# ＤＢ流通におけるデータ抽出方式に関する考察

5Ｅ－8

奥村昌和　池田哲夫　岸本義一

ＮＴＴ情報通信研究所

## １．はじめに

近年、分散システムを構築する上での現実的かつ有力な手法として、ＤＢの非同期更新（レプリケーション）と呼ばれる新しい手法が注目されている。レプリケーションの機能を備えたソフトウェア製品は、近年急速にその数を増やしている。しかし、多くのレプリケーションソフトはその利用にソフトと同じベンダのＤＢＭＳを組み合わせることを必要としている[1]。

原本ＤＢと複製ＤＢを異なるＤＢＭＳ／マシン上で構築するマルチベンダ環境でのレプリケーションの実現のためには、マルチベンダに対応した①原本ＤＢの更新データ（差分データ）の抽出処理、②複製ＤＢのデータ構造へのデータ変換と流通、および、③複製ＤＢへの差分データの格納処理が必要となる。

上記の３つの処理のうち、②データ変換と流通の機能については我々が開発しているＤＢ流通基本システム（ＤＢ－ＳＴＲＥＡＭ）で実現可能である[2]。しかし、①及び③のデータ抽出格納処理に関しては、システム対応にＡＰを作成する必要がある。

本論文では、上記の抽出／格納処理のうち、特に差分抽出処理の作成を容易とする方式について提案する。

## ２．前提条件

（１）対象とするＤＢ

市販されている主要なリレーショナルＤＢＭＳを対象とする。

（２）ＤＢとのインタフェース

市販されている主要なＤＢＭＳでサポートしている以下の機能の提供を前提とする。

①検索、更新機能

②ログ機能

## ３．差分データ抽出法

差分データ抽出法は以下の４つの方式が考えられる。この４つの方式を整理、評価する。

### ３．１　差分データ抽出法の整理

①全データ比較法

更新時の処理：特に必要なし。

抽出時の処理：スナップショットを取得し、前回のデータ抽出時のスナップショットと比較することで更新されたレコードを検出し、抽出する。

②差分フラグ法

更新時の処理：各レコード上に更新種別（追加、更新、削除）、更新順序番号、更新前のキー値を示すフラグを設定する。

抽出時の処理：更新種別が設定されているレコードを抽出する。抽出が完了した後、更新種別が削除のものはレコードを削除し、更新種別が追加および更新のものについてはフラグを削除する。

③差分テーブル法

更新時の処理：更新のあったレコードを差分テーブルへコピーするとともに、更新種別、更新順序番号および更新前のキー値を付与する。

抽出時の処理：差分テーブルから更新データを抽出し、その後差分蓄積テーブルからレコードを削除する。

---

A Data Extract Method for a Data Delivery System
Masakazu OKUMURA, Tetsuo IKEDA, Giichi KISHIMOTO
NTT Information and Communication Systems Laboratories

④更新ログ法

更新時の処理：特に必要なし。

抽出時の処理：ＤＢＭＳが出力するＤＢ更新ログを解析し、差分データを抽出する。

### 3．2　評価

４つの差分データ抽出法を①更新時の処理時間、②更新処理と抽出処理の併存、③抽出時の処理時間、④抽出処理のためのディスク使用量、および⑤実現上の問題点の観点で評価し表１に示す。総合的に評価すると、データベーストリガ機能を用いることで実現上の問題（表１⑤）が解決でき、かつ処理性能（表１①～④）が良い差分テーブル法が最も優れていることがわかる。

表１．差分データ抽出法の比較

| 比較項目 | 全データ比較 | 差分フラグ法 | 差分テーブル法 | 更新ログ法 |
|---|---|---|---|---|
| ①更新時の処理時間 | ◎ 更新処理への影響なし | ○ フラグの設定処理分だけ，処理時間増加 | △ 差分テーブルへのデータ格納の処理時間分増加 | ◎ 更新処理への影響なし |
| ②更新処理と抽出処理の併存 | × スナップショット作成時，更新処理不可となる | △ 抽出処理中は，更新処理不可となる | ○ 抽出処理中は，更新処理不可となる | ◎ ログファイルを切替えることで並行して実行可 |
| ③抽出時の処理時間 | × 全データを比較する必要がある | × 全データのフラグのチェックが必要がある | ◎ 差分テーブルのみの参照で差分データ抽出可能 | ○ 更新ログの構成に依存する |
| ④抽出処理のためのディスク使用量 | × 原本ＤＢの２倍のディスク容量が必要 | ○ フラグ分のディスク容量増加 | △ 更新頻度が高い場合にディスク容量増加 | ◎ 新たなディスク容量の増加はない |
| ⑤実現上の課題 | × スナップショット取得の高速化 | △ 更新発生の検出法が必要 データベーストリガ機能で対処可能 | △ 更新発生の検出法が必要 データベーストリガ機能で対処可能 | × ＤＢＭＳ依存のログ情報のデータ構造の開示，及び個々のデータ構造に対応した抽出処理 |

## 4．差分データ抽出の簡易な実現方式

### 4．1　DB-STREAM

ＤＢ－ＳＴＲＥＡＭは、データ流通機構の構築期間の短縮化をはかることを目的として開発されている。その特徴はデータ流通に必要な基本的な部品（メソッド）をあらかじめ用意し、その組み合わせ（シナリオ）をＨＭＩ（ヒューマンマシンインタフェース）で非手続的に記述することでデータ流通機構を構築できることである。

### 4．2　差分データ抽出の簡易な実現方式

ＤＢ－ＳＴＲＥＡＭの設計思想（シナリオ／メソッド方式）に基づき、かつ一部の既存メソッドを利用することで最低限の記述量で差分データ抽出を実現する方式を提案する。処理フローを図１に示す。

スキーマ検索メソッド、差分テーブル定義メソッドおよびトリガ定義メソッドをそれぞれＤＢＭＳ対応に用意しておく。それらのメソッドの組み合わせと必要最低限のパラメータ（対象ＤＢＭＳ、差分抽出を行うテーブル名およびキー項目）をＨＭＩを用い非手続的に指定することで差分データ抽出を簡易に実現する。

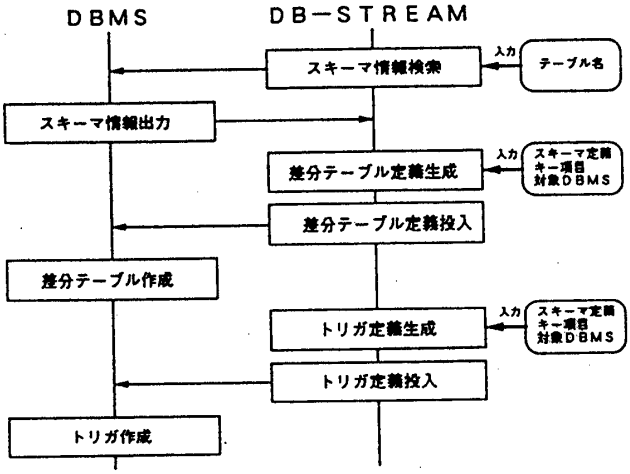


図１．処理フロー

## 5．おわりに

本論文では差分データ抽出を行う４つの方法を分類整理および評価した。また評価の結果優れていることが判明した差分テーブル法について、簡易に実現する方式を提案した。

今後はプロトタイプシステムを作成し、提案した方式の有効性を確認する予定である。

**参考文献**

[１] Yongdong Wanget al. :Data Replication in a Distributed Heterogeneous Database Environment: An Open System Approach, Conf Proc Annu Phenix Conf Comput Commun, 1994.4

[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, 1993