

## 3L-1

垂直分散データベースにおける  
重複データの処理について

高野 智 鶴岡 邦敏

日本電気(株)C &amp; C システム研究所

## 1. はじめに

情報化社会の進展と共に、従来の集中的なデータ管理にかわり、企業内の各部門ごとに必要なデータを蓄積・管理し、他の部門のデータはネットワークを介してアクセスする分散データベースへのニーズが高まっている。企業の組織管理と情報の利用形態とを考えると、例えば、上位部門が下位に位置する部門の情報を一括管理したり、また下位部門が上位部門の膨大なデータの一部をローカルに格納して処理をするというように、各部門のデータベース中のデータ間に階層関係が存在する場合が多い<sup>[1]</sup>。このように、データベース間の階層関係に着目し、それらのデータと関係をシステム側で管理するものを「垂直分散データベース」と定義する<sup>[2]</sup>。

本稿では、垂直分散データベースにおける部分コピー(重複データ)の処理に関する基本機能をあげ、それらの機能の実現方法について説明する。

## 2. 重複データ管理機能

## 2.1 データ導出機能

リモートサイトのデータ検索にはデータ転送のコストがかかるので、リモートサイトデータのうち頻りにアクセスする部分のデータは、ローカルサイトに格納する方が転送コストと実行時間の面で効率がよい。リモートサイトの表の部分コピーをローカルサイトのデータベースに作成することを「導出」と呼ぶ<sup>[1]</sup>。

導出は、ローカルサイトのデータベースであらかじめ定義されたSQL文(導出規則)を導出元サイトへ転送し、その検索結果を受け取り、表を作成することにより実行される。導出規則には、導出する属性、導出元の表名、導出条件を指定する。以下に例を示す。

```
SELECT X FROM TAB WHERE C; (式1)
```

ここで、Xは導出する属性の集合、TABは導出元の表、Cは導出条件を表わす。Xの指定方法によって、重複データの形態が異なる。

すべての属性を指定(\*指定)……水平導出  
いくつを選択して指定……垂直導出  
垂直導出の場合は、後述する機能を実行するためにタプルを一意に識別するキー項目は必ず導出する必要がある。

On Processing Replicated Data in Vertically Distributed Databases

Satoshi TAKANO, Kunitoshi TSURUOKA  
NEC Corporation

## 2.2 ローカルデータ優先検索機能

本機能は、リモート検索の際に要求データの一部がローカルデータベースに導出されていれば、自動的にローカルなデータを検索し、残りのデータだけについてリモート検索を行なう機能である。これにより、リモート検索の転送コストが低減され、またデータ導出以前のアプリケーションプログラムの修正を行なうことなしに、ローカルな重複データを利用できる。

## 2.3 重複データ更新機能

データ導出により作成される表と、導出元の表との間にはデータの一意性が保たれる必要がある。本機能は、重複データの更新をアプリケーションで意識的に行わなくても、システム側で自動的に処理する機能である。

## 3. 実現方法

## 3.1 ローカルデータ優先検索機能

システムは、リモート検索の要求が出されると以下のようなステップを実行する。

1. 指定されたリモート表と重複部分を持つ表がローカルに存在するかどうか調べる。
2. 導出規則を検索し、検索命令をローカル用とリモート用に変換する。
3. ローカル、リモートの検索結果をまとめて、アプリケーションに返す。

導出の形態により、ステップ2、3の実現方法が異なる。

## 3.1.1 水平導出の場合

水平導出では、導出元の表のすべての属性が導出先に存在するので、要求属性の部分は変更する必要がない。

ステップ2では、表名をローカルな導出表に変更し、検索条件と導出条件をANDで結び新たな検索条件として、ローカル検索を実行する。リモート検索では、要求データのうちローカルに存在しないものだけを要求すればよいので、検索条件と導出条件の否定をANDで結び新たな検索条件としてリモートサイトへ転送する。ステップ3では、ローカル及びリモートの検索結果の併合(UNION)を行なう。

## 3.1.2 垂直導出の場合

垂直導出では、すべての属性を導出していないので、検索命令に現れる属性がローカルに存在しない場合がある。そのような場合には別の方法をとる必要がある。これについて図1を用いて説明

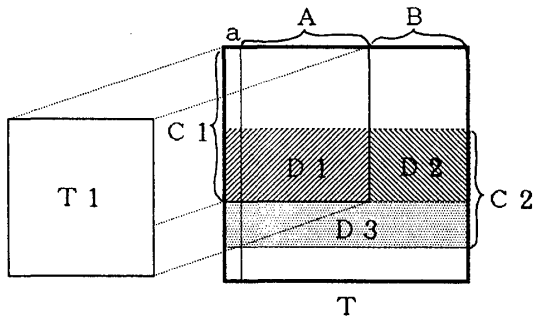


図1 ローカルデータ検索の概念図

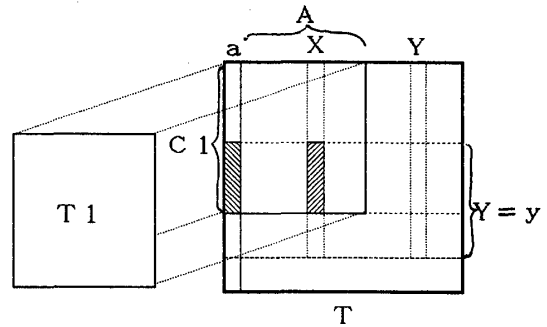


図2 重複データ更新の概念図

する。

リモートサイトの表Tから、以下のような導出規則でT1という表がローカルに存在していると仮定する。

`SELECT a, A FROM T WHERE C1;` (式2)

aはタプルを一意に識別するキー項目である。ここで、次のようなリモート検索を要求したとする。

`SELECT * FROM T WHERE C2;` (式3)

この場合、要求データは図1の(D1+D2+D3)であり、そのうちD1だけがローカルに存在するデータである。そこでステップ2では、式2と式3から次のようにして検索命令を作成する。

• D1部分の検索命令

`SELECT a, A FROM T1 WHERE C1 AND C2;`

• D2部分の検索命令

`SELECT a, B FROM T WHERE C1 AND C2;`

• D3部分の検索命令

`SELECT * FROM T WHERE NOT C1 AND C2;`

そして、ローカル及びリモートで検索命令を実行する。

ステップ3では、D1の検索結果とD2の検索結果をキー項目aを結合属性として結合処理(join)を行い、その結果とD3の併合をとる。

以上のような方法により、タプルを識別するためのキー項目が含まれていれば、任意の部分の導出に対して、リモート検索の際にローカルな重複データを利用することが可能となる。

### 3.2 重複データ更新機能

#### 3.2.1 水平導出の場合

更新された表から水平導出されている重複データには、更新表のすべての属性が含まれるため、その更新は前述の3.1.1の場合と同様に、表名を重複データ表に変更し、更新条件と導出規則をANDで結び新たな更新条件とすればよい。

#### 3.2.2 垂直導出の場合

ここで問題になるのは、更新条件に含まれる属性が重複データに導出されていない場合である。このような場合、重複データ側で更新タプルを識別できないからである。そこで、そのような更新が発生した場合は、次のように対処する。

図2に概念図を示す。図1の場合と同様に、式

2の導出規則によりT1という表がTの重複データとして存在すると仮定する。ここで、次のような更新命令がTに要求されたとする。

`UPDATE T SET X=x WHERE Y=y;` (式4)

更新条件に含まれる属性YがT1に導出されていないことを認識すると、Tにおいて式2の導出条件と更新条件を満たすタプルのキー項目を検索する。

`SELECT a FROM T WHERE C1 AND Y=y;` (式5)

そして、式5の検索結果をもとにT1を更新する次のような命令を作る。

`UPDATE T1 SET X=x WHERE a IN {式5の結果};`

そして、これをT1の導出先に転送する。

以上のような方法により、タプルを識別するためのキー項目が含まれていれば、任意の部分の導出に対して、重複データの自動更新が可能となる。

### 4. おわりに

本稿では、垂直分散データベースにおける重複データに着目し、検索時のコスト削減方法と重複データの自動更新方法について述べた。現在、前章で説明した機能を持つ垂直分散データベース管理システムを開発中である。本稿では、単一表からの部分導出のみを扱ったが、結合導出についても検討している。今後、コスト評価規準を定め、処理の最適化、並列化などの機能を検討し、開発中のシステムに実装する予定である。

### 謝 辞

本研究開発は、通産省工業技術院大型プロジェクト、「電子計算機相互運用データベースシステムの研究開発」の一環として、日本電気㈱がNEDOから委託を受けて実施したものである。

なお、本研究内容に関して貴重な御意見を頂いた日本電気㈱基本ソフトウェア開発本部の浜館昌樹氏および稲葉亜矢子嬢に感謝致します。

### 文 献

- [1] 鶴岡, 高野「垂直分散データベースのモデルと基本機能について」, 情処データベース・システム研究会研報, Vol.90, No.36, 77-2, 1990. 5.17
- [2] 高野, 鶴岡「垂直分散データベースの遅延更新のための一方式」, 第41回情処全国大会講演論文集(4), pp.172-173, 1990.9.4-6