に得られた収益が昨年度の同じ月のものと比較してどれほどであったかを調べることができます。

Red Brick Warehouseの限界テスト

OLAPアプリケーションでは、ますます高い性能と機能性が求められています。データ・ウェアハウスまたはデータ・マートの1クエリーは、データ入力やデータ検索などを伴う標準的なオンライン・トランザクション処理(OLTP)クエリーの数千倍、数百万倍もの(質の異なる)作業量が必要となります。

本資料では、OLAPやその他のビジネス・インテリジェンス機能に特化したRDBMSであるIBM Red Brick Warehouseの優れた性能について述べます。ここで取り上げる事例は、スタースキーマ・データベースのスケラビリティ・ベンチマーク・テストの結果です。このテストは、お客様の環境において、管理された条件のもとでお客様によって実施されたものであり、実稼働型データ・ウェアハウス・アプリケーションをシミュレートするよう作成されたものです。

ハイライト

低コストでハイ・パフォーマンス

シングル・ステップのロード処理

1990年の初頭より市場に登場したRed Brick Warehouseは、標準として発展していくクライアント/サーバー・コンピューティング環境ベースの意思決定支援アプリケーションに特化したオープンなリレーショナル・データベース・テクノロジーです。Red Brick Warehouseは、低いTCO(トータル・コスト・オブ・オーナーシップ)でハイ・パフォーマンスの分析データベースとして揺るぎない地位を築きました。インターネット・ポータル大手Terra Lycos社のデータベース・オペレーション部門マネージャー、Frank Flynnは次のように語っています。「IBM Informix® Red Brick Warehouseで情報を統合することで、Terra Lycosのビジネス・インテリジェンス・アプリケーションの効果が増した結果、当社の戦略目標に沿って、より多くの情報による意思決定を行えるようになりました」

コスト面で有利なUNIX®やMicrosoft® Windows NT®のオペレーティング・システムと組み合わせさせたRed Brick Warehouseは、意思決定支援とデータ・ウェアハウスなどのビジネス・インテリジェンス・アプリケーションを類似のシステムよりはるかに安価に提供できます。Red Brick Warehouseのようなソリューションをご利用いただければ、規模の大小を問わず、あらゆる企業で強力なビジネス・インテリジェンス・アプリケーションを実現し、設備投資に対する大きな見返りを得ることができます。

データ・ウェアハウスへの入力

データ・ウェアハウスへのデータ入力(ロード処理)は、データ・ウェアハウス・アプリケーションを立ち上げて維持していくうえで、まず最初の非常に重要なプロセスです。ロード処理およびロード性能を最適化することで、ビジネス上の2つの大きなメリットが得られます。まず、ロード可能な時間帯に、データ品質を劣化させずに必要な全データをウェアハウスに取り込むことができます。次に、巡回性を伴うビジネス・ニーズにマッチするよう、データのロード頻度を多くすることができます。

高品質のデータだけをウェアハウスに入力するには、データのロード処理を以下のタスクで構成します。

入力ソースからのデータの読み出し、およびデータ・ウェアハウスへのデータの書き込み。

新データ・キーが追加される時点での内部インデックス構造の増補。

投入データにより参照整合性違反がないかチェック(例えば、部品xの売り上げが入力された場合、その部品xには部品の説明、ソース、コストなどが付与されていないと見なされます)。

必要に応じてEBCDICからASCIIへの変換、タイム・フォーマットなどの性能データの変換。

ウェアハウスへの新規データ投入時における、異なるデータ・サマリー・レベルまたは階層構造の自動維持。

ハイライト

複数のタスクを実行する1つのスクリプト

Red Brick Warehouseを使えば、上記のロード処理のタスクをすべて単一のステップとして実行できます。Red Brick Warehouseは、1つのスクリプトで以下のタスクを実行できるように設計されているのです。

- データベースへのデータのロード。
- 参照整合性を維持するためのレコード・チェック。
- インデックスとアグリゲートの作成。
- データの変換。
- 内部インデックス構造の増補。

こうした特長によって、ビジネス・インテリジェンス・アプリケーションの開発時間と、関連するオーバーヘッド・コストが削減されます。

優れたロード性能

ロード処理に加え、データ・ウェアハウス・アプリケーションを成功に導くうえで絶対に必要な条件はロード性能です。意思決定支援対応のデータベースを更新するには、データを頻繁に一括ロードしなければなりません。したがって膨大なデータを扱うことから、当然ながら高性能のデータ・ローダーが必要になります。IBMは下記タスクの実行効率の大幅アップを目指してデータ・ローディング機能を設計しました。

高速のデータ・ロード

- インデックス作成、参照整合性チェック、およびロードの並列処理。
- 新データをロードする時点でのインデックス作成。
- オプション機能として、参照整合性チェックで不適格となった入力データの自動行生成(不良データの調停時間を大幅に短縮)
- ディスクへの書き込みに先だって行われる、メモリー内データおよびインデックスのソート最適化(ウェアハウスにデータを格納させるためのディスク書き込みおよび読み出し回数を削減)
- データベースを連続的に使えるようにする、ブロック・レベルのバージョンング(高速でデータを一括ロードしているときでも、クエリーを実行)
- 新規データをロードする時点での複数インデックス作成(新規データに対して複数インデックスを並行作成)

ハイライト

Red Brick Warehouseのサイズが増大すれば、データのロードは効率化する

Red Brick Warehouseサーバーのシングルパス・ローダーは、入力データを読み取ると、内部のストレージ・フォーマットへのデータ変換、参照整合性チェック、インデックス生成、行レコードとインデックス・ノードからディスクの物理的ストレージ・ユニットへの書き込みを行います。ロード処理後、データベースはクエリー受付可能な状態となります。

HP/CompaqのAlphaServer GS320システムで稼働させたRed Brick Warehouseに関する共同研究では、Red Brick Warehouseはデータ・ロード、クエリー性能、スケーラビリティのいずれでも高得点を収めました。この共同研究は、最大1000人のユーザーを持つ2TB、4TB、8TBのデータ・ウェアハウスを測定対象として、Red Brick WarehouseのPOPS (proof of performance and scalability) ベンチマークを利用しています。全体的に見て、Red Brick Warehouseのサイズがスケールアップすると、ロード速度は向上することが分かります。このベンチマーク・テストの結果については、図Aのロード性能チャートを参照してください。

データ・ポイント	テーブル名	GB/時間	ロード行数	経過時間	行数/分
2TB	Daily sales	200.0	8,483M	10:17:38	13,735,651
4TB	Daily sales	229.8	17,060M	17:59:57	15,797,533
8TB	Daily sales	279.0	34,121M	29:39:27	19,172,922

図A. ロード性能ベンチマーク

顧客フィードバックの効率的な分析によって製品の改善を目指す

Red Brick Warehouseのロード性能を活用している例としては、ある企業のケースが挙げられます。この会社は、データ・ウェアハウスでRed Brick Warehouseを使って製品に関する顧客フィードバックを分析しています。Red Brick Warehouseが誇る高速のデータ・ロードとクエリー応答は、顧客データを参考にして製品開発と品質管理を進めるこの企業にとって重要な役割を果たしています。Red Brick Warehouseによって製品開発者は迅速にフィードバックを得られるので、より適切な対応が新しい製品や既存の製品で可能となります。

ハイライト

Red Brick Warehouseは、大量のデータを管理するために設計されている

ディスク・スペースの必要量が少なければ、コストは減る

大容量のデータ・ストレージ

データ・ウェアハウスのサイズは一般に10～100ギガバイト台で、医療や金融などの業界ではテラバイト台に近いデータもいくつかあります。データ・ウェアハウスのテーブル1つで数億のデータを格納することもできます。したがって大量のデータを効率よく格納する能力の有無は、データ・ウェアハウス・アプリケーションの運用コスト総額に大きな影響を及ぼします。

Red Brick Warehouseはデータ・ウェアハウス・アプリケーションに最適化されているため、非常に優れたデータ・ストレージ効率を利用することができます。IBM Red Brick Warehouseは圧縮インデックス構造を使用しており、インデックス格納に必要なディスク・スペースを最小限に抑えています。Red Brick Warehouseの最適化インデックス作成アルゴリズムは、初期インデックス作成時には全面的にブロック・スペースの割り当てを、その後は増分インデックスの更新を採用しています。Red Brick Warehouseは、数値データのストレージにコンパクトなバイナリー数値データ型を使用しています。

Red Brick Warehouseのユニークなインデックス構造とストレージ機能によってユーザーは、Red Brick Warehouseのデータ・ウェアハウスで必要となるディスク・スペースを最小限にすることができます。したがって、データ・ウェアハウスが大きくなると、ストレージ・サブシステムはコスト面で不利になりますが、Red Brick Warehouseはコスト効果が高いものとなります。詳細については、図Bを参照してください。この図は、8TBの生データを格納するのにRed Brick Warehouseが使うストレージの合計容量を示しています。

テーブル名	行数	データ・ストレージ	インデックス・ストレージ
Daily sales	34,121,091,680	8,271,782,144K	938,052,656K
Daily forecast	3,574,993,950	140,196,072K	38,511,552K
Total		8,411,978,216K	976,564,208K

図B. 8TBの生データで必要なストレージ容量

ハイライト

複雑なクエリーも処理可能

OLTP RDBMSテクノロジーを使用したデータ・ウェアハウス・アプリケーションでは、集約データが大きなネックとなりがちです。各エンド・ユーザーが詳細データにアクセスできるほどの集約データを用意することは、システム性能面で多大な犠牲を強いることとなります。データ・ウェアハウジングで役立つOLTP RDBMS製品は、集約レベルのデータ維持をサポートするために、多くの場合200%から300%もの余分なオーバーヘッドを必要とします。これに対し、IBM Red Brick Warehouseでは既存の詳細(生)データに対する複雑なクエリー処理の効率が基本的に高いため、集約データの格納・メンテナンスに必要なディスク・スペースおよびITリソースに関連するオーバーヘッドを飛躍的に減少させることができます。

クエリーへの迅速な応答

エンド・ユーザーの視点でとらえた場合、クエリー性能はデータ・ウェアハウスの利用容易性や有効性を最も大きく左右します。繰り返し行うクエリー・プロセスを維持するためには応答時間の短いことが要求されます。発見プロセスを辿っていくとき、アナリストや企業トップはビジネスの特定の定量的側面を「ドリルダウン」します。

スタースキーマ設計の利点

例えば、「北米市場において製品Xと製品Yとでは、本年度と昨年度で売り上げ状況はどう違うか」といった典型的な質問について考えてみましょう。この種の複雑なクエリーを処理する必要性から、リレーショナル・データベースの最適化アルゴリズムに対する要求は過大なものとなります。RDBMSは多くのテーブル間のジョインを処理しなければなりません。

Red Brick Warehouseは複雑なクエリー処理に最適化されており、これまでに複数テーブルのジョインが必要なクエリーに対して、予測可能な高速クエリー処理を確実に実行してきました。この機能は、クエリー中心の立場から見たデータを提供するスタースキーマに基づいています。スタースキーマでは、情報がファクトとディメンションという2種類に分類されます。ファクトは分析の対象となる中心的なデータ要素であり、ディメンションはファクトの属性です。

ハイライト

STARjoinおよびSTARindexテクノロジー

Red Brick Warehouseは、複雑なクエリーの処理を高速化するためにSTARjoinとSTARindexというテクノロジーを採用しています。STARjoinは、比類のないジョイン性能を特長とする、高速のシングルパス複数テーブル・ジョインです。STARindexは、ファクト・テーブルのディメンションとそのディメンションを含む行との関係を示す、高度に圧縮された情報であり、それによってジョイン性能が大幅に向上します。STARjoinアルゴリズムは、STARindexを使って、特定のジョインに必要な行のすべてを効率的に識別できます。データ・ウェアハウスの性能を低下させる要因としては、想定されないクエリーからさまざまなユーザーのロード、整合性の欠けたデータに至るまで、さまざまなものがあります。それでも、Red Brick Warehouseの広範囲に及ぶスタースキーマ処理ソリューションは、現実世界の問題に正面から取り組むことを可能にしてくれます。

予測可能でリニアに上がるクエリー性能は、データ・セットが大型化しユーザー数が増えるにつれ、企業のシステム要件の予測能力向上という恩恵をもたらします。Red Brick Warehouseでは、キャパシティー・プランニングが容易です。複数テーブルのジョイン試験では、IBM Red Brick Warehouseは複雑なクエリーに対する結果をコンスタントに1分未満の応答時間で返していますが、これはSTARindexとSTARjoinのクエリー高速化アルゴリズムによる成果です。

ビジネス・インテリジェンスを実現する、実績あるプラットフォーム

データ・ウェアハウジングで成功を収めたRed Brick Warehouse

世界中のほとんどの企業は、他社との競争で、洞察に富んだビジネス情報は有能な従業員や高品質の製品に比肩するほど強力な武器になるという認識をしています。こうした認識に支えられて、データ・ウェアハウジングなどの意思決定支援テクノロジーの人気は高まる一方です。あらゆるデータ・ウェアハウスの中核を成すRDBMSは、そうしたテクノロジーとして脚光を浴びています。データ・ウェアハウジングに特化したRDBMSに対してDBAが求めるものは主として、効率的なロード処理、ハイ・パフォーマンス、クエリーに対する高速の応答、そして大量データを格納する能力です。ビジネス・インテリジェンス・アプリケーション用の分析データベースであるRed Brick Warehouseがこうした条件を満たしていることは、一連のベンチマーク・テストで実証済みです。Red Brick Warehouseを導入した企業は、優れた操作性および投資効果を高める低いITCOによって、必要なハイ・パフォーマンスを得ることができます。

IBM, e(ロゴ)business software, Informix, Red Brickは、IBM Corporationの商標。
Microsoft, Windows NTは、Microsoft Corporationの米国およびその他の国における商標。
UNIXは、The Open Groupがライセンスしている米国およびその他の国における登録商標。
他の会社名、製品名、サービス名等は、それぞれ各社の商標または登録商標。



日本アイ・ビー・エム株式会社
〒106-8711 東京都港区六本木3-2-12
09-02 Printed in Japan

仕様は事前の予告なしに変更することがあります。製品、サービスなどの詳細については、弊社もしくはIBMビジネス・パートナーの営業担当員にご相談ください。



6588-2393-00