

OSI RDAを用いた分散データベース・アクセスの実現 (2) — 分散処理メカニズム —

5H-5

西山 智 杉山 敬三 小花 貞夫

国際電信電話株式会社

1. はじめに

データベース・システムの増加に伴い、データベース通信を行うためのプロトコルの必要性が高まり、OSIにおいてもRDA(遠隔データベース・アクセス)として標準化が進められている。RDAでは、サーバ/クライアントモデルに基づく1対1のデータベース・アクセスを規定している。

筆者らは現在、リレーショナル型データベース(RDB)を用いたRDAプロトコルの実装を進めており^[1]、その応用として分散データベース・アクセスを実現している^[2]。本稿では、その分散処理メカニズムについて報告する。

2. 分散データベース

以下では、物理的に異ってはいるが通信回線等で接続されているシステム(以下ノードと呼ぶ)間に分散格納されており、データベース利用者からデータの分散を意識しないで操作可能なデータベースを分散データベースと定義する。

2.1 RDAを用いる場合の前提条件

各ノード間の通信プロトコルとしてRDAを用いる場合、分散データベースとして以下のような前提条件を仮定することができる。

- (1) 各ノードは通信回線により疎結合されており、“通信コスト” >> “データ処理コスト”である。
- (2) 各ノードに対する操作分割/結果合成は、利用者からの操作要求を受けたノードしか行えない。(RDAサービスでは操作要求は行えるが、“手続き”は転送できないため)

さらに、ここでは以下の条件を追加した。

- (3) RDAで転送するデータ操作言語としてSQLを用いる。従って以下ではデータベース管理システムとしてRDBを用いて議論を進める。
 - (4) 各テーブルは1つのノードにのみ格納されており、複数のノードに分散格納されていることはない。また、同様に同じデータを複数ノードでコピーを持つこともない。
 - (5) 分散処理を行うための知識として、①テーブルと格納ノードの対応関係、②テーブルの構造(各カラムの名等)を持つが、テーブル内のデータに関する知識(タプル数等)は持たない。
 - (6) (5)の知識、すなわちスキーマは固定とする。
 - (7) ノードは任意のノードに直接通信可能とする。
- 以上の条件より、ここでの分散データベースのアクセス・モデルは図1のようになる。

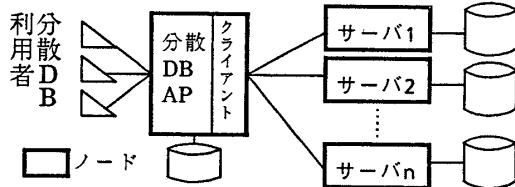


図1 分散データベース・アクセスモデル

2.2 分散処理の形態

利用者からの操作により図2に示すような処理形態がある。

- ① 利用者からの操作に含まれるテーブルが全て自ノードにある場合、自ノードでその操作を実行し結果を利用者に返す。
- ② 利用者からの操作に含まれるテーブルが全て他の1つのノードに含まれる場合、そのノードに操作を転送し、結果を利用者に中継する。
- ③ ①②以外の場合、操作を解析し必要なテーブルのデータを各ノードに対して要求し、それらの結果をもとに自ノードで結果合成を行い、利用者へ返送する。

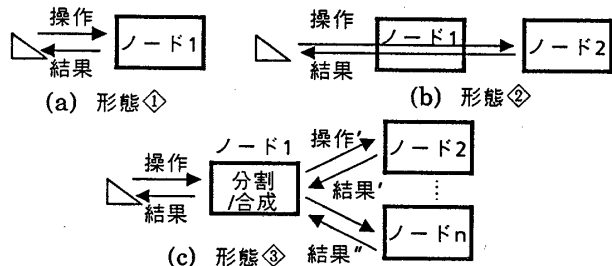


図2 分散処理の形態

2.3 操作分割/結果合成のアルゴリズム

2.2節③の場合、操作の分割と結果の合成処理が必要となる。2.1節(1)の条件より、操作の処理コストを抑えるためには、データ処理量が多くなっても各ノード間のデータ転送量を小さくすればよい。ノード間の転送データ量を減少させるためには、次のような方法が考えられる。

- a) データに関する知識から処理ノードを選択する
- b) 準結合(semi-join)を適用する
- c) 転送を行う前にノード内で、射影(projection)、選択(selection)を行いデータ量を絞る

2.1節の条件に基づけば、(a)は(5)により、また(b)は(2)によりここでは適用することが出来ない。従って、(c)の手法のみを適用する。また、(2)より

操作分割/結果合成を行えるノードは利用者からの操作要求を受けたノードに限定される。

これらの点を考慮すると、操作分割/結果合成の処理は次のようになる。3)及び4)の手順は基本的に関係演算での最適化を行う処理^[3]と等価である。

- 1) 利用者の操作要求(SQL)を解析し検索木を得る。
この際、操作中のORは和集合(union)を用いた検索木に変換する。
- 2) 検索木に対して他のノードで処理可能な部分(検索部分木)を決定する。この際、同一ノードの複数のテーブルが直積(cartesian product)の関係にある場合、直積のまま、まとめて部分木とする。また、副照会(sub-query)については、一番ネストの深い副照会がそれに連続したネストの浅い副照会と同じノードのテーブル群に関するものであればまとめる。他の場合は副照会の1段毎に分割する。
- 3) 検索木の選択の内、そのノードのテーブル群に関するものを、その検索部分木まで落とす。この時、副照会は直積/選択/射影と扱う。
- 4) 各検索部分木に射影を追加する。この射影は、その部分木から木を上にある他の射影まで辿り、その射影及び途中の選択に出現したその部分木に関する属性の集合として得られる。この際も、副照会は直積/選択/射影と扱う。
- 5) 各検索部分木から、他ノードに対する操作を生成する。
- 6) 自ノードで結果合成に用いる操作を生成する。
例えば、図3に示す操作から操作分割/結果合成を行うと図4の例になる。

```
SELECT name,projname
FROM emp,proj,emp-proj
WHERE emp.emp# = emp-proj.emp#
AND emp-proj.proj# = proj.proj#
AND (projname = 'RDA'
OR projname = 'FTAM'
AND proj.proj# IN ( SELECT user-proj#
FROM parts
WHERE parts-name = 'ACSE'));
```

図3 操作例

```
ノード2へ SELECT projname,emp-proj.emp#
結果work1 FROM proj,emp-proj
WHERE emp-proj.proj# = proj.proj#
AND projname = 'RDA';

ノード2へ SELECT projname,emp-proj.emp#,proj.proj#
結果work2 FROM proj,emp-proj
WHERE emp-proj.proj# = proj.proj#
AND projname = 'FTAM';

ノード3へ SELECT user-proj#
結果work3 FROM parts
WHERE parts-name = 'ACSE';

合成操作 SELECT name,projname
FROM proj,work1
WHERE emp-proj.proj# = work1.proj#
UNION
SELECT name,projname
FROM proj,work2,work3
WHERE emp-proj.proj# = work2.proj#
AND work2.proj# IN ( SELECT proj#
FROM work3);
```

図4 操作分割/合成操作例

2.4 操作分割/結果合成の実現

今回の実装では、結果合成の処理はRDBの機能を用いている。このため、途中結果の格納用にRDB上に一時的な作業テーブルを生成する必要がある。

RDAはトランザクション機能を提供するために、RDBのトランザクション機能を利用しているが、今回用いたRDBのORACLEでは、作業テーブルの生成/削除等によりスキーマ変更がなされると、自動的にトランザクションがコミットされてしまう。このコミットをさけるために、今回の実装では、作業用テーブル生成/削除を別のプロセスとして実行させている。図5に操作分割/結果合成処理の流れを示す。

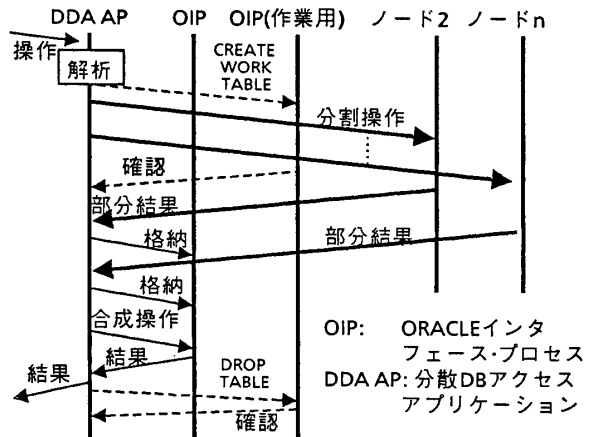


図5 操作分割/結果合成処理の流れ

3. 考察

- (1) 今回の分散データベース・アクセスでは、タプルの水平/垂直分散を実現していない。通信手段としてRDAを用いる場合、操作分割/結果合成を行うノードが利用者から操作要求を受けたノードに限定されるので、タプルの水平/垂直分散を許した場合もこの手法の延長で実現できる。この場合、2.3節2)での処理可能な部分木は単一のテーブルに限られて、直積にまとめて扱うことはできなくなる。その制限を付加するだけで、垂直分散については、2.3節5)で対応するテーブルのうち、その射影に関するノード群に操作を転送すればよい。また、水平分散については、選択条件がその水平分散条件と空にならないノード群に転送すればよい。
- (2) 効率的な分散データベースを実現するためには、データの分散に関する知識を利用して、最適な処理ノードを選択する必要がある。しかし、RDAでは処理の依頼が行えない。より効率的な分散処理メカニズムの実現には、RDAに加えて、“処理手続”を転送できるプロトコルが必要であろう。

4. おわりに

本稿では RDAを用いた分散データベース・アクセス・アプリケーションの分散処理の実現手法について報告した。最後に日頃御指導頂くKDD上福岡研究所 小野所長、浦野次長、コンピュータ通信研究室 鈴木室長に感謝します。

参考文献

[1] 西山,杉山,小花,“OSI RDA実装の基本設計”,第37回情報全大,1988
 [2] 小花,西山,杉山,“OSI RDAを用いた分散データベース・アクセスの実現(1) - 全体概要 - ”,本全国大会予稿
 [3] J.D.Ullman,“Principles of DATABASE SYSTEMS”, pp. 268 -283