

## 赤レンガ倉庫6丁目2番地



### 第10回 ステップ2-DWHの構築、ステップ3-情報の分析

第10回目の今回は、ソリューション（テンプレート）を使った実装に関する後半の部分である、DWHの構築、分析について触れて行きたいと思います。その前に、今回で本連載はひとまず最終回となります。これまでなかなか皆様にご紹介する機会の少なかったRed Brick Warehouseの技術情報を、このように1年に渡りご紹介することができたのも、皆様からの励ましの言葉と、貴重なご意見が私たちの推進力になったと思っております。

6丁目2番地の全てのページをプリントアウトして会議に参加された方が複数の会社でいらしたとか、実際の構築でテンプレート情報が参考になったというようなお話をうかがうと、本当に1年間二人でやってきてよかったと思っております。今後は別の形にはなるとは思いますが、引き続きRed Brickの技術情報を提供できればと思っております。それでは最終回になりますが、後半の部分を始めさせていただきます。

#### ステップ2 DWHの構築

DWHの構築では、通信業務用に特化したソリューションテンプレートをベースとして使用します。まず最初にBIデザイン社が構築したDWHに使用したデータモデルのご紹介と、データベースエンジンとして採用しましたDWHに特化したデータベースであるRed Brick Warehouseの特徴的な機能をご紹介します。

#### スタースキーマ構成

BIデザイン社が業種に特化したテンプレートとして使用していますデータモデルを、スタースキーマ型モデル、あるいはディメンジョナル型モデルといいます。

#### コンテンツ

- ↓ [ステップ2 DWHの構築](#)
- ↓ [ステップ3 分析](#)
- ↓ [まとめ](#)

#### 関連リンク

- [IBM Red Brick Warehouse 製品ページ](#)

#### 執筆者

ソフトウェアサポート & サービス  
IMテクニカルサークルス  
堤 保晴

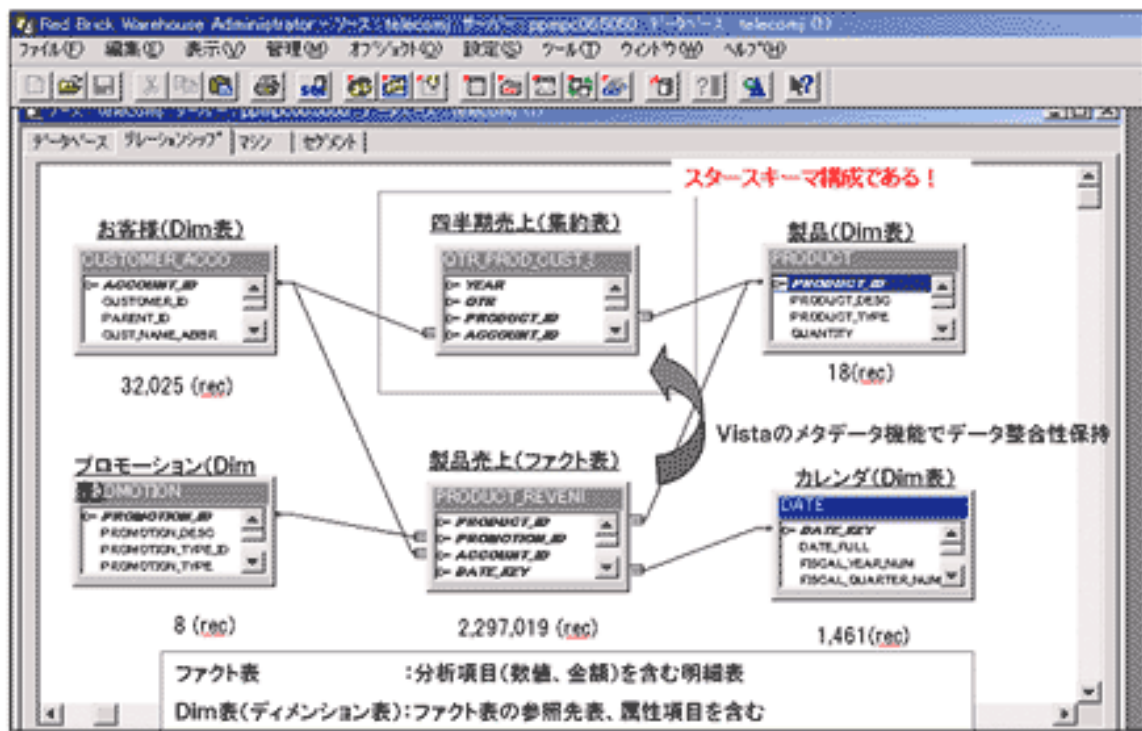


Red Brick Warehouseはこのスタースキーマモデルを実装することで、製品本来のパフォーマンスを引き出すことが可能となる製品です。スタースキーマモデルについて、図に沿って説明いたします。まず図の中心にある明細情報を管理するテーブルをファクトテーブルと言います。図の中では「製品売上」がそれに該当します。そのファクトテーブルの周りに配置されているのが、ディメンジョンテーブル、あるいはマスターテーブルと言います。「製品」「プロモーション」「お客様」「カレンダー」の各テーブルがそれに当たります。

ファクトテーブルにはディメンジョンテーブルに関連付けるキー項目と分析対象値を含み、ディメンジョンテーブルにはファクトテーブルを参照するためのキー項目を含みます。これはディメンジョンテーブルの主キーをファクトテーブルの外部キーとして定義し、テーブル同士を関連付けることを意味します。つまり、ファクトテーブルである「製品売上」テーブルには、製品の購入や利用によって得られた売上金額の明細情報が格納され、「製品」「プロモーション」「お客様」「カレンダー」などのディメンジョンテーブルには、分析対象とするデータを直感的に理解できるようなテーブル名と項目を定義します。

またこの図では集約テーブルとして、「四半期売上」が定義されています。この集約テーブルは四半期ごとの明細情報の集計結果がテーブルとして管理されており、Red BrickのVista機能により利用されます。このようにファクトテーブルを中心とした星型のデータ構造をスタースキーマ構造といます。

## データモデルの説明



Red Brick Administrator モデル参照画面  
[<クリックして拡大>](#)

## 集約テーブル管理

Red Brick Warehouseでは、システム自身に集約情報を管理する機能があり、Vistaという名で提供されています。

Vistaは事前計算された集約結果を使用することで、クエリー性能を向上させるための機能です。

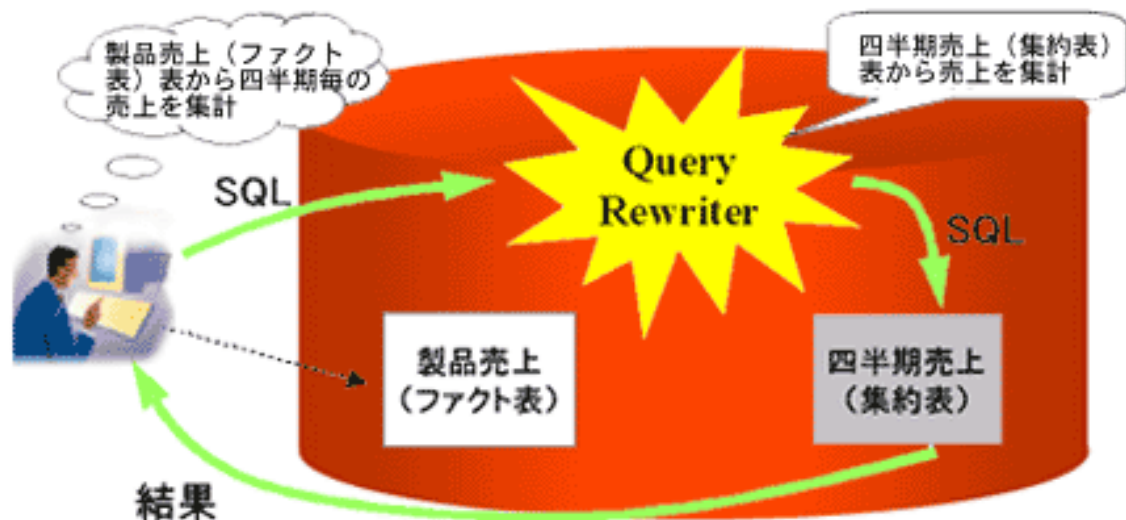
Vistaに関しても、図について説明いたします。

通常、ユーザーからの参照要求は点線矢印に沿って(Vistaを使用しない状況)、売上明細表にクエリーを行います。例えば、毎日のデータから四半期の平均売上を計算する場合、膨大な数の行をグループ化して計算する必要があり、その場合には、多くの時間とリソースを必要とします。そのような条件の場合、Vistaによる集約表の利用を検討します。集約表には製品売上金額を事前に計算した集約データを格納し、Vistaはユーザーからのクエリー要求をユーザーに意識させることなく、集約表へのクエリに書き換えを行います。集約表の更新タイミングは、データロードやデータの更

新処理（挿入、更新、削除）の際に自動的に行われます。また、Vistaの管理機能として、GUIによる管理ツールがあります。この管理ツールを使って、これまでのクエリーのログ情報から、より効果的なクエリーの書き換えを提案するアドバイザー機能も提供しています。

## Vista機能の説明

製品売上（ファクト表）のクエリーを、四半期売上（集約表）へのクエリーに置換し、ユーザーに意識させることなく 高速に実行  
アドバイザー機能により、作成すべきVista構成を提案



## Star Index, Star Joinについて

IBM Red Brick Warehouse は製品のリリース当初から、スタースキーマ構造をシステムのアーキテクチャーとして採用していました。このデータ構造は、ER図のようなトランザクションDBが採用するデータ構造に比べて、テーブル間の関係が直感的に理解でき、ユーザが求める結果を得るための参照が簡単にできるという利点があります。Red Brickは、このデータ構造の上に独自のインデックス、ジョイン技術を提供することで、分析系システムの要求課題である高速なクエリーを実現しています。ここではOLTPのデータベースと比較して、内部処理なテーブルジョインの違いについて説明します。

### OLTPのデータベースの場合

OLTPのデータベースで複数のテーブルをジョインさせて結果を求めるような参照要求の場合には、pair-wise Joinという方式でジョインが行われます。このジョインでは2つのテーブルをジョインさせて、該当結果を一時テーブルに収めます。

OLAPデータベースは、この処理を結果を得るためにジョインさせる

テーブルの全てに対して行い、その最終結果をユーザーに返すこととなります。

pair-wise Joinの場合、ジョイン対象となるテーブルの数マイナス1の一時テーブルが作成され、数分のテーブルに対するI/Oが発生します。

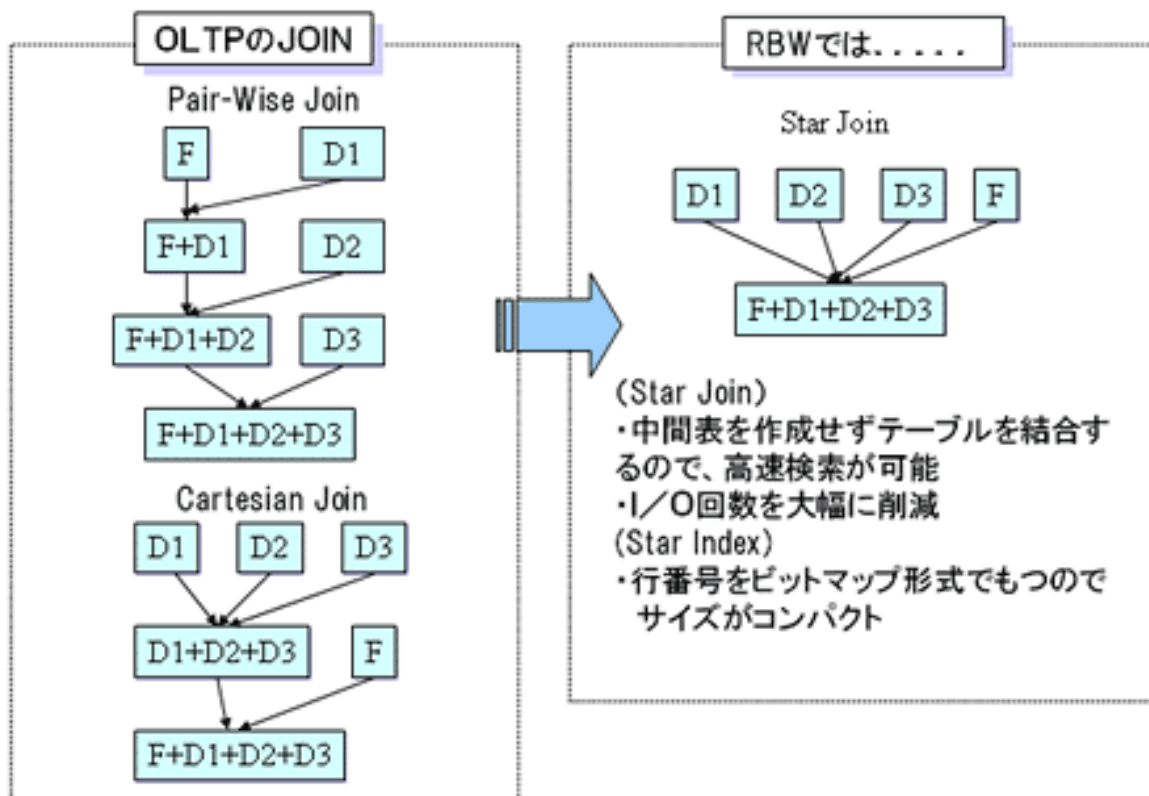
この場合、参照対象のテーブルが多くなればなるほど、多くのサーバーリソースが必要となり、期待した時間での参照が不可能になる傾向があります。ただ最近では多くの製品で改善が見られ、カーテージョインという方法によるジョインが行われてきています。この方法は、2つのテーブルをジョインさせながら、結果を絞り込んで行くのではなく、ディメンジョンテーブルにあるすべての該当データを1つの一時テーブルとして作成し、その一時テーブルとファクトテーブルをジョインさせて結果を得る方法です。この方法ですと、複数の一時表を作成する必要がないため、多くのリソースやI/Oを必要としない参照が可能となります。

#### Red Brickの場合

Red Brickでは、Starスキーマ構造のデータ構造の上にStarIndexを使って、該当するテーブル間のジョインを一度に定義します。このジョイン定義をStar Joinといい、Red Brickによるテーブル参照の基本的な機能として、提供されています。

Star Joinにより複数のテーブルをジョインさせるために定義するIndexがStarIndexで、作成されたIndexはシステムにより管理されます。Red Brickからの参照要求は、あらかじめ該当するテーブルがすべてジョインさせた結果からなるIndexであるStarIndexに対して検索を行います。この場合には、参照要求のあった全てのテーブルに対して直接1回のパスで参照が行われ、結果を返すことが可能です。そのことは、OLAPデータベースのような中間表の作成を必要とすることなく、またパフォーマンスへの最大障害であるディスクI/Oに関しても軽減する形での参照が行なわれることを意味します。このことがRed Brickの高速参照を実現している理由です。従って、Red Brickを使った高速クエリーを実現するには、事前のStarIndexの設定によるStarJoinの実装が必要であることがわかんと思います。また、Red Brickはあらゆる傾向のデータに高速処理を実現するために、StarIndex以外にも複数のユニークなindexを用意しています。Target Indexもその一つで、粒度の荒いデータを参照する場合には、該当する項目に対して定義することで、効果を上げることができずし、StarIndexと一緒に使用することで、より効果的なクエリーパフォーマンスを実現することが可能となります。

## Star Joinの説明



## ステップ3 分析

それでは、次にステップ3による分析についてお話しいたします。

前述までの内容からRed Brickはデータ分析に特化した高速なりレシヨナルデータベースであることはおわかりいただけたと思います。ただ、実際に業務等で分析を行う場合には、Red Brickにアクセスするクライアントのアプリケーションが必要になります。簡単な分析ですとExcelのシートを使ってデータを表示したり、グラフ化して資料化したするのが一般的ですが、マーケティング部門などが専門に分析をするような場合には、分析専用のソフトウェアを使う必要があります。今回は分析専用のソフトウェアであるBrio Queryを使用して作成した分析画面を見ながら、いくつかの例を示して情報の分析を行ってみたいと思います。

まず、下記の図について説明いたします。この図は、Brio Queryを使用して作成した画面で、分析者が、今期の売り上げについての傾向を分析するために必要な情報を1画面として作成しています。このように、Brio Queryのような、分析用アプリケーションを使うことで、分析者が必要な分析結果をわかりやすくまとめて表示することが簡単に可能になります。

ここでご紹介する分析画面は大きく2つのWindowから構成されています。左側のWindowには分析対象条件のメニューを、右側のWindowには分析者がメニューごとに、分析対象データをグラフ化して表示しています。



次の画面は、分析用のクライアントアプリケーションであるBusinessObjectsにより、分析結果をグラフ化して表示しています。BrioQueryで使用したIBM Red Brick Warehouseに存在するデータを利用して、製品別売り上げトップ10、製品収益率、各サービスごとの収益率を棒グラフ化して表示しています。

この例では、各サービスごとの収益率を示す棒グラフがあります。製品全体の収益率をよりブレイクダウンさせた形の内容で、今期の収益の内訳から、収益を上げている製品とそうでない製品の特定が行えます。たとえばこの結果を元に今期以降の各サービスのプロモーション戦略を立て直すことも可能になります。

このように、分析用アプリケーションを使用することで、全体的な評価をする場合にデータをグラフ化することで、データへの分析を直感的に理解することができ、また表示されたグラフを評価して、より具体的な詳細情報を簡単に入手することが可能となります。

Red Brickを分析エンジンのデータベースとして使うことで、分析者が求める、あらゆる分析条件に敏速に対応できるソリューションを提供することが可能になります。

## 分析結果の検証 ~ BusinessObjects編 ~



BusinessObjectsアプリケーション画面

## [<クリックして拡大>](#)

### まとめ

このシリーズの最後は、ソリューション（テンプレート）の実装について9～10の2回にわたり、お話しさせていただきました。リレーショナルデータベースでありながら、DWHの世界で多くのアドバンテージと利用実績を上げることができる製品であるRed Brick Warehouse は、その専門特化した製品戦略の確かな方向性と、ソフトウェア技術に裏打ちされた製品クオリティの高さにより、多岐多様の業種による実績と、PC環境から数十テラバイト環境までの幅広いシステム環境での運用を可能にしています。また、オープンなユーザインターフェイスによる高速なデータアクセスは、分析用アプリケーションの複雑な要求にも敏速に対応できる仕組みを幾つも備えています。ユーザ要求に応じて適切な機能を使うことで、分析者が必要とする情報を最適な形でご提供することが可能な製品であることを、今回のシリーズの中で、ご理解いただけたと思います。

Red Brick Warehouseの最新版であるバージョン6.3を今年の上半期に、リリースを予定しています。より精選されたパフォーマンス技術を追加実装しており、複雑に変化拡張する市場要求に対応できる製品クオリティを維持し続けていることを表明しています。まもなく皆様にもその機能をご紹介する機会があると思いますので、ご期待ください。それでは長い間、ご購入いただきありがとうございました。