

DB流通における整合性保証法*

6E-8

塩原 寿子†

村田 達彦‡

NTT情報通信網研究所§

1 はじめに

業務毎に独立に開発されてきたDBの中には、重複したデータを持つものがある。これらのDB間における比較的自律性の高い相互運用 [1] の実現方式としてDB流通機構を提案した [2]。本稿では 構造の異種性の観点から整合性を保証する方法について報告する。

2 複数DB間の整合性とDB流通機構

業務毎に開発されてきたDBの中には、重複したデータを持つものがあり、個別管理によるDB間の矛盾が問題となっていた。ただしDB間整合性を保証するために相互運用を考える場合、個々の業務DBが性能などの要因から自律的な運用を前提に作られていることを考慮する必要がある。

ところでこれらの重複データをもつ業務DBは、データの発生源となる原本DBとこれを利用するDBとに分けられることが多い。このような場合には、原本DBでの更新を他のDBに自動的に伝播することにより全体の整合性を保つことができる。

個々のDBの自律性を前提に更新の伝播を実現する方法として、更新データを他のDBに自動的に流通するDB流通機構を提案した。(図1)

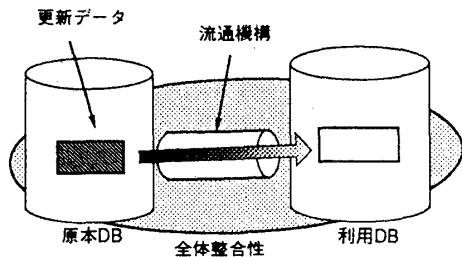


図1: 流通機構による更新の伝播

3 流通機構における異種性

既存DB間の相互運用をする場合にはDB間の異種性が問題となる。DB流通機構でも次の3つの異種性を解決しなければならない [4]。

- 名前レベルの異種性
- 値レベルの異種性
- 構造レベルの異種性

名前レベルの異種性とは、テーブル名や項目名が異なることで、値レベルの異種性とは個々のデータの表現形式が異なることである。これら2つに関してはすでに研究がなされており [3] 流通機構で適切な変換を施せばよい。

3つ目の 構造レベルの異種性とは、DBによって管理する実体の捉え方が異なることである。

個々のDBはそれぞれの業務の特質に応じた設計がされており、重複するデータであってもそれらの構造は一般的には異なる。図2にその例を示す。

本稿ではこの構造レベルの異種性に関する問題とその解決法について報告する。

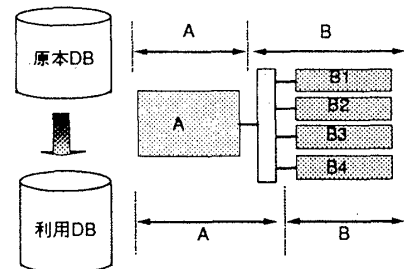


図2: DBによる実体の管理単位の違い

4 構造の異種性の解決

DBの更新は、そのDBが矛盾しないような一定のデータのまとまりで行なわれる。DBによって構造が異なれば、このまとまりも一般に異なる。したがって流通機構は、流通された更新データを、流通先DBを更新できる単位に変換する必要がある。

このデータの単位を流通機構では出力ユニットと呼び、原本DBから流通される更新データを入力ユニットと呼ぶとすると

*Global Integrity on the Data Delivery Mechanism

†Shiohara Hisako

‡Murata Tatsuhiko

§NTT Network Information Systems Laboratories

1. 入力ユニットと出力ユニットが 1 : 1 に対応。
2. 1つの入力ユニットが複数の出力ユニットに分解される。
3. 複数の入力ユニットの結合が1つの出力ユニットになる。
4. 上記の組合せ。

という場合があり、流通機構ではユニットの組み替えを行なう必要がある。このうち結合は、結合すべきユニットの選択において流通機構特有の問題がある。

5 流通ユニットの結合における問題

結合すべきユニットの組合せの選択は、結合条件を指定することによって行なう。この結合条件は、そのユニットの持つデータ項目の値で指定することができる。

しかし、流通機構は更新されたデータを受けとるだけなので、結合をいつまで続ければ良いかわからない、つまり、結合処理を終了できないという問題がある。

この問題をユニット間の基数関係という観点から考察する。

簡単のため、AとBという入力ユニットから、Xという出力ユニットができる例を考える。(図3)

AとXの基数関係は、1つのXを構成するAの個数を C_A 、1つのAから作られるXの個数を R_A とおくと、 $A : X = C_A : R_A$ と書ける。Bについても同様とすると、AとBのXに対する結合関係は、

Aは C_B 個のBと R_A 回結合、

Bは C_A 個のAと R_B 回結合、

となる。

これらの数が全て一定であらかじめわかる場合は、この数を用いて結合処理の終了を示すことができる。しかし、このうちの1つでも不定でわからない場合は、流通機構だけでは結合処理の終了がわからない。

6 問題の対処法

前節の問題に対し、以下の2つの対処法を提案する。

基数指定法 結合すべき相手ユニットが流通されるまで流通機構上で一時蓄積しておき、指定された回

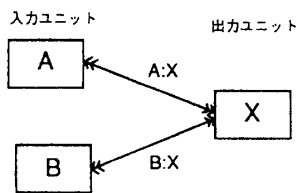


図 3: 基数関係

数だけ結合する方法。

範囲指定法 結合すべきユニットのまとまりを流通元に提示してもらう方法。例えば、1つのファイルに格納して流通する、など。流通機構は、このまとまり内でのみ結合処理を行なう。

基数指定法は、基数関係があらかじめ既知の時に適用できる。ただし、あらかじめ既知でない場合でも、ユニットのデータ内に基数関係に関する情報を保有する場合であれば適用が可能である。

それに対し、範囲指定法はあらゆる場合に適用可能である。しかし、流通元システムが出力先に合わせて流通しなくてはならないという実現上の困難がある。

これを表 1にまとめる。

表 1: 流通機構におけるユニット結合法

ユニット間の基数関係		対処法
あらかじめ既知		基数指定法 OR 範囲指定法
未知	ユニット内に情報保有	基数指定法 OR 範囲指定法
	上記以外	範囲指定法

7 まとめ

DB流通機構における異種構造の変換方法、特に結合問題の対処法について検討した。

今後の課題としては、提案したDB流通機構における結合処理を記述する言語機能等について検討していきたい。

参考文献

- [1] C.Batini et al. "Report of the Workshop on Semantic Heterogeneity and Interoperation in Multidatabase System", ACM Computing Surveys, Vol.18, No.4, pp.323-364.
- [2] 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
- [3] 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
- [4] W.Kim et al. "Classifying Schematic and Data Heterogeneity in Multidatabase Systems", IEEE COMPUTER Dec. 1991 pp.12-18.