4W－7

論理ジャーナルを用いた複製データベースオンライン同期方式

森下　慎次　　小林　伸幸
NTT情報通信研究所

1．はじめに

ネットワークの普及により、データベースサービスにも広範囲からのアクセスに高速に対応できる能力が要求されている。この要求に応える方式として、分散データベースシステムが有力視されていたが、分散データベースは処理が複雑になりすぎるため実装が難しいという問題点がある。

この問題に対処するため、分散データベース実現のネックになっている複数サイトのデータベースを同期更新するという制約を非同期更新に緩和した複製データベースが注目を集めている。

そこで本報告では同一内容のデータベースすなわち複製データベースをシステム内に複数持つことにより負荷分散/危険分散を図る複製データベースの実現に必要なオンライン同期方式を提案する。

2．複製データベース

2．1　複製データベースとは

複製データベースとは、同じデータを持つ複数のデータベース群のことであり、マスタサイトのデータが更新(insert/delete/update)された場合、複製サイト全てに更新が伝播される形態になっているデータベースである。

2．2　複製データベースの効果

複製データベースでは、複数のサイトに同じデータを分散して保持することにより、アクセスを分散させる効果とサイト障害時に対するフォールトトレランス向上という２つの効果がある。

A Method of Logical Replication for Databases
Shinji MORISHITA, Nobuyuki KOBAYASHI
NTT Information and Communication Systems Laboratories
1-2356 Take,Yokosuka,Kanagawa 238-03,Japan

2．3　対象モデル

複製データベースシステムは、図１に示すようにマスタサイトにあるマスタデータベースと複製サイトにある複製データベースとで構成される。

本報告では、マスタサイトと複製サイトとの関係を1:nとしたモデルを使用する。

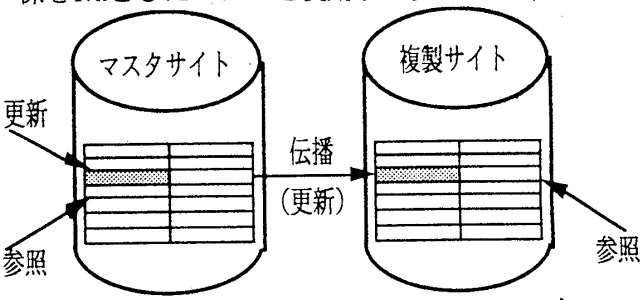

図１．複製データベースのイメージ

(1) マスタサイト

マスタサイトは、本モデルではユーザからの更新処理を許されているサイトであり、更新の順序性の問題からシステムで１つとしている。

(2) 複製サイト

複製サイトでは、負荷分散を実現するためにユーザからの検索処理のみを受け付けており、更新処理は図１にあるように全てマスタサイトからの伝播により行う。ただし、データベースの物理的格納位置/構造の変更が伴う再構成/再編成等の保守作業に関しては、マスタサイトから作業結果を伝播するにはあまりに転送情報量が膨大になるため、複製サイトで独立して行うとする。

本モデルではデータベース保守を各サイトで独立して行うため、マスタサイトのデータベースと複製サイトのデータベースの物理構造が等しくない場合が発生しうる。そこで、本報告では、物理構造の異なる複製サイトにマスタサイトの更新情報を反映する方式を報告する。

3. 論理ジャーナルを用いた同期方式

マスタサイトで行われた更新処理(update, insert, delete)を複製サイトに伝播させるための転送情報としては、現状、(1)通常ジャーナルに変更時刻/更新種別を付加したもの(2)更新要求(SQL)に似たフォーマットの２種類が存在する。しかし、方式(1)では、マスタサイトと複製サイトの物理構造が異なる場合に適応できないという問題があり、方式(2)ではマスタサイトでの負荷増が大きくなるという問題がある。そこで本報告の提案方式は、この２方式の中間案で、通常ジャーナルをベースに複製サイトにおいて通常SQL処理のルートを行うための情報を付加したジャーナル（論理ジャーナル）を使用する。

本方式のイメージ図を図２に示す。以下、マスタサイト、複製サイトにおける処理イメージを説明する。

(1) マスタサイト（論理ジャーナル作成処理）

通常更新処理のジャーナルを出力する際に更新に使用したキー情報をジャーナルに追加して伝播情報、すなわち論理ジャーナルを作成する（論理ジャーナルの概略については図２を参照）。

本方式では、論理ジャーナルを作成する処理において、通常発生するジャーナルを使用するためにマスタサイトでの負荷をあまり増加させずに処理を実現することができる。

(2) 複製サイト（伝播情報反映処理）

複製サイト側では、マスタサイトから転送されてきた伝播情報を反映する処理を行う。その際、論理ジャーナルに含まれている論理位置情報（インデックス名、キー値）を元に通常のSQL処理ルートを通してデータベースの更新処理を行うため、マスタサイトと複製サイトのデータベースの物理構造の違いに関係なく同期処理が行えるとともに、複製サイトでの障害復旧用の通常ジャーナルを取得することも可能になる。このため、複製サイトは、各サイト単位でリカバリを実行できる。

4. おわりに

本報告では、複製データベースの同期処理に論理ジャーナルを使用することにより、マスタサイトの負荷増を抑え、かつ、複製データベース間の物理構造の違いに関係なく同期処理を行う方式を示した。

参考文献：
日経エレクトロニクス no.609 (Page101)
「複製ファイルの非同期更新が
分散データベースの中心技術に」
1994.6.6

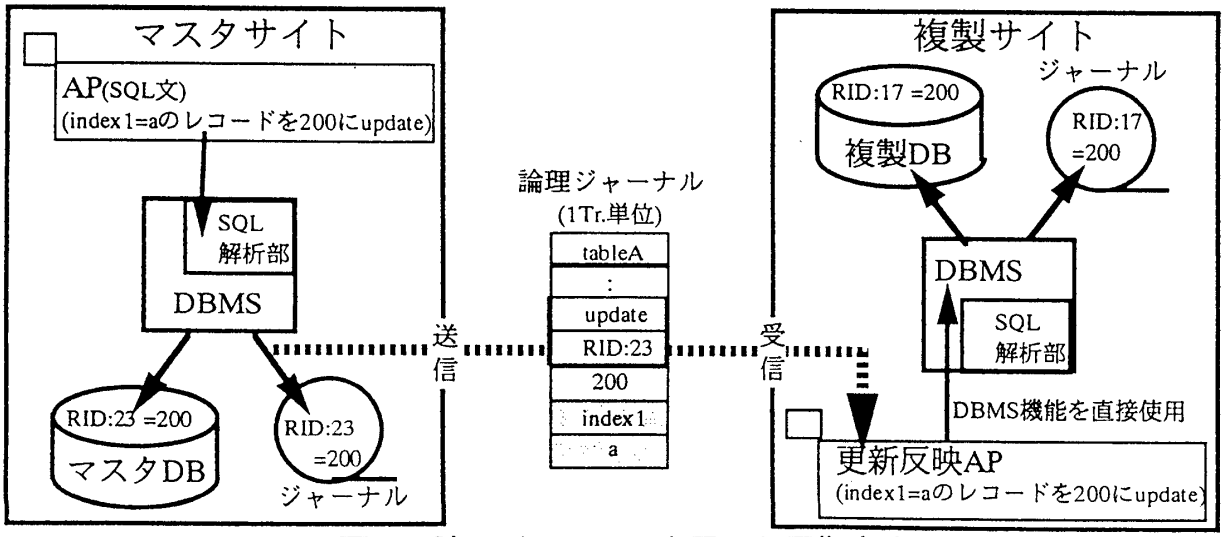

図２．論理ジャーナルを用いた同期方式