

DWH 構築におけるメタデータの移行について

佐久間孝夫, 森川修一, 金田典久

三菱電機(株) 情報技術総合研究所

1. はじめに

業務データから DWH (データウェアハウス) を構築するためには ETL (Extract、Transform and Load) 機能を持ったツールが使用される。我々は自社で開発した ETL ツールを保持しているが、フラットファイルだけでなく様々なデータソースに対応するために、また、いろいろな操作を行うために高機能な ETL ツールが要求されていた。そこで、他社で開発した ETL ツールに移行することとした。そのため、ETL ツールで定義したメタデータ (テーブル定義情報、列定義情報、ジョインなどの処理定義情報など) も移行しなければならなかった。このメタデータの移行をスムーズに行うために移行ツールを開発した。

本稿では、今回、我々が開発した移行ツールの移行方式について述べる。

2. メタデータを移行した ETL ツール

自社で開発した ETL ツール DIAPRISM/DSF のメタデータを Informatica 社の PowerMart という ETL ツールに移行する。それぞれの ETL ツールは次の項目以降で述べるメタデータを保持している。

2.1 DIAPRISM/DSF

DIAPRISM/DSF は自社で開発した ETL ツールであり、連携図を図 1 に示す。DIAPRISM/DSF は独自フォーマットのメタデータとして以下の定義情報を持っている。

- ・ ソース定義 (列定義情報)
- ・ ビルド定義 (テーブル定義情報、ジョインや列の切出し等の処理定義情報)

これらの定義情報を元にフラットファイルからデータを読み込み DIAPRISM/AQL (大福帳データベース) にデータをロードする。

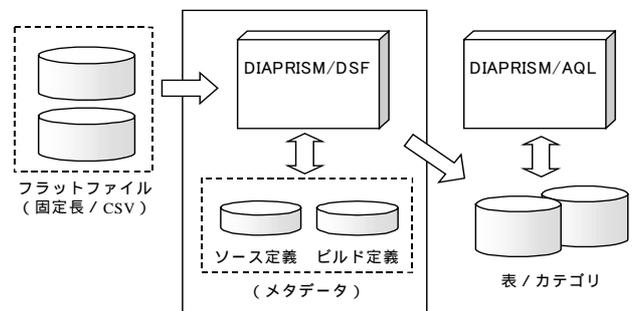


図 1 DIAPRISM/DSF 連携図

2.2 PowerMart

PowerMart は Informatica 社が開発した ETL ツールであり、連携図を図 2 に示す。PowerMart は XML (Extensible Markup Language) を使用してメタデータとして以下の定義情報を持っている。

- ・ マッピング定義 (列定義情報、テーブル定義情報、ジョインや列の切出し等の処理定義情報)
- ・ ワークフロー定義 (マッピング定義で定義した処理の動作を決定する変数定義)

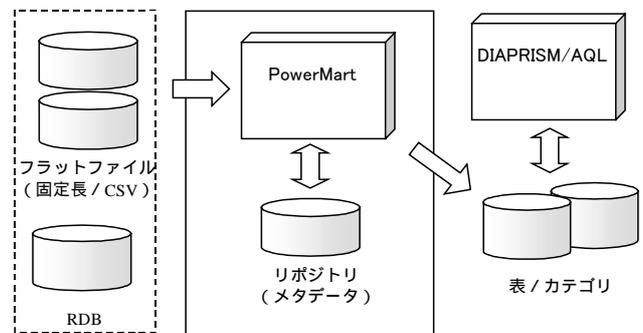


図 2 PowerMart 連携図

The migration of meta data for DWH
Takao Sakuma,

Shuichi Morikawa and Norihisa Kaneda,
Mitsubishi Electric Corporation,

Information Technology R&D Center
5-1-1 Ofuna, Kamakura,

Kanagawa, 247-8501, Japan

これらのメタデータを元に各種のデータソースから DIAPRISM/AQL にデータをロードする。

3. メタデータ移行時の課題

メタデータの移行時には、以下の 2 つの課題があった。

- (1) 定義情報の内容が異なるため、1対1の変換ができない。
- (2) 内部表現が異なるので変換が複雑になる。

課題(1)では ETL ツールが異なるために、ジョインするためには DIAPRISM/DSF ではビルド定義として 2 行の定義情報で済んだものが PowerMart ではマッピング定義の複数の定義として 10 数行必要になる。

課題(2)ではデータの内部表現形式が異なるために移行する場合に型変換の処理を追加するための定義を加えなければならない。

4. 移行ツール

上記課題を解決するために移行ツールは図 3 に示す以下の機能を有している。

- (1) メタデータを解析する機能
- (2) 共通定義情報
- (3) XML ファイルとしてメタデータを生成する機能

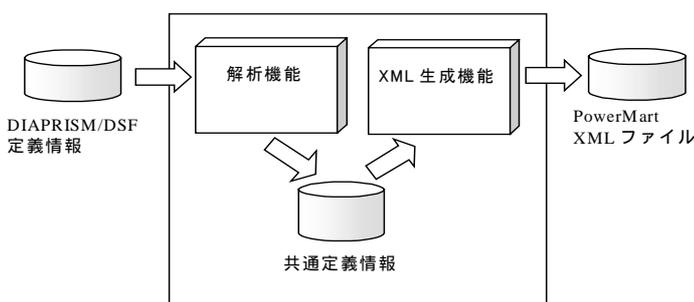


図 3 移行ツール

4.1 メタデータを解析する機能

メタデータを解析する機能は

DIAPRISM/DSF の持つメタデータ(定義情報)を解析するものである。複数のファイルで構成されているので、複数のファイルからメタデータを読み込み、整合性をチェックし、必要な情報を取り込む。まず、列定義情報を読み込む。次にテーブル定義情報を読み込み、最後にジョインや切り出し列などの処理定義情報を読み込む。これらの読み込んだ情報を共通定義情報として構造体に取り込む。

4.2 共通定義情報

共通定義情報は汎用化されたデータ構造を持っている。列定義情報、テーブル定義情報、処理定義情報などを持っている。

4.3 XML ファイルとしてメタデータを生成する機能

共通定義情報を元に XML ファイルを作成する。XML ファイルには列情報、マッピング情報、ワークフロー情報などを記述する。まず列情報を記述する。次にマッピング定義情報を記述し、最後にワークフロー定義情報を記述する。マッピング定義情報では、列情報のソース定義、ターゲット定義、ジョインなどの処理を指定するトランスフォーメーション定義を行う。ワークフロー定義では、マッピング定義を実行するためのセッション定義と動作状態を決定する変数の定義を行う。

5. 結果

移行ツールにより、DIAPRISM/DSF のメタデータを PowerMart に移行することができた。DIAPRISM/DSF のメタデータを読み込み、共通定義情報に変換することにより、スムーズにメタデータを移行することができた。

6. おわりに

本稿では、DWH を構築するために使用する ETL ツールのメタデータを移行するために開発した移行ツールの方式について説明した。