

## DB統合のためのスキーマ変換

3V-6

黒坂正則

草加学

石塚匡哉

NTT情報通信研究所

## 1.はじめに

通信網管理業務には伝送路管理、バス管理、回線管理等の網階層別の管理業務及び公衆網管理、専用線網管理等のサービス別の管理業務等、多種多様な業務が存在する。従来、通信網管理のためのデータベース(DB)はこれらの業務目的別に構築され運用されてきた。各DBは使用されているDBMSが異なるだけでなく、各々運用する部署が異なるために、お互いに重複したデータが格納されており、著しい場合には、同一の設備を示すデータの表現形式が異なる等の問題がある。近年の複雑化かつ高度化した通信網管理業務では、各部署が自己の保有するデータのみで業務を遂行することは困難となり、ほかの部署のデータが必要になってきているため、各システムの持つDBを相互に参照し、連携する必要性が生じてきた。

本稿は、異なる目的及び手段で構築された、通信網管理業務DB群を集約し、散在するデータを一つのプラットフォーム上に展開するDB統合処理の一過程である、スキーマ変換の考え方を示す。

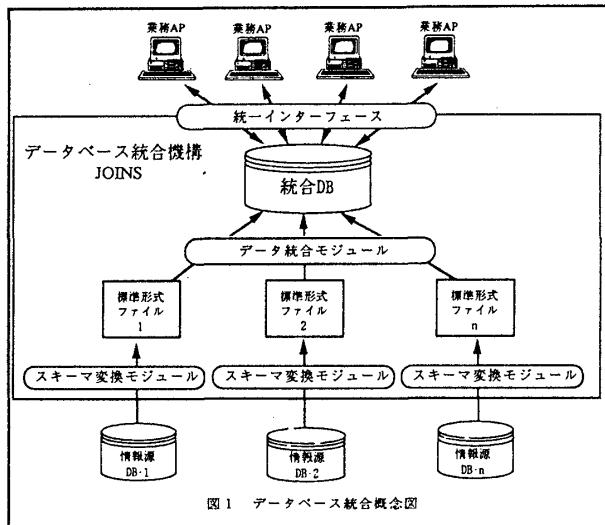
## 2.DB統合の条件

複数のDBを一つのRDBMSに統合するとき、異なるアーキテクチャで構築されているので、情報源から直接に統合処理を実施するのは困難である。従って既存DB群のデータ構造、表現形式の差異を吸収するために、統一した標準形式ファイルに変換する。この標準形式ファイルは統合処理の最終結果である統合DBと、名称及び表現形式は同じである。図1にDB統合の概念図を示す。

## 3.スキーマ変換処理

## (1)機種/OS/DBMSの違いの吸収

情報源のDBから共通プラットフォーム上に流通させるとき、入力としてDBMSのファイルを直接取得せず、各DBMSからブレーンテキスト形式で出力させたダンプファイルを使用することで対応することができる。



## (2)データの表現形式の違いの吸収

各データ項目の型式（数値、文字列等）、長さ、カラム値を各システムに固有のものから、共通の表現形式へ変換する必要がある。実際の変換処理は変換モジュールのロジックで実現できるが、カラム値の変換については頻繁にコードの追加修正があるので、簡単に対応できるように変換テーブルを用いて行う。

## (3)データモデルの違いの吸収

図2及び図3は通信網管理のための用途の異なる2種類のDBのER図であり、システムAとシステムBはともにディジタル伝送路網のデータを格納するDBである。

図4はシステムBの構造を基本に各情報源システムの違いを吸収し、統一したデータモデルである。AのPRIVATEPATH,D2STAGEPATH,D1STAGEPATH,PUBLICPATHの4つのテーブルとBのPASUに格納されていたディジタルバスの実体情報は標準形式モデルではPATHへ格納する。AのHANDLINGGROUPEの実体情報はデータの属性がディジタルバスであるのでPATHに集約する。また、HANDLINGGROUPEをPATHに集約したことによりCIRCUITPATHに格納されていた関係情報はPATHPATHまたはCIRCUITPATHへ分割する。

データモデルの変更は情報源から1レコードづつ

The Schema Transformation for Database Integration

Masanori Kurosaka, Manabu Kusaka, Masachika Ishizuka  
NTT Information and Communication Systems Laboratories

3-9-11 Midori-cho Musashino-shi Tokyo 180, Japan

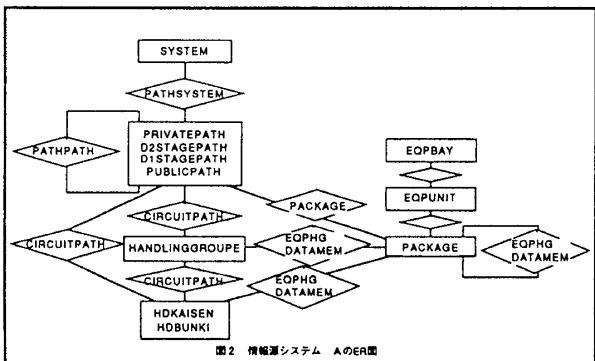


図2 情報源システム A のER図

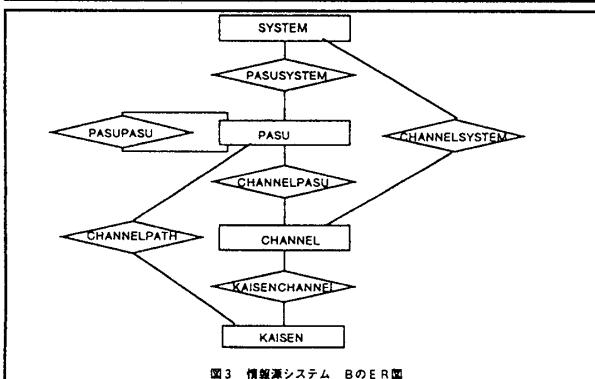


図3 情報源システム B のER図

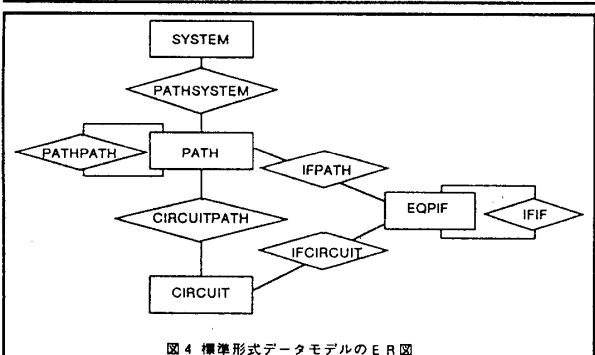


図4 標準形式データモデルのER図

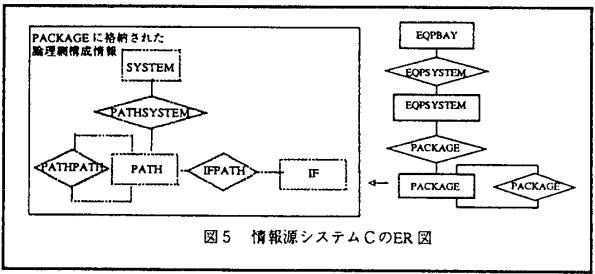
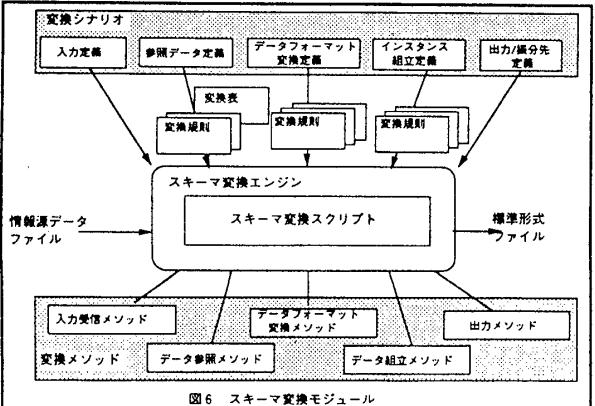


図5 情報源システム C のER図



データを読み出し、あるカラムに格納されているデータの属性を判断して、統一モデルにマッチするように振り分ける必要がある。

#### (4)隠されたインスタンス

図5は設備管理に使用されているDBのER図である。このDBは装置情報を格納しており、その構成情報であるEPQBAY,EQPSYSTEM,PACKAGEの3テーブルをもつ。

図5の左側に点線で示した実体、関係レコードは装置に収用されるディジタルバスの情報であり、PACKAGEの中に装置情報とともに格納されている。これらの隠されたディジタルバス情報は情報源の1レコード(PACKAGE)から、複数の実体(IFとPATHまたはIFとSYSTEM)を作成する。同様にもPACKAGE同士の接続を判断してPATHとPATH、PATHとSYSTEMの関係情報をPATHPATH、PATHSYSTEMを作成する事ができる。

#### (5)重複して格納されているインスタンス

情報源の1レコード内に複数のエンティティを格納しているDBがある。例えば、実体情報とその実体に関する関係情報が同一のレコードに格納されていて関係情報が独立のレコードとして管理されていないDBである。図2のPACKAGEは装置の実体情報及び、装置とディジタルバスとの関係情報を同一のテーブルに格納している。その場合、複数の関係を持つ実体情報は1テーブル内の複数のレコードに重複して格納されており、このテーブルから実体情報に関する情報を抽出すると、同じ内容のデータができるので、重複データは削除する。

#### 4.スキーマ変換モジュール

図5にスキーマ変換モジュールの概要を示す。情報源DBから出力されたテキストファイルを入力として、共通プラットフォームとなるDBMSに直接コードが可能な標準形式ファイルを出力する。

#### 5.まとめ

実際の通信網管理業務DBを情報源として、変換処理を実行したところ、DB統合処理が実行可能な標準形式ファイルを作成可能であることが確認できた。しかし、実際の統合処理には、各情報源の精度や統合のタイミングなど、まだ解決しなければならない課題がある。