

DWH 向けデータベースにおける JDBC ドライバの実装

森川修一，秩父かおり，高山茂伸，八木綾子

三菱電機株式会社

1. はじめに

データウェアハウス(DWH)の利用目的が高度化し、企業内での適用範囲が拡大している。従来の部門レベルにおける業務分析・意思決定支援システムから、企業全体のシステム統合化を目的とした情報管理基盤として、DWH の位置付けが変わりつつある。業務データを全社的に統合することにより、経営状況の可視化や事業上の問題検知、付加価値創造など、活きた業務情報に基づいて経営活動を行おうというものである。

このような DWH の役割の変化に伴い、DWH システムでは基幹系システムとの相互運用性や BI ツールとの接続性、データのリアルタイム連携などの機能に対する必要性が増している。三菱 BI サーバ DIAPRISM においても接続性・相互運用性の強化を実施しており、その一環として JDBC ドライバの開発と提供を行っている。本稿では DIAPRISM の JDBC ドライバ実現方式と、DWH での利用における JDBC 機能に関する一考察について述べる。

2. DIAPRISM

DIAPRISM は集計・検索処理の高速性を特長とする DWH 向けデータベースエンジンである。データは大福帳と呼ばれる表形式フラットファイルに格納される。データベース処理には SQL に相当する DIAPRISM 用のデータベース操作言語 AQL を使用する。AQL は DWH 向け操作言語として、次のような特徴を持つ。

- DML としては SELECT のみ提供。
更新 (INSERT/DELETE/UPDATE) は提供せず。
- SELECT は SQL92 準拠。
DWH に対応した拡張機能あり。
ROLLUP, CUBE 機能
SQL99 の OLAP 機能の一部に相当
カテゴリ機能
データの階層構造の定義

3. アプリケーション インタフェース

DIAPRISM はデータベースエンジン部分のサーバ OS として Windows、UNIX(Solaris, HP-UX)に対応している。アプリケーション開発用のプラットフォームは、次の理由により Windows 環境のみを対象としていた。

- 主な利用形態はアプリケーション/パッケージなどによる単独の DWH システム。
- 企業クライアントの主流は Windows 環境。

しかし DWH 適用の拡大化に伴い、要求されるシステム構成が広がっている。XML/Web サービスに対する関心も高い。アプリケーションサーバとの連携の要求もあり、DIAPRISM では J2EE 及び Microsoft .NET Framework への対応を視野に入れている。この第一歩として Java 対応を目的とした JDBC ドライバの開発を行った。

Windows 環境では、データアクセス用のアプリケーションインタフェース(API)として、用途に応じた 2 つの API (DIAPRISM 専用 COM コンポーネント及び ODBC ドライバ)を提供している。

表 1. データアクセス用 API

	専用 COM	ODBC ドライバ
機能	DIAPRISM 全機能	標準的な SQL, ODBC 機能範囲
用途	専用パッケージ/アプリケーション	一般 BI 製品

JDBC ドライバを使用した汎用 BI 製品がまだ普及していないことから、DIAPRISM 専用 EJB などによる対応も考えられたが、

- DIAPRISM 固有のスキルを要求しない
 - 既存のフレームワークが流用可能
- という点から JDBC ドライバの開発を行った。

4. JDBC ドライバ実現方式

JDBC ドライバはネイティブプロトコルによる Type4 の Pure Java ドライバとして実現している。DIAPRISM サーバとはソケット(java.net)により通信を行う。ソケットを用いたアクセスパスは、既に ODBC ドライバにおいて確立されており、これを利用している。ODBC ドライバと同一のプロトコルを使用することにより、サーバ側の開発量を削減している。

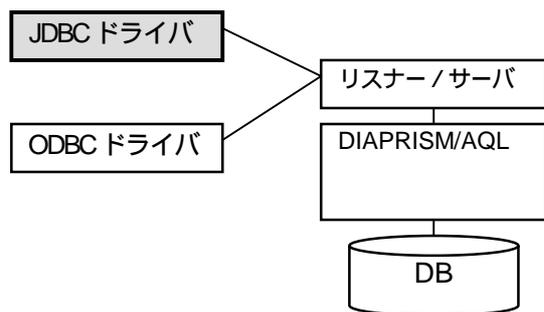


図 1. 構成図

JDBC ドライバが提供する機能は JDBC 3.0 仕様を基にしている。今回の開発では JDBC のコア API (java.sql)、オプションパッケージ API (javax.sql) のうち、コア API について実現している。

表 2. サポート API

インタフェース	サポート
Array	×
Blob	×
CallableStatement	
Clob	×
Connection	
DatabaseMetaData	
Driver	
ParameterMetaData	
PreparedStatement	
Ref	×
ResultSet	
ResultSetMetaData	
Savepoint	×
SQLData	×
SQLInput	×
SQLOutput	×
Statement	
Struct	×

DIAPRISM には更新操作がないため、トランザクションに関連した Savepoint は未サポートとなっている。また、DIAPRISM がないデータ型のインタフェースについても未サポートである。

オプションパッケージ API は、DataSource やコネクションプールなど機能的に有用である。しかし、性能やスケーラビリティを確保するための仕組みが必要であり、検討課題としている。

5. DWH 用 API としての JDBC

JDBC を DIAPRISM 用の API としたとき、DWH 向けの拡張機能に対する操作性が重要となる。

ROLLUP, CUBE 機能

GROUP BY 句への ROLLUP、CUBE 指定は、特に問題はない。ROLLUP、CUBE 指定による結果セットの操作に関しては、次の課題が挙げられる。

- ・ 集計レベル単位のデータ取得
結果セットの全データを取得した後、アプリケーションが選別する必要がある。
- ・ 集計行の判定

SELECT リストに TAG 関数 (GROUPING 関数相当) を追加し、その値で判断する必要がある。

ROLLUP、CUBE 機能により集計レベル毎の SELECT 文実行が抑止できるものの、ドリルダウン/ロールアップ操作をユーザに提供するには、集計レベル単位でデータを操作する必要がある。この操作はアプリケーション側でそれぞれ作成しなければならない。これは DIAPRISM に限らず、SQL99 の OLAP 機能を利用する上でも同様の課題であると思われる。

このような DWH の分析でよく用いられる操作に関しては、操作インタフェースを共通化し、JDBC ドライバ側で提供することで、アプリケーション開発の効率化と、DB 処理の効率化が図れる。例えば ResultSet か RowSet のサブクラスを用意して、次のような機能を提供する方法が考えられる。

- ・ 集計レベルに応じた結果セットの生成
- ・ 集計レベルの切替 (ドリルダウン/ロールアップ)
- ・ 集計行の判定

カテゴリ機能

カテゴリをデータ検索で使用する場合、SELECT 文にカテゴリ関数を記述することで行う。カテゴリ機能という点では、その内容が結果セットに反映されるのみであり、特に問題はない。

カテゴリ情報の取得に関しては、カテゴリ用のメタデータ取得インタフェースは JDBC 仕様にはない。しかし、システムテーブルを公開することで、システムテーブルの検索により情報取得は可能になる。

6. おわりに

DIAPRISM の接続性強化の一環として JDBC ドライバを実装した際の実現方式と、DWH として利用する場合の JDBC の課題について述べた。