

分散DBMS DREAMの  
問い合わせ処理

3H-5

萩野正樹、小寺 誠、坂本明史、正田定幸

沖電気工業株式会社

1. はじめに

分散データベースシステムにおいて、問い合わせ処理は主要な要素と言える。

SQLで書かれたアプリケーションプログラム(AP)は、分散データベースシステムDREAMにおいても、集中型データベースシステムREAM[KANK89]においても、何の変更もなく実行できなければならない。そのためには、APが複数のサイトにあるすべての表を利用できること、APから渡される問い合わせを効率良く実行できる最適な戦略を提供すること、そして、問い合わせ処理でポトルネックになりがちな通信処理を効率良く行えること等の問題がある。

本稿では、分散データベースシステムDREAMの問い合わせ処理の基本方式と処理構造について報告する。

2章でこれらの課題を達成するための基本方式を、3章でその処理構造を述べる。

2. 基本方式

2.1. 表名の扱い

DREAMシステムではサイト自律性を実現するため、表名はサイト内で一意とし、かつ、「サイト名:表名」の一意名により表をシステム内で一意に識別する。また、格納位置の透過性を実現するために、一意名に対し、サイト名を除いた一意名(表名)を別名として定義する。DREAMシステムは一意名と別名をデータディレクトリで1対1に対応づけて管理する。

問い合わせ処理では、DREAMシステムはSQLで指定した表を別名として解釈する。そして、データディレクトリを参照して別名に対応する一意名を調べる。このとき、別名がデータディレクトリに定義されてなければ自サイトの表として解釈する。このような方法で、DREAMシステムは表の格納位置を決定し、APはSQLでシステム内の任意のサイトの表にアクセスできる。

2.2. 最適化

DREAMシステムが、サイト間ジョインなどのサイトにまたがる演算を実行する場合、問い合わせの最適化を行う。DREAMシステムは、問い合わせの中に出てくる表の格納位置とスキーマ情報を上述した方法でデータディレクトリから参照する。次に、参照した情報から、複数サイトの表を含む問い合わせを、各サイトに閉じた表を検索する問い合わせに分解する。その際、問い合わせ処理が効率良く実行されるように最適化を図る。すなわち、インデックスの有無、検索条件のセレクトィビティとおよそのタプル数、データの所在(自サイト/他サイト)、転送/検索データ量などから処理コストを評価し、問い合わせ処理の戦略を確定する。

DREAMシステムは、ネステッドループ法を基本として問い合わせを処理する。その際、どの表をインナに置き

、どの表をアウトに置くかは、最適化で作成されたプランに基づき決定される。

2.3. 問い合わせの実行

DREAMシステムでは、分解及び最適化された問い合わせはパイプライン処理[斎坂89][HIKH85][LUCA85]される。パイプライン処理を実現するために、非同期通信プリミティブによるプリフェッチ機能を用意している。以下にプリフェッチ機能によるパイプライン処理の概要を示す。

図1に、サイトAのデータをサイトBが処理する場合のプリフェッチによるパイプライン処理を示す。サイトBはサイトAからタプルを受け取ると、タプルの処理を行う前に次のフェッチ要求を出す(①)。サイトBはフェッチ要求の応答を待たないで受け取ったタプルの処理を行う。処理が終了したら次のタプルを受け取り(②)、受け取ったタプルを処理する前に次のタプルの要求を出す(③)。

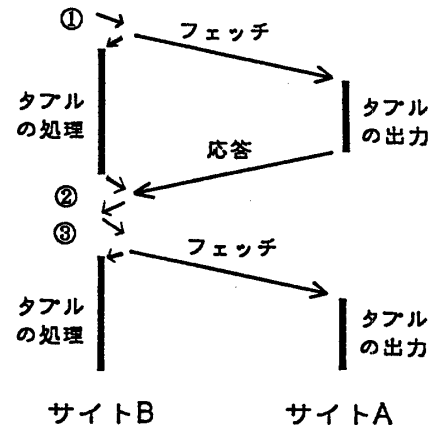


図1 プリフェッチ機能によるパイプライン処理

このシーケンスからわかるように、サイトBで受け取ったタプルを処理している間に、サイトAはサイトBからのフェッチ要求を処理する。つまり、サイトAとサイトBは並列に処理が進む。さらに、DREAMシステムでは、並列処理の効率を良くし、通信オーバーヘッドを削減するために、転送するタプル数組をブロック化して行う。これによって、問い合わせ処理の実行における通信オーバーヘッドを削減する。

3. 処理構造

3.1. 処理手順

図2に利用者のアプリケーションプログラムの実行手順を示す。SQL文は、SQL埋め込みホストプログラムの規則[SQL]に従い、親言語で記述されたアプリケーションプログラムの中に埋め込まれる。親言語にはCとCOBOLがある。このプログラムは、SQLプリプロセッサにより対応する手続き呼出しを含む親言語のソースプログラムに変換される。さらに、親言語の処理系によって実行形式プログラムが作成される。SQLプリプロセッサなどのプ

Query Processing in DREAM.

by Masaki OGINO, Makoto KOTERA, Akifumi SAKAMOTO  
and  
Sadayuki HIKITA

Oki Electric Industry Co., Ltd.

プログラミングシステムはREAMとまったく同様のものを用いることができる。

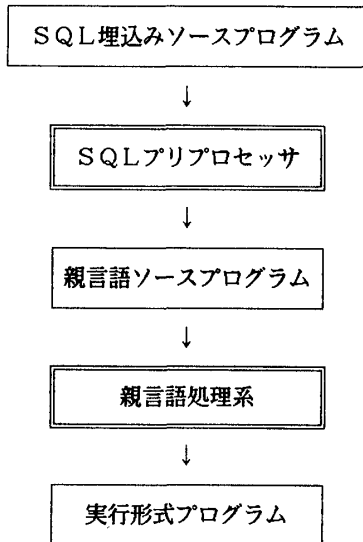


図2 アプリケーションプログラムの作成過程

作成された実行形式プログラム中では、問い合わせに使う表の格納位置は未解決なため、表の所在に関する最適化処理は行われていない。問い合わせは実行時にDREAMに渡され、表の格納位置の決定がなされ、問い合わせの処理が行われる。

この方法では、すでにREAM用に開発されたアプリケーションプログラムは再コンパイルすることなく分散環境で実行できる。例えば、REAMを含むシステムで1つのサイトにデータベースが入らなくなった場合、サイトを1つ追加して、表のいくつかを新しく追加したサイトに移動だけで、REAMのアプリケーションプログラムは変更なく利用できる。

### 3.2. 問い合わせ処理部の構成

DREAMのクライアントは、問い合わせ処理部、トランザクションマネージャ、サイトマネージャ及びデータベースアクセスプロトコルモジュールからなっている[HYKS89]。ここでは、問い合わせ処理部の処理構成について述べる。図3に問い合わせ処理部の構成を示す。

#### (a) データディレトリマネージャ

データディレトリマネージャは、データディレトリを検索して、指定された表名の解釈を行い、格納サイトを決定する。また、オブティマイザに表のスキーマ情報を提供するために、格納サイトにスキーマ情報のアクセスを行う。

#### (b) オブティマイザ

オブティマイザは、まず、データディレトリマネージャを呼び出して、処理に必要な表のスキーマ情報を得る。次に、SQLのカーソル宣言に対応して問い合わせの分解と最適化を行い、問い合わせ実行部への入力になる実行プランを生成する。生成したプランは、ループのネストの順番とタプルサブステイテューションを行う列の情報などからなる。

#### (c) 問い合わせ実行部

問い合わせ実行部は、オブティマイザで作成した実行プ

ランに従って各サイトのカーソルからの行のフェッチ、パラメタの転送を制御する。また、ORの条件評価もここで行う。

一方、サーバはローカルなデータベース管理システムであるREAM、トランザクションマネージャ、問い合わせ処理部及びデータベースアクセスプロトコルモジュールからなる[HYKS89]。サーバの問い合わせ処理部は、基本的にREAMの機能として実現されており、インタフェースの整合だけを行う。

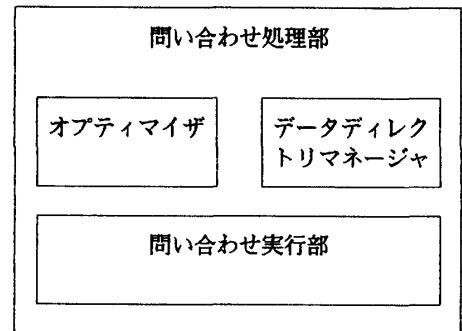


図3 クライアントの問い合わせ処理部の構成

### 4. おわりに

本稿では、DREAMシステムの問い合わせ処理方式について報告した。

現在、筆者らは実装したシステムを用いて評価をしている。これまでの評価から、DREAMの問い合わせ処理は実用レベルに達していると言う感触を得た。

今後は、この評価結果を基に最適化とパイプライン処理の効率化等について検討を進める予定である。

### 《参考文献》

- [KANK89] Kawakami, S., Nakayama, T., Kashiwabara, K., Hikita, S., "REAM: An SQL Based and Extensible Relational Database Management System", Proc. International Symposium on Database Systems and Advanced Applications, pp. 166-170, 1989
- [LUC85] Lu, H., Carey, M., "Some Experimental Results on Distributed Join Algorithms in a Local Network", Proc. VLDB85 pp. 292-304, 1985
- [WOY076] Wong, E., Youssefi, K., "Decomposition - A Strategy for Query Processing", TODS vol. 1, No. 3, 1976
- [WIDA82] Williams, R., Daniels, D., "R\*: An Overview of the architecture", Improving Database Usability and Responsiveness, Academic Press, 1982
- [HIKH85] Hikita, S., Kawakami, S., Haniuda, H., "A Stepwise Approach to Distributed Database Systems by Database Machines", Proc. 1985 ACM SIG SMALL Symp. Small Syst., pp. 18-24, 1985
- [斎坂89] 斎藤, 坂本, 川上, 疋田, "パイプラインによるサイト間結合演算の並列処理", 情報処理学会第38回全国大会, pp. 885-886
- [坂疋89] 坂本, 川上, 吉田, 疋田, "DREAMシステムについて", 情処研究会報告89-DBS-73, 1989
- [SQL] 日本工業標準調査会, "データベース言語SQL", JIS X 3005, 1987