

並列DBサーバシステムにおけるDB再配置方式の提案

1 F - 1

原 憲宏*、河村信男*、根岸和義*、正井一夫**、宮崎光夫**
 *日立製作所システム開発研究所、**日立製作所ソフトウェア開発本部

1.はじめに

高速通信ネットワークで接続される複数のプロセッサからなる並列コンピュータ向けのクライアントサーバ型のDBサーバの必要性が高まってきている。並列DBサーバは、従来、単体プロセッサで処理時間のかかった大量検索などのDB処理を高速に、しかも短時間で処理できる。並列DBサーバのサポート方法として次の2通りがある。

■Shared-Everything方式

複数のプロセッサ間でDBを共用管理する。

■Shared-Nothing方式

各プロセッサにDBを分割管理させ、プロセッサ間では共有しない。

本並列DBサーバシステムでは、スケーラビリティおよびシンプルなシステム設計を支援できるShared Nothing方式を選択した。

本稿では、後者のShared Nothing方式(ディスク非共有型)のDBサーバにおける分割された表の再配置方式を提案する。

2.並列DBサーバのシステム構成

本システムは、図1に示すように各プロセッサに割り当てられた以下の4つのサーバから構成される。

(1)フロントエンドサーバ(FES)

SQLを受けて、SQLを分解してDB処理要求を並列化する。

(2)ジャーナルサーバ(JS)

BES等で行った更新履歴情報(ジャーナル)を管理する。

(3)辞書サーバ(DS)

DB定義情報を管理する。

(4)バックエンドサーバ(BES)

FESより指示されたDB処理要求にしたがってDBをアクセスする。

以上、述べたように目的に応じてサーバを分けることにより、構成変更の容易化ならびに運用の簡便性を実現する。

また、本並列DBサーバのDBの表の分割方法は、次のものがある。

■キー・レンジ分割方式

表のある特定の列の値(範囲)でBESに分割する。

■ラウンド・ロビン分割方式

複数のBESにデータ量が均等になるように分割する。

3.並列DBサーバシステムの課題

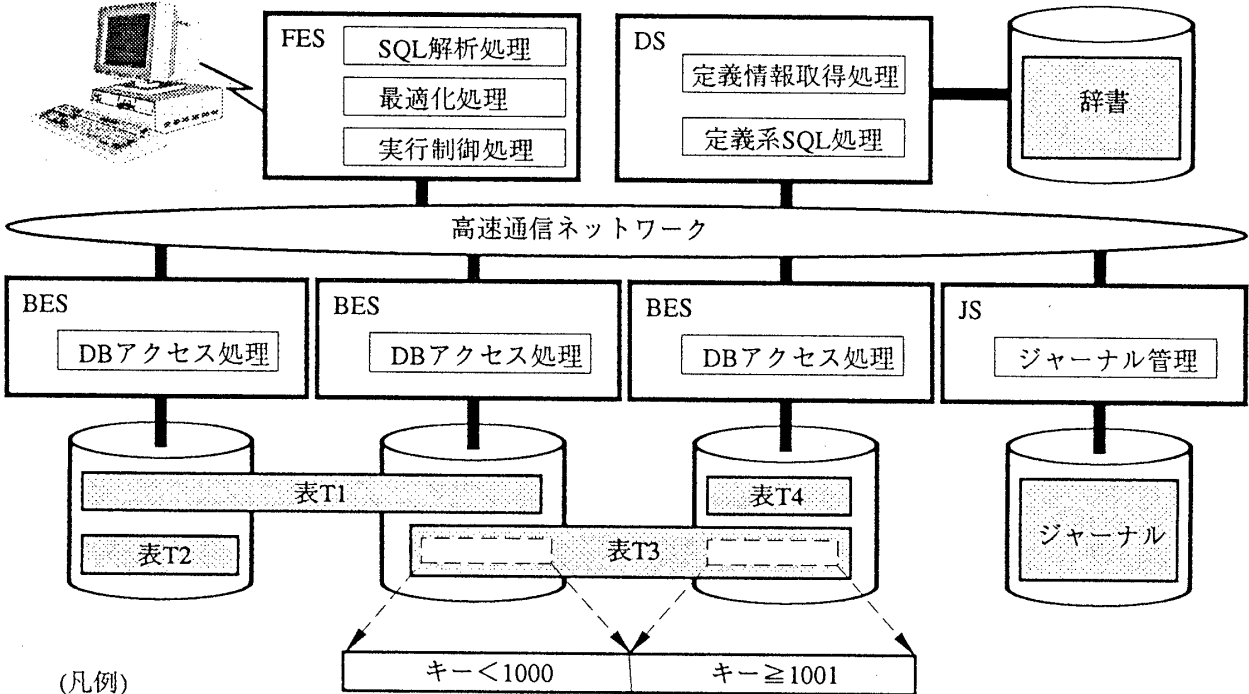
並列DBサーバシステムでは、次の課題について考慮する必要がある。

(1)データの片寄りとは負荷バランスの対応

データの片寄りによる負荷バランス調整のためのデータ再配置方法が必要となる。

(2)結合処理の並列処理方法の最適なアクセス手順の選択

複数のBESに分割された表を結合するためのBESの選択方法および結合処理のパイプライン化。



(凡例)

FES:Front-End Server,JS:Journal Server,DS:Dictionary Server,BES:Back-End Server

図1 並列DBサーバシステムの構成

Proposal of Data Reallocation Method for Parallel DB Server System

*Norihiro HARA,*Nobuo KAWAMURA,*Kazuyoshi NEGISHI,**Kazuo MASAI,**Mitsuo MIYAZAKI

*Hitachi Systems Development Laboratory,**Hitachi Software Development Center

本稿では、前記課題のうち、負荷バランスへの対応策としてデータの再配置方式について述べる。

4.DB再配置方式

4.1 DB分割時の課題

並列DBサーバにおいて、DBの表を分割して格納する場合、キー・レンジ分割方式を採用すると、ユーザはデータをどのBESに分割したかを明確にすることができる。その反面、表を分割する場合に各BES毎のアクセス頻度が均一になるように分割しておかなければならないという欠点を持つ。

また、運用によって、DB内のデータの追加、削除等によって物理的な格納の配置が乱れ、さらに各BESに存在するデータ量にも変化が生じるので各BESのアクセス頻度の均一化が崩れ、特定のBESに負荷が集中したり、逆に負荷がほとんどかからないBESが出現してくる。

4.2 DB再配置方式

複数のBESに分割された表のデータを再配置時にアクセス状況に応じてデータ量、およびアクセスの負荷の両面について不均衡な状況をなくし、常に表に対するアクセスの負荷が各BESで低い負荷となるようにデータベースの再配置を行う。図2に示すような処理の流れにしたがってDBの再配置を行う。

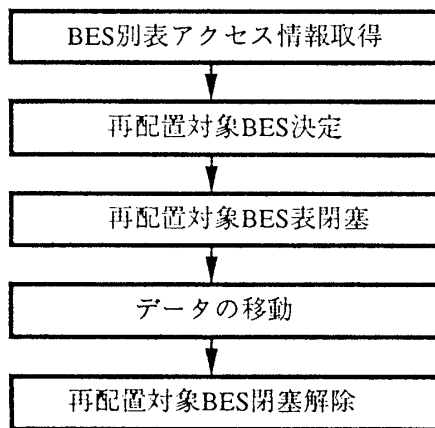


図2 DB再配置方式の処理の流れ

データベースの再配置を行う場合の方法として次の方法がある。

■アクセス頻度が高いBESの表のデータを分割する方法

■アクセス頻度が低いBESの表のデータを統合する方法

(1)アクセス頻度が高いBESの表のデータを分割する方法

DB再配置時に、ある表のアクセス頻度が高いBESに格納したデータを表の分割方法によって他のBESにデータを分割するようになる。

■キー・レンジ分割の場合

表の分割方法がキー・レンジ分割の場合は対象となったBESのキー・レンジをさらに細分化してキー・レンジに分けて分割する。

■ラウンド・ロビン分割

表の分割方法がラウンド・ロビン分割の場合は対象となったBESのデータ量を分割する他のBESとデータ量が均等になるように分割する。

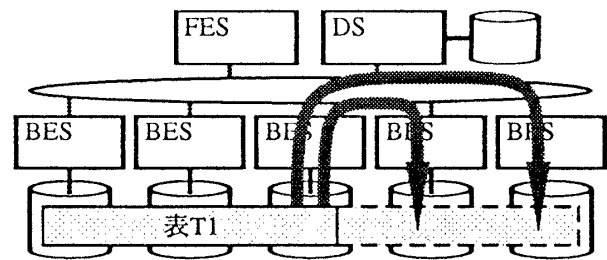


図3 アクセス頻度が高いBESのデータ再配置方式

(2)アクセス頻度が低いBESの表のデータを統合する方法

DB再配置時に、ある表のアクセス頻度が低いBESに格納したデータを表の分割方法によって他のアクセス頻度の低いBESのデータと統合するようになる。

■キー・レンジ分割の場合

表の分割方法がキー・レンジ分割の場合は対象となったBESのキー・レンジを隣接する他のBESのキー・レンジにデータを統合する。

■ラウンド・ロビン分割

表の分割方法がラウンド・ロビン分割の場合は対象となったBESのデータ量を均等に他のBESに分配する。

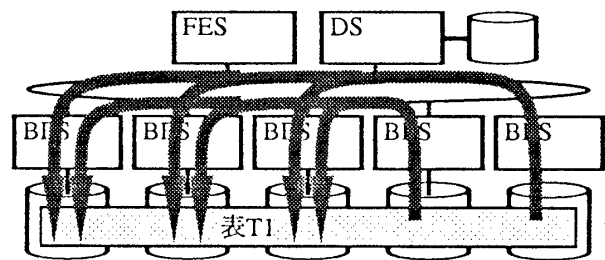


図4 アクセス頻度が低いBESのデータ再配置方式

5.おわりに

並列DBサーバにおけるDB再配置方式を提案した。

(1)アクセス頻度が高いBESの表のデータを分割して他のBESへデータを移動する方式

(2)アクセス頻度が低いBESの表のデータを他のBESの同じ方のデータへ統合する方式

これにより、複数のBESに分割された表に対するアクセスのボトルネックを解消し、スループットを向上することができる。

【参考文献】

- 1.David.J.DeWitt他,"GAMMA A High Performance Dataflow Database Machine",VLDB,1986
- 2.Gregory Piatetsky Shapiro他,"Accurate Estimation of the Number of Tuples Satisfying A Condition", ACM-SIGMOD,1984
- 3.「並列マシン向けDBMS技術、90年代半ばの実用化めざす」、日経エレクトロニクス、1993.7.19(No.586)