

# データ更新を伴う Operational OLAP

大場 達生

(株) ビーコン インフォメーション テクノロジー

東京都渋谷区恵比寿南1-6-10恵比寿 MFビル14号館

[oba@beacon-it.co.jp](mailto:oba@beacon-it.co.jp)

データウェアハウスは一般に、基幹系システムやオペレーショナルデータストアから必要なデータを抽出し、これをロードすることにより構築される。データウェアハウス内のデータには恒常性が求められ、データウェアハウスの代表的な利用形態である OLAP はふつうデータウェアハウスに参照アクセスを行う。本稿ではデータウェアハウスに更新を含めたアクセスを行う、Operational OLAP について考察する。

## Operational OLAP with Write-Back

Tatsuo Oba

Beacon Information Technology Inc.

Ebisu MF Bldg No.14, 1-6-10 Ebisu Minami, Shibuya-ku, Tokyo

[oba@beacon-it.co.jp](mailto:oba@beacon-it.co.jp)

Data warehouse in general is built by extracting data from the mission critical systems and loading the data. The data stored in data warehouse must be 'Non-volatile', and the data access method of OLAP is read-only OLAP, a typical way to access data warehouse, this paper focuses Operational OLAP that allows updating the data in data warehouse.

### 1. はじめに

従来の基幹システムに蓄積されたデータを活用することに主眼をおいた「データウェアハウス」の概念が唱えられてから久しい。現在までに多くの企業や組織でデータウェアハウスの構築が進んでいる。その多くは、基幹系・業務系のシステムに蓄積されたデータを、データの活用を目指すかたちで抽出、また必要に応じて加工し、エンドユーザに提供するという形態である。一般に、エンドユーザはデータウェアハウスにより提供されたデータを、業務上必要とされる資料などの作成に、つまりレポートに使用している。ここでは、データウェアハウスの準備されたデータには参照のアクセスしか必要とされない。

しかしながら、データウェアハウスの普及、またそれに伴ったエンドユーザのデータ活用の気運の高まりにつれて、データウェアハウス内のデー

タ更新を伴う新しい利用形態の概念が生まれている。

本稿では、従来までのデータウェアハウスの概念と、データの更新を伴う新しい利用形態を、データマート、OLAP の側面から比較し、データ更新を伴う Operational OLAP という概念について考察する。

### 2. データウェアハウスの定義

データウェアハウスという概念の提唱者である W.H.Inmon は、データウェアハウスの定義として、次の4つの特徴を定めている\*1。

- ・ サブジェクト指向
- ・ 統合化
- ・ 時系列
- ・ 恒常性

「サブジェクト指向」については、いわゆる基幹

系システムにおける「アプリケーション指向」とは対照的に位置づけられるものである。基幹系システムは、たとえば受発注や購買、予約の受け付けなど、アプリケーションに依存するデータを扱う。これに対しデータウェアハウスは、そもそもデータの有効活用を実現するために生まれたものであり、ここで扱われるデータは、一般に意思決定支援システムなどで用いられるための顧客の注文や管理会計、製品毎の販売実績といった類の利用目的別のデータである。

「統合化」は、データウェアハウス内に管理されるデータについて一貫性を保証するものである。先の「サブジェクト指向」に関連するが、データウェアハウス構築のもととなるソースデータは基幹系システム内のデータである。ここでは「アプリケーション指向」でデータが管理されているため、たとえば会計システムにおいては金額が円単位で管理されているのに、販売管理システムでは千円単位ということがある。両システムからデータを抽出してデータウェアハウスを構築するとき、金額の単位の統一しなければならない。ほかにコード体系の相違などもしばしば存在し、データウェアハウスの構築にあたっては、「データクレンジング」と呼ばれるデータの統合化作業が必須とされる。

「時系列」は、データウェアハウスに管理されるデータがある時点でのみ正確であるということの意味している。基幹系システムはトランザクションデータを処理するが、ここではユーザがアクセスしたまさにその瞬間のデータは正しい。つまりその瞬間のデータは現在値であり正確である。データウェアハウスには、基幹系システムで管理されるデータがスナップショットで移される。したがって、スナップショット時点のデータはその時点で現在値であり正確であるが、トランザクション処理を行わないため、常に現在値を持つわけでは

ない。また、基幹系システムでは、トランザクション処理の効率を高めるために、可能な限り取り扱うデータ量を小さくするという指向から、データを短い範囲に限定する。一方でデータウェアハウスでは、一定間隔、つまり日、週、月という単位で基幹系システムからのスナップショットデータを取り込み、長期間の履歴をもつように管理する。これは、データウェアハウスがそもそも意思決定支援システムなどの代表されるようなデータの活用、分析のために構築され、利用されることからきているものである。データを活用、分析するためには時系列で管理されていることが望ましい。

また、この定義のなかでは、「データウェアハウスのデータは一度記録されると更新されることがない」とされている。

「恒常性」は、データウェアハウスに管理されるデータの特性である。データウェアハウス内のデータについては、2種類の操作、つまり「データの取り込み」と「データへのアクセス」しかなく、データウェアハウスには「更新」はないとされている。

### 3. データの恒常性

一般に、データウェアハウスは次の手順で構築・運用されることが多い。

- ① 基幹系システムからのデータの抽出
- ② データの統合化(データクレンジング)
- ③ データウェアハウス(のデータベース)へのデータロード
- ④ データウェアハウス内のデータ参照

①は、データウェアハウスを構築するためのデータを、それが蓄積されている基幹系システムから抽出する作業を示す。

②は、①で抽出されたデータの統合化作業を意味する。データウェアハウスにロードするデー

データは、ふつう基幹系システム内の多種のアプリケーションにわたって存在しており、データに一貫性がない。このため、データウェアハウスへのロード前に、データの統合を行う。

③は②で準備されたデータを、データウェアハウス内のデータベースにロードする作業である。この段階でデータウェアハウス内に利用可能なデータが存在することになる。

④は、構築されたデータウェアハウスのデータにエンドユーザがアクセスすることを意味する。

データウェアハウスはこのような手順で、構築され、またこの繰り返しが運用手順と考えられる。これはデータウェアハウスの4つの定義に忠実にしたがった構築手法であり、ここではデータの流れは一方通行である。データウェアハウス内のデータは、常に基幹系システムなどの既存のシステムから抽出されたものであり、エンドユーザからのアクセスは参照のみである。したがって、データウェアハウス内のデータの恒常性は保証される。

さて、一般にはデータウェアハウスは上記の手順で構築・運用されるものであるが、最近の応用事例では、必ずしもデータウェアハウス構築・運用のなかでデータの流れが一方通行であるというわけではない。

#### 4. OLAPとデータマート

データウェアハウスは、そもそもデータの活用、有効利用を目的に提唱されたものであるが、実際にデータウェアハウスを構築したとしても、エンドユーザが提供されたデータを本当に活用できるかどうかは疑問である。その原因は次のように考えられる。

エンドユーザはコンピュータシステムの専門家ではない。したがって、基幹系システムから「サブジェクト指向」の考えに基づいてエンドユーザが必要とする目的別のデータをデータウェアハウス

として準備しても、例えばそこにリレーショナルデータベースによるデータ管理があった場合、エンドユーザは必要とするデータを得るには、SQLなどの問い合わせ言語を学ばねばならない。つまり「操作性」がデータ活用を実現する課題となる。また、「サブジェクト指向」に基づいたデータウェアハウスであっても、刻々と変化し、かつ多様なエンドユーザの要求には柔軟に対応することが困難である。このような場合には、統合的なデータウェアハウスよりも、「サブジェクト指向」をさらに進めた「目的別のデータ利用環境」が求められる。

このような課題について、先の「操作性」を解決するのがOLAPであり、また「目的別のデータ利用環境」を実現するのがデータマートである。データマートは小規模なデータウェアハウスと捉えることができる。データマートは、単にエンドユーザの多様な要求に対する小規模データウェアハウスという位置づけだけでなく、中規模以上のデータウェアハウスで表面化する、「構築のための所要時間が許容範囲に収まらない」という問題や、エンドユーザの要求に対する「レスポンス時間の悪化」という問題を解決するアプローチとしても利用される。

#### 5. OLAPアプリケーション

データウェアハウスの構築・運用については、これまでも述べているとおり、エンドユーザがデータを有効活用できる環境を実現することが主目的であり、一般的にはここにOLAPという利用形態が採り入れられている。OLAPは、E.F.Coddが当初12のルール<sup>\*2</sup>をもって提唱した概念であり、エンドユーザのデータ活用における操作性を重視したものであって、データウェアハウスと密接な関係をもつ。

データウェアハウス内のデータを利用する

OLAP アプリケーションには、2つの形態が存在する。

参照型アプリケーションは、前述のとおり基幹システムから抽出されたデータを、データウェアハウス(またはデータマート)内にロードし、このデータにアクセスするものであるが、アクセスは参照のみにとどまる。一般にほとんどの OLAP アプリケーションはこの形態で利用されており、「販売実績」や「管理会計」、「顧客の商品購買履歴」といったようなデータを参照するような事例が実現されている。エンドユーザの利用形態は、帳票や会議資料の作成、リポーティングであり、また EIS(役員情報システム)や DSS(意思決定支援システム)としての分析作業である。

対話型アプリケーションは、エンドユーザがデータウェアハウス(またはデータマート)内のデータに、参照だけではなく更新を含めたアクセスを行うものである。実現されている事例としては、「予算編成」、「販売計画」、「コストの配賦」などがある。このようなアプリケーションの構築には、多次元データベースのアプローチが有効であり、OLAP による実現に向いている。エンドユーザは、準備された実績データを参照し、これをもとに計画データを入力する。あるいは事前に準備された計画あるいは予測値に修正を加える。これらの手続きは、アクセスするデータウェアハウス(またはデータマート)内のデータを更新するものであり、ここではデータウェアハウスの恒常性という考え方はない。

## 6. Operational OLAP の実現

データウェアハウス(またはデータマート)の利用形態である OLAP について、とくにデータの更新を伴う利用形態を Operational OLAP と定義する。

Operational OLAP は OLAP アプリケーシ

ョンのなかでも対話型アプリケーションでの利用形態である。

対話型アプリケーションでは、事前にデータウェアハウス(またはデータマート)に準備されたデータを参照するのみでなく、これらに修正を加えたり、または新たにデータを入力するといった更新処理が伴う。

OLAP を実現する多次元データベース製品の代表的なものである「Essbase(エスペース)<sup>※3</sup>」は、Operational OLAP を実現するための機能として必須とされる「データの送付(ライトバック)」、また「動的計算」の2つを提供している。以下、Operational OLAP について、Essbase を例に考察を行う。

「データの送付」については、そもそも対話型アプリケーションではアクセスするデータウェアハウス(またはデータマート)内のデータに更新を行うものであり、これを実現するものである。Essbase では、エンドユーザが利用するフロントエンドとして Microsoft Excel や Lotus 1-2-3 で利用可能な「スプレッドシートアドイン」を提供している。これらの利用により、エンドユーザはスプレッドシート上で、Essbase データベースで実現されるデータウェアハウス(またはデータマート)からデータを取得(参照)することが可能である。これに加え、取得したデータをスプレッドシート上で修正したり、あるいは新しくデータを入力したうえで、「データの送付」機能を利用すれば、アクセスしている Essbase データベース内のデータを更新(ライトバック)することが可能である。

「動的計算」は、「データの送付」機能でデータベースに書き戻された(更新された)データをもとに別のデータを動的に計算(導出)するものである。

Essbase は多次元データベースを管理する DBMS であり、データベースは次元から構成さ

れ、次元は階層から構成される。「データの送付」機能を用いて、例えばある営業担当者の売上予算データを更新した場合、その営業担当者が属する部門の集計値、また部門が属する全社合計の集計値も変化しなければならない。Essbase をはじめとする多次元データベースでは、データの更新や入力に伴い、このような集計値を計算する「計算処理」を実行しなければ、正しい集計値を得ることができない。

また、データベース内に「利益」という項目があり、これは「売上」から「経費」を差し引いて導出されるものであるとする。この場合についてもやはり「売上」データを更新すると、結果的に「利益」データも変化しなければならないため、やはりここで「計算処理」を実行する必要がある。

このように多次元データベースでは、データの更新に伴って「計算処理」を実行することにより、初めて多次元データベース内でのデータの一意性が保証される。したがって、データの更新後には必ず計算処理を実行する必要がある。言い換えれば、エンドユーザはデータの更新を実行すると、計算処理を実行しなければならない。つまり、Operational OLAP を実現するには、「データの送付」と「計算処理の実行」の2つの処理を手続き的に実行する必要がある。また、計算結果を確認するために再度データの取得を行う必要がある。

Essbase がもつ「動的計算」機能は、エンドユーザによるデータの送付の後、再度データの取得を要求すると、その時点で動的(自動的に)に計算を行うものであり、これを利用することによりエンドユーザは明示的に「計算処理」を実行することなく、計算結果を瞬時に得ることが可能となる。

一般に明示的に実行する計算処理は、とくに指定を行わない限りデータベース全体を対象に計算(集計や導出項目の計算)処理を実行するた

め、処理時間はデータベースサイズに依存する。つまりデータベースの規模が大きいと、この処理には数分から数時間という範囲の時間を要求される。これに対し「動的計算」は、エンドユーザが取得を要求する範囲のみを計算対象とする。ふつうエンドユーザが業務で利用するデータは一時にスプレッドシート1枚分程度であり、多次元データベース全体から考えるとごく僅かである。したがって計算対象となるデータは限定されているので、「動的計算」は瞬時に計算結果をエンドユーザに返す。

「動的計算」機能がない多次元データベースでは、計算処理に長時間を要するため、やむを得ず計算処理の終了を待ったり、夜間バッチ処理での対応が要求される。よって、近年のスピードが要求される経営環境のもとでは、データの更新が必要な対話型アプリケーションの構築には不向きである。逆に考えれば、こうした対話型アプリケーションでは、「動的計算」機能がもつのも重要であり、Operational OLAP に必須とされるものである。

また、E.F.Codd の提唱した OLAP の12のルールには「マルチユーザサポート」も含まれる。データの更新が伴う Operational OLAP では排他制御を含む「マルチユーザサポート」が求められる(Essbase は対応している)。

## 7. まとめ

多次元データベース製品 Essbase は、「データの送付」、および「動的計算」機能を提供することにより、新しい OLAP アプリケーションの可能性を広げてきている。これまでの OLAP アプリケーションは、従来のデータウェアハウスの一概念である「恒常性」をそのまま継承し、基幹系システムから抽出されたデータを、利用目的別にデータウェアハウス、あるいはデータマートとして管理

し、エンドユーザはそれに対し参照のみのアクセスをするという参照型アプリケーションの利用がほとんどであった。こうした利用が普及する一方で、「データの送付(ライトバック)」機能を利用する対話型アプリケーションが、企業・組織内の計画業務に直結する「予算編成」、「販売計画」や「コストの配賦」といったアプリケーションで実現されてきている。また、ここに「動的計算」という機能が付加されたことにより、リアルタイムにデータを更新することが可能となっている。本稿では、データウェアハウス、あるいはデータマート内のデータを更新する形態の OLAP について、とくに Operational OLAP という概念を定義し、「データの送付」と「動的計算」の2つの機能が必須であるととした。Operational OLAP では、データウェアハウス、データマート内のデータを更新する。従来の考えでは、ここに管理されるデータには「恒常性」が求められていたが、Operational OLAP では「恒常性」はない。ここでは、基幹システムから抽出され、ロードされたデータを、今度は逆にエンドユーザが更新・修正するという形態となる。また、最近の事例では、Operational OLAP で直接データウェアハウス内のデータを更新しないものの、限定したデータマート内のデータを更新・修正し、その結果をデータウェアハウスや基幹システムに反映するというデータの流れも現れている。

対話型アプリケーションの登場は、Essbase などのどちらかといえばデータウェアハウスというよりもむしろ OLAP を実現するための製品が、「データ送付」機能を持っていて、これを応用する形で実現されはじめたものである。そもそも、おおもとを辿ればデータウェアハウスから切り離されたデータマートで実現されていたものであり、データの更新は限定され、閉ざされた環境のなかのみで行われていた。しかし最近になって、「動的

計算」のような製品の機能強化、またスピード経営が重視される経営環境とあわせて、対話型アプリケーションで作成される企業内において重要な計画値、予測値などのデータ(数字)を、データウェアハウスに反映、あるいは逆に基幹システムに取り込むようなデータの流れが検討されてきている。

このように Operational OLAP で実現されるアプリケーションでは、データウェアハウスのデータの恒常性は必要とされない。そもそもデータウェアハウスという概念の根底をなすものは、「データの有効活用」である。現在の OLAP アプリケーションの実現範囲を見てみると、Operational OLAP のように、準備されたデータを活用して新たなデータを作成する業務が見受けられるが、これこそが真の「データの有効活用」であり、このときデータウェアハウス内のデータの恒常性を保つ必要性は考えられない。また、従来の一方通行のデータの流いでデータウェアハウスを構築するセオリーはもはや唯一の構築手段ではない。今後、Operational OLAP による計画や予測の作成、シミュレーションといったアプリケーションが普及すると、そこにはデータウェアハウスやデータマートを実現する OLAP ではなく、基幹業務に利用される OLAP という世界も見えてくるのではないだろうか。

#### 【参考文献・その他】

1. W.H.Inmon & Richard D.Hackathorn, "Using the Data Warehouse", John Wiley & Sons International, 1994 (藤本康秀監訳『よくわかるデータウェアハウス活用』、インターナショナル・トムソン・パブリッシング・ジャパン、1996年)
2. 石井義興『データ・ウェアハウス』日本経営科学研究所、1995年
3. Hyperion Solutions Corporation 開発の多次元データベース製品