

## 通信網管理情報のデータ統合機構

木原 民雄 中川 純一 山中 康史 寺中 勝美

NTT情報通信網研究所

通信網管理情報は、複数の管理システムのデータベースに蓄積され利用されている。これらの様々な異種性を持つデータを統合して他のデータベースに流通させるための機構を提案し、その構成法について明らかにした。

データ統合処理を行うために、ユーザは統合シナリオを記述する。統合シナリオが、予め用意された統合のためのメソッドを呼び出すデータ統合エンジンを駆動する。現実 to 同一の実体を表現している複数のインスタンスが同時に存在するパターンによって、統合処理の場合分けを自動的に行う。これにより、複数のインスタンスの待ち合わせをしたり、指定条件によってデータ項目の値を選択して統合インスタンスを生成することが容易にできる。

## A Mechanism to Integrate Network Management Data

Tamio Kihara, Junichi Nakagawa, Yasushi Yamanaka, Katsumi Teranaka

NTT Network Information Systems Laboratory

E-mail: kihara@nttops.min.ntt.jp

In actual network management system it has various information models and keeping style. It is necessary to combine heterogeneous information. A newly proposed mechanism translates, delivers and integrates network management data of different databases. This mechanism has the scenarios and the integration methods that are prepared function modules. These scenarios call integration methods to drive the data integration engine. In the scenario, integration processes are classified into existence pattern of instances. It is possible to wait for the lack of instance, and to generate new instance from every data item.

## 1. はじめに

通信網の管理情報は、複数の管理システムのデータベースに蓄積され利用されている。通信網は、新しい設備や管理技術が段階的に導入され、徐々に発展してきている。管理システムは、それぞれの管理業務の範囲やその分担の仕方や、通信網が構築されてきた経緯によって、管理情報の蓄積形態やデータ構造が互いに異なっている場合がある。

より高度で巨視的な通信網管理を行うために、これらの様々な異種性を持つ情報を組み合わせて利用したいという要求が高まっている。このためには、分散して配置された既存の複数のデータベースのデータを統合しなければならない。自律的に運用されている複数のデータベースの更新を、自動的に他のデータベースに反映する機構が必要であり、このとき現実に同一の実体を表現している複数のインスタンスはひとつに結合したい。

このために通信網管理情報を変換し、流通させ、統合する機構を提案し、その構成法を示す。

## 2. 解決すべき課題とデータ統合の方針

データ統合の基本方針を示す。データベース間のデータ流通は、データ変換処理とデータ統合処理に

よって段階的に行う(図1)。ここでは、各管理システムが、繁忙時間外に流通先データベースに必要なデータをファイル転送で流通させる形態を想定している。

流通されるデータは以下のような状況にある。

- (1) それぞれの管理システムでデータ表現などが異なる。
- (2) 複数の管理システムに、現実には同一の実体を表現するインスタンスが重複して保持されている。
- (3) 更新データを集信するタイミングにばらつきがある。

まず、データ変換処理によって、各流通元データベース間の実体や項目の名称、データ項目の値表現、構造の異種性を解消する。データベース間のデータ流通を実現するデータ流通システムが提案されており<sup>[3, 7]</sup>、そのデータ変換機能を利用し、データ項目の対応づけ、コード変換、構造変換などを行い、標準形式のファイルを生成する。この標準形式は、通信網管理の対象となる管理実体ごとに、そのインスタンスを一意に識別し、実体間の関連づけを特定できるように定義したものである。このとき現実に同一の実体を表現するインスタンスは、同一の識別項目(キー)を持つようにそれぞれデータ変換を行う。

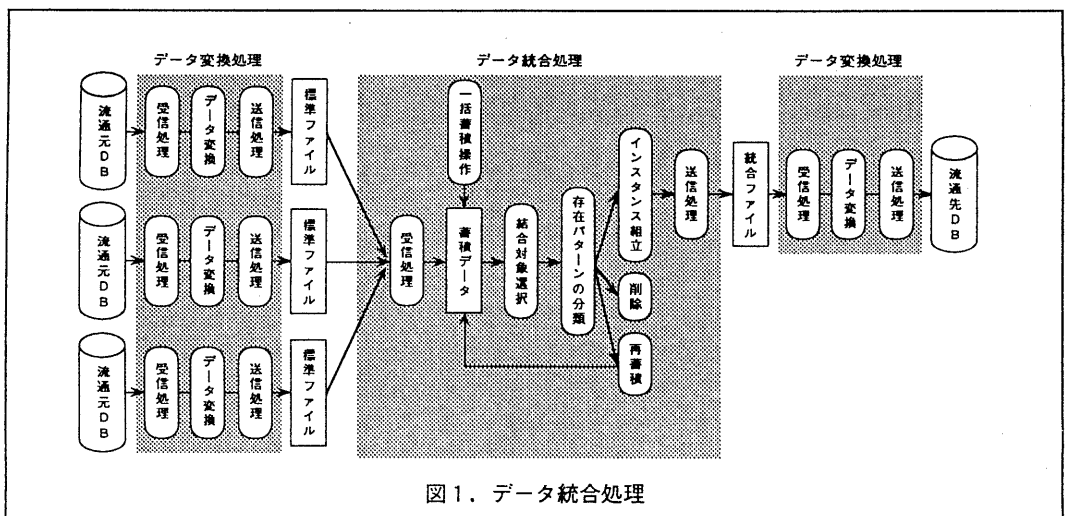


図1. データ統合処理

データ統合処理は、現実同一の実体を表現する複数のインスタンスが、ある時間内に統合機構に到着することを前提としている。このとき、次のような問題がある。

(1) データ統合機構は、流通元及び流通先データベースの全体を検索することはできない。流通してきたデータを受け取るだけなので、統合対象となるデータが揃うのをいつまで待たばよいか決めなければならない。

(2) 現実同一の実体を表現するインスタンスが複数あったとき、そのデータ結合の方法は、各データベース間の管理業務の範囲やその分担仕方等の知識がなければ決めることができない。

(3) 各インスタンスで重複したデータ項目があったとき、値の不一致や、値が不明なことがある。

キーの値が同じ複数のインスタンスは、可能な限り一つのインスタンスに結合して送り出したい。データ統合処理を何回か試行するうちに、ある結合すべき複数のインスタンスが揃うタイミングがあるはずである。

### 3. データ統合に必要な機能

データ統合機構によって以下の機能を実現する。

(1) 各データベースから流通してくる現実同一の実体を表現する複数のインスタンスの時間差を吸収する。その統合処理でインスタンスの結合ができなかったものは、再蓄積して、次回以降の統合処理の試行を待つことができる。統合すべきインスタンスが全て揃うのを待つのが原則だが、待ってもデータが来なかった場合の処理が記述できる。待ち状態として許容する条件を記述することができ、時間切れによる判定と、そのデータに対する統合試行回数

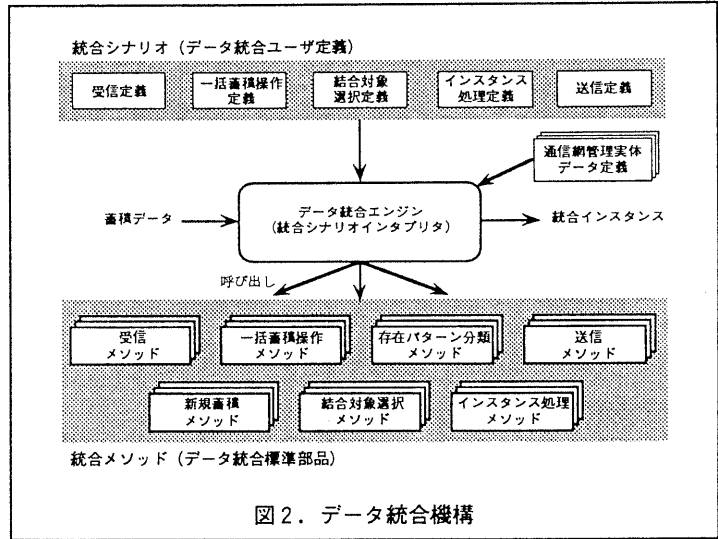


図2. データ統合機構

による判定などができる。

(2) キーによるデータ結合を行い、あるキーに対するインスタンスの存在/非存在を調べ、インスタンスの存在パターンによる分類を自動的にを行い、その存在パターン毎にひとつずつのインスタンスについて処理を行うことができる。統合シナリオを記述するユーザが、インスタンスの存在パターンの場合分けによって処理パターンを分けて記述することができれば、より容易な統合処理を実現することが可能となる。例えば、統合インスタンスに必要なインスタンスがないパターンのとき、一括して再蓄積するというような処理が簡単に記述できる。

(3) 同一のキーを持つ各インスタンスからデータ項目を選択し、統合インスタンスを生成する。各インスタンスで重複したデータ項目がある場合は、そのデータ項目の値に依存した選択条件を指定することができる。

### 4. データ統合機構の構成

データ統合に必要な機能を実現するために、次のような特徴を持つデータ統合機構を提案する。

(1) データ統合のための機能が部品化されている。これらの部品を統合メソッドと呼び、基本的な機能を組み合わせて呼び出すことによって統合処理の

各ステップを実施できるようにする。

(2) ユーザが記述する統合シナリオに従って統合処理をする。統合シナリオは、統合内容についての宣言的記述がされている。これにより多様なデータ統合を可能にし、新規のデータ統合に関わるユーザの負担を軽減させ、統合内容の変更も容易にできる。

データ統合機構では、データ統合エンジンが統合シナリオを逐次解釈し、予め用意された統合メソッドを呼び出し、駆動する(図2)。

統合シナリオについては、記述言語を新たに作成した。この記述言語において、if文などの制御構造、代入文、統合メソッドの関数呼び出しを行うことによって統合処理を実現する。

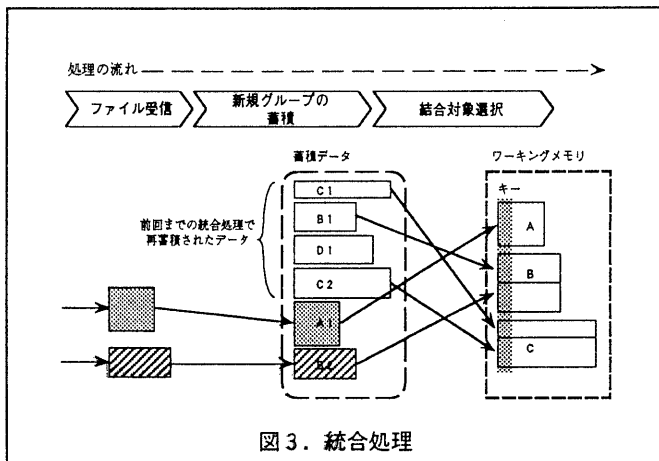
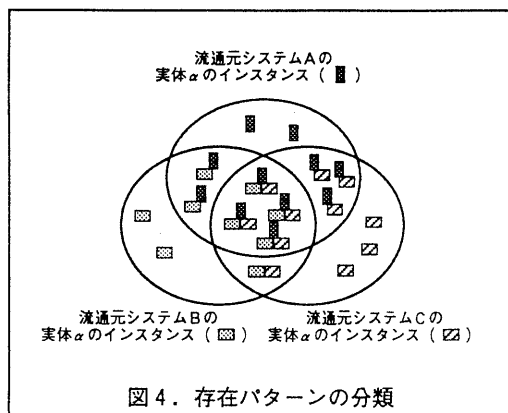
データ項目定義などのメタデータは、ユーザによって用意されているものとする。

## 5. データ統合処理

データ統合処理は以下の通り行う。各処理内容は統合シナリオの記述に従って行われる。

### (1) 受信処理

統合対象となるデータをファイルとして受信する。



### (2) 新規グループの蓄積

受信したファイルは、ある統合シナリオの統合対象となるデータとして、蓄積ファイルに蓄積される(図3)。蓄積ファイルには、前回までの統合処理の試行で再蓄積された複数世代のデータが合わせて蓄積されている。1つの蓄積ファイルに複数のグループのインスタンスが蓄積されている。グループとは、流通元データベースからある時点でファイル転送されてきたデータを、実体の種類ごとにまとめたものである。統合処理においてはこのグループを一括して処理する。このまとまりに対して、統合のための管理属性を持たせて、統合処理中に参照/更新するようにしている。その管理属性は、以下の通り。

- ・統合シナリオ名、蓄積ファイル名
- ・流通元データベース名、流通先データベース名
- ・実体名
- ・流通元配信日付(流通元データベースからこのデータが配信された年月日時刻)
- ・統合試行回数(このグループのインスタンスに対して統合処理が行われた回数)

### (3) 一括蓄積操作

蓄積データに対して、一括蓄積操作を行う。例えば、一定の期間を経過してしまったグループのデータの一括削除などができる。

### (4) 結合対象選択

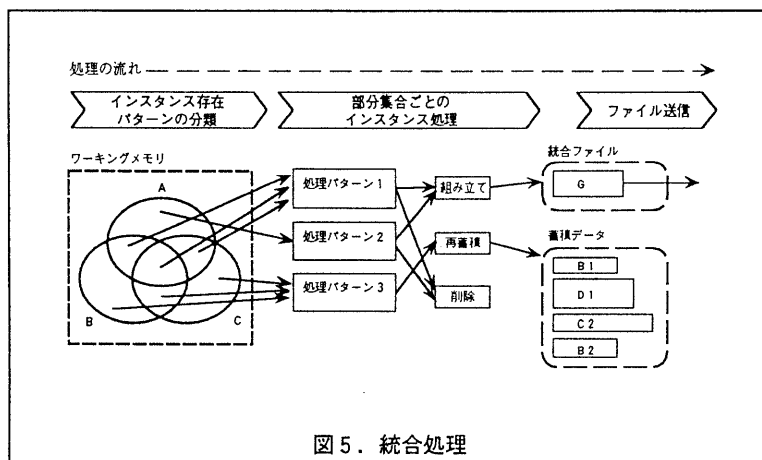


図5. 統合処理

結合対象となるデータを選択する(図3)。蓄積ファイルの中から、指定された流通元データベースのある実体のインスタンスを複数のグループにわたって全て読み込み、指定された識別項目をキーとしてハッシュし、ワーキングメモリに展開する。結合したい複数の流通元データベースの実体についてこの処理を行う。

(5) インスタンス存在パターンの分類

結合対象として選択されたデータについて、各流通元データベースのある実体の同一キーのインスタンスが同時に存在するかしないかのパターンによって分類する(図4)。この分類処理は、ユーザが統合シナリオで明示的に記述する必要がない。結合対象を選択した結果として、ワーキングメモリ上で自動的に行われる。

(6) 部分集合ごとのインスタンス処理

インスタンスの存在パターンによって分類したインスタンスの集合に対して、部分集合ごとに違う処理パターンのインスタンス処理を行う(図5)。統合シナリオにおける部分集合の指定の仕方は、流通元データベース名の組み合わせなどで行う。指定のない存在パターンについてはデ

フォルト処理を用意する。各処理パターンでは、その部分集合の同一のキーの複数のインスタンスを同時にfetchし、順次処理をする。ひとつひとつのインスタンス処理には、次の3つがある。

・再蓄積

今回の統合処理で再度、結合対象として選択されるようにインスタンスを蓄積ファイルへ再蓄積する。

・削除

受信待ちとして許容できなかった場合や、データの誤りが発見された場合に、統合インスタンスの生成を諦めて蓄積ファイルから該当インスタンスを削除する。ログに保存し、ユーザへ結果を通知する。

・インスタンス組み立て

同一キーのインスタンスの各データ項目の値を選択し統合インスタンスを生成する(図6)。

各々の処理は、統合シナリオ中の条件文と統合メソッドによって、場合分けして実行される。インスタンス組み立てにおいて、データ項目間の単純な一括コピーのような処理は、一括代入のための統合メソッドを呼び出せば良い。

データ項目の値を選択するとき、そのとき存在した複数のインスタンスで値が異なることも考えられるが、統合メソッドによって次のような条件を指定できる。

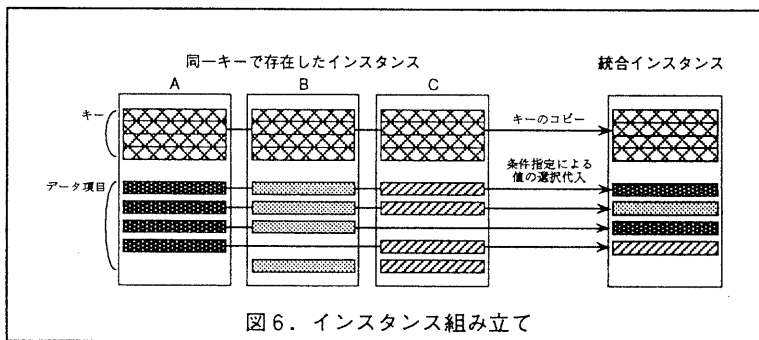


図6. インスタンス組み立て

- ・あるデータ項目の値を参照してその結果優先順を決める。
- ・配信日時がより新しいものを選ぶ。
- ・予め指定された流通元データベースごとの優先順に従う。
- ・データ項目の値で多数決。

このとき、複数の条件がある場合、それらの間には適用順がある。統合インスタンスのデータ項目の値が得られない場合のデフォルト値の代入も条件指定する。

何度か統合処理の試行して、最も望ましいインスタンスの存在パターンにならなかった場合、その時点で揃っているインスタンスから、不完全ながらも統合インスタンスを生成してしまう処理も可能である。

#### (7) 送信処理

統合したデータをファイルとして送信する。

### 6. 通信網管理情報のデータ統合実験

今回、データ統合エンジンをperlにより作成して、現実の通信網管理情報（伝送路、パス情報）の統合実験を行った。データ統合エンジンは、統合シナリオを解釈するインタプリタを持ち、perlスクリプトを生成する。各統合メソッドは、perlのサブルーチンであり、インスタンスの存在パターンは、perlの連想配列に展開するようにした。

対象とした流通元データベースは3種類であり、インスタンスの存在パターンによる統合処理の場合分けを4通りとした。複数回の統合の試行で、再蓄積、削除、インスタンス組み立ての動作を確認した。

### 7. おわりに

分散して配置された既存の複数のデータベースの通信網管理情報を変換し、流通させ、統合するデータ統合機構を提案し、その構成法を示した。この機構は、予め用意された統合のためのメソッドを呼び

出すデータ統合エンジンを駆動する統合シナリオを記述することで、データ統合処理を行う。現実と同一の実体を表現している複数のインスタンスの存在パターンによる統合処理の場合分けを自動的に行うことによって、複数のインスタンスの待ち合わせをしたり、指定された条件によりデータ項目の値を選択して統合インスタンスを生成することが容易にできることが特徴である。今回、現実の通信網管理情報のデータ統合を行う実験システムを構築し、その有用性を確認した。

### 参考文献

- [1] Batini, C. et al : A Comparative Analysis of Methodologies for Database Schema Integration, ACM Computing Surveys, Vol.18, No.4, 1986.
- [2] 星野隆, 他 : DB流通におけるデータ変換方式について, 情報処理学会第46回全国大会, 1993.
- [3] 池田哲夫, 他 : DB流通の基本方式について, 情報処理学会第46回全国大会, 1993.
- [4] 木原民雄, 他 : 通信網管理情報のデータ統合規則, 情報処理学会第47回全国大会, 1994.
- [5] Kim, W. et al : Classifying Schematic and Data Heterogeneity in Multidatabase Systems, IEEE COMPUTER, Dec. 1991.
- [6] Sheth, A.P. et al : Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases, ACM Computing Surveys, Vol.22, No.3, 1990.
- [7] 塩原寿子, 他 : DB流通基本システム(DB-STREAM)とその適用, 情報処理学会マルチメディア通信と分散処理研究会64-10, 1994.
- [8] 山口治男, 他 : 通信網管理のための情報インテグレータ, 電子情報通信学会秋季全国大会, 1990.
- [9] 山下亮, 他 : 網オペレーション情報のインテグレーション, NTT R&D, Feb. 1993.