

多次元データベースとRDB（OLAPの紹介）

2P-1

豊島一政

(株) ソフトウェア・エージ OLAP事業部

1. はじめに

OLAP (On-Line Analytical Processing) という言葉が最初に登場したのは、1993年に発表されたE. F. Codd, S. B. Codd, C. T. Salley共著による論文である。そこには、今日のように多岐にわたる問題を解決するにはRDBMSだけでは不十分である、との記述がある。さらに、E. F. Coddは、RDBMS開発の本来の目的は1960年代終りのシステムに存在する問題を解決し、DBMSをより魅力的なものにすることにあつた、と記述している。RDBMSに欠けている機能を提供するのがOLAPコンセプトであり、そのベースになっている技術が多次元データベースである。ここでは、RDBMSの限界を明らかにし、それを補完する多次元データベースの特性を紹介する。

2. マネージメント・レベルの問題

OLAPが得意とするのは、企業のマネージメント・レベルに生じる問題である。マネージメントの実際の問題を小売業の例で説明することにする。あるお店に一人の客がやってきた。店の中の商品をあれこれ見ていたが、しばらくして店員に声をかけた。「このセータが気に入ったんだけど、私には少し小さいようだ。もう一回り大きなサイズがありますか。それに、もっと明るい色が欲しいのだが。」店員は、早速パソコンに向かって何やら検索を始める。「お客様。本店のほうに大きいサイズがございます。明日にでも取寄せることができます。しかし、色の方はこれだけです。いかがいたしましょうか。」店員が行った検索の作業はマネージメントの問題ではなく、現場の問題、すなわちオペレーショナルな問題である。それは、指定されたサイズと色の商品があるかないかを確かめる問題であり、限られた範囲のデータが対象になっている。このような問題はRDBMSに適している。

しかるに、マネージメントの問題はもっと包括的である。どの商品の売行きが良く、どの商品の売行きが芳しくないのか。また、今年はどうな色のものが良く売れているのか。どの商品の在庫が問題であるのか・・・などなどである。この例で分かるように、マネージメントの問題には明らかな検索条件のようなものは存在しない。また、一つの疑問が別の疑問を呼ぶ。思考過程は連続的に変化してゆく。結局、データ全体が対象となる。データは様々な形で集計され、比較される。このような問題を解く過程がOLAPであり、多次元データベースの支援範囲ということになる。では、なぜ、このような問題をRDBMSで解くことが困難なのだろうか。それを考えてみる。

Multi-Dimensional Data base v.s. RDB (Introduction of OLAP)

Kazumasa Toyoshima

OLAP Division, Software A.G.

Shinjyuku L Tower, 1-6-1, Nishi-Shinjyuku, Shinjyuku-ku, Tokyo 160, Japan

3. RDBMSがOLAPに適さないわけ

3-1. RDBMSは多次元関係を効率良く管理できない

- 2次元構造** RDBMSは本来2次元の構造を持ったものである。多次元を表現しようとする、データを互いに関連する複数のテーブルで管理することになる。ビルの3次元モデルを多数の平面図に描くのに似ている。そのビル全体のイメージをこれらの平面図から想像するのが大変なことは容易に理解できる。
- SQLの難しさ** 多くのテーブルの内容や関連をエンドユーザが正しく理解し、それをSQLで記述することは難しい。
- パフォーマンスは決定的に悪化する** 複数のテーブルを結合するためのジョインと呼ばれるオペレーションは、パフォーマンスを低下させ、レスポンスタイムに影響を与える。
- 直積型では資源を食う** そこで考えられる直積型（予めジョインされたテーブルをもつ）では、スペースを無駄づかいし、結局パフォーマンスを低下させる。

3-2. 分析機能が限られているRDBMS

- 機能上の制約** SQLでは、レコードの同じフィールドをSUMしたり、AVGしたりはできるが、一つのテーブルの一つのレコード内のいくつかのフィールドや、いくつかの列をわたっては同じオペレーションができない。
- フロント・エンド・ツール** 結果として、フロント・エンド側で多くのプログラムを書かなくてはならない。

3-3. データを簡単に集計できないRDBMS

RDBMSはロールアップやドリルダウンが容易ではないため、連続した思考過程を支援できない。

4. 多次元データベースの特性

多次元データベースは、既述された問題を解決する仕組みを提供している。すなわち、データを多次元構造の下で、圧縮した形で管理している。それによって生まれる特性は次の様である。

- 集約** データの集約とはデータをまとめて、集計する機能をいう。企業の組織や製品の体系などは、階層構造で表現されるのが普通である。どのようなレベルの階層、また、その組み合わせでも容易に集計できるのが多次元データベースの特性の一つである。
- ドリルダウン** データの分析を行う場合、まず、会社全体を眺め、部門や営業所別・商品別に展開してゆくような手順を取ることが多い。レベルの上から下へ掘り下げてゆく分析手法をドリルダウンという。
- スライス・アンド・ダイス** データを分析するとき、様々な視点から切込んでゆく。売上げを商品別や地域別、顧客層別・・・といったふうに分析するような場合である。ちょうど、サイコロを転がすのに似ている。これをダイシングといい、現れた面を一枚一枚開いてゆくのをスライシングという。多次元データベースにおけるキーワードがこのスライス・アンド・ダイスである。

5. 結論

巨大化する企業データ、複雑化する企業データを的確に分析するには、場面に合った分析ツールを選択する必要がある。その意味で、多次元データベースはRDBMSを補完しているといえる。