4Q-4

機能ディスクシステム第2版における集計演算処理の考察

中野 美由紀 喜連川 優 高木 幹雄
東京大学 生産技術研究所

1. 始めに

機能ディスクシステム (Functional Disk System with Relational database engine : FDS-R) は，プロセッサ本体と二次記憶システム間のI／Oボトルネックの問題に着目し，二次記憶システムを単なる記憶媒体ではなく，それ自身がデータに対して高レベルな処理機能を持つことにより，関係データベース・システムとしての二次記憶の性能向上を図ったものである。特に，大容量のステージング・バッファと複数台のプロセッサを導入することで，大規模データ処理を効率良くかつ高速に行っている。また，関係代数演算を支援する専用ディスク・コントローラを新しく開発し，専用ハードウェアによる”機能”も実装している。すでにＦＤＳ－Ｒ第１版の試作機を開発し，基本性能についての計測を行い，一般の商用データベース・システムと比較して高い性能を得られることを確認した(1)。続いて第１版で得られた知見をもとに新たにＦＤＳ－Ｒ第２版を開発し，大規模リレーションに対する関係代数演算の処理方式の実装を行い，200MB 程度のリレーションを用いた性能評価の結果，他のデータベース・マシンと比較して十分高い性能を確認した(3)。本報告では，第２版試作機上での集計演算の計測結果を用いて行った解析について報告する。

2. ＦＤＳ－Ｒ第２版における集計演算

2. 1 測定環境

以下に述べる測定結果は，プロセッサは，MC68020 ４台（ローカル・メモリ1MB），ステージング・バッファ６MB，８インチディスク１台を用いて計測された。プロセッサ台数は特に記述しない場合には４台である。ステージング・バッファは，大きさを変化させて測定する。測定用のデータベースのレコード長は１２８バイト固定である。今回の測定では，集計演算の問合せを試作機上に実装されている問合せ解析実行システムのＱＵＥＬサブセット・システム(2)を用いてＱＵＥＬステートメントとして実行している。

性能測定に用いた問合せは以下の通りである。

```
range of e is relation1
retrieve(e.a1, sum=sum(e.a2 by e.a1))
```

フィールドa1,a2 はともに4 バイトの整数である。フィールドa1はパーティション・サイズ分の重複があり，このフィールドによりデータがクラスタリングされる。

集計演算は，実際に扱う必要のあるデータは小さいため，大規模リレーションにおいては，ほぼその処理時間はデータI/O コストとなる。一方，ステージング・バッファ上でのデータ処理では，クラスタリングなどの状況で処理性能が大きく異なる。ここでは，大規模データの処理性能としてNested Loop 方式とGRACE Hash方式をリレーション・サイズとステージング・バッファ・サイズを変化させて測定するとともに，１タスク・サイクルの処理性能としてパーティション数，クラスタ数，プロセッサ台数を変化させて測定を行った。

2. 2 測定結果

(1) リレーション・サイズによる処理時間の変化

図１にリレーション・サイズを変化させた場合のNested Loop 方式とGRACE Hash方式の処理時間の変化を示す。ステージング・バッファ・サイズは64 KB，クラスタ・サイズは平均10タプル，パーティション数はリレーション・サイズの0.1 に設定されている。この計測では，ステージング・バッファを越えるリレーション・サイズを対象としており，図からわかるようにGRACE Hash方式ではリレーション・サイズに比例して処理時間が増加しているが，Nested Loop 方式ではリレーション・サイズの二乗で処理時間が増えている。性能評価に用いた集計演算では，処理の対象となるデータがソースデータと比較して小さいために処理時間はほぼＩ／Ｏコストで占められている。特に，ステージング・バッファを越えるデータが対象となる場合は，演算処理時間のほとんどはデータのＩ／Ｏ時間である。

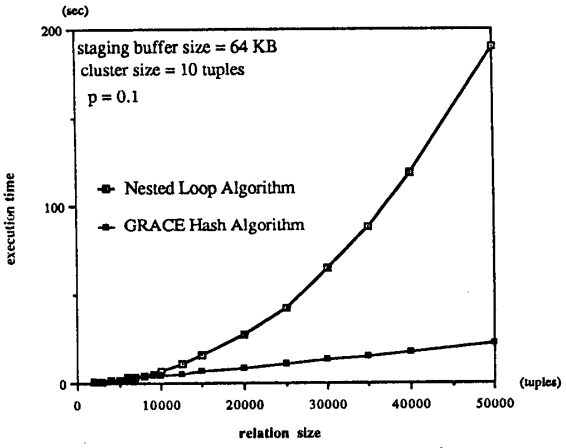

図1．集計演算：リレーション・サイズによる処理時間の変化

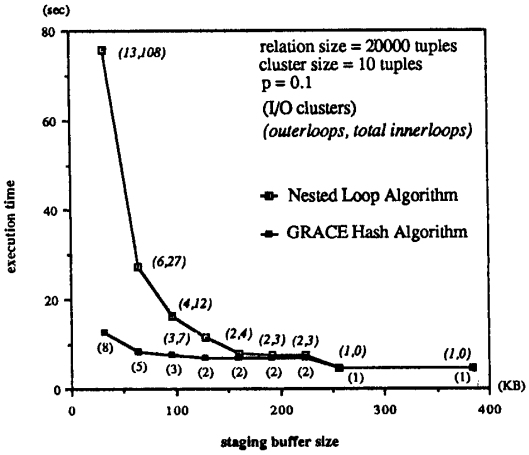

図2．集計演算：ステージング・バッファ・サイズによる処理時間の変化

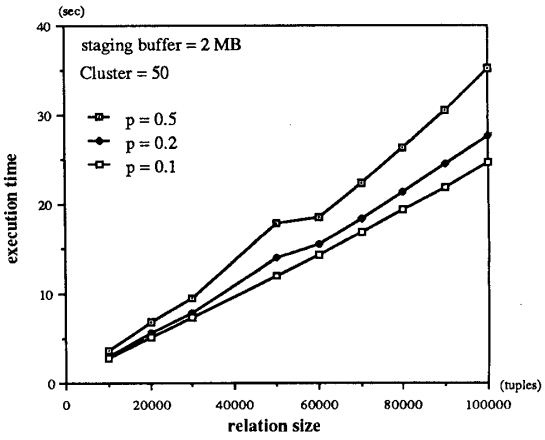

図3．集計演算：パーティション数による処理時間の変化

Aggregation Query on FDS-R2
Miyuki NAKANO, Masaru KITSUREGAWA, Mikio TAKAGI
Institute of Industrial Science, University of Tokyo

(2) ステージング・バッファ・サイズによる処理時間の変化

図2にステージング・バッファ・サイズを変化させた場合のNested Loop方式とGRACE Hash方式の処理時間を示す。リレーション・サイズは20000 タプル，クラスタ・サイズは平均10タプル，パーティション数はリレーション・サイズの0.1 に設定されている。また，図中の()内にGRACE Hash方式の場合はI/O クラスタ数を，Nested Loop 方式の場合はイタリックで外側ループの回数と全内側ループの回数を示す。図3からわかるようにNested Loop 方式と比較してGRACE Hash方式の処理時間が小さい。これはフィルタリング・ファクタが非常に小さいためで，文献(4)で述べたI/O コスト式から推定された結果と同じになる。また，ステージング・バッファ・サイズが小さくなるとNested Loop 方式では急激に処理時間が増加する。これは，ステージング・バッファが小さくなるとその二乗に反比例してネストループ回数が増えるためである。

(3) パーティション数による変化

図3にパーティション数を変化させた場合の処理時間の変化を示す。ここでは，1タスク・サイクルですべてのデータが処理される場合の結果を示す。ステージング・バッファ・サイズは2MB，クラスタ・サイズは平均50タプルとした。パーティション数はそれぞれ，リレーション・サイズの0.1, 0.2, 0.5 と変化している。つまり，リレーションを10000 とすると，各々100, 200, 500 のパーティションができる。処理時間はこの順で増加している。これは，すでに第1版の性能評価で述べているように(1)，パーティション数が増えることにより1クラスタ内でのフィールドa1の比較回数が増えるためである。

(4) クラスタ・サイズによる変化

図4にクラスタ・サイズを変化させた場合の処理時間の変化を示す。リレーション・サイズ100000タプル，ステージング・バッファ1MB とした。また，パーティション数はリレーション・サイズの0.01, 0.1, 0.2, 0.5と変化している。図からわかるように，処理時間はクラスタ・サイズの変化に比例して増加する。これは，クラスタ・サイズが大きくなると1クラスタ内のパーティション数が多くなり，フィールドa1の比較回数が増えるためである。

(5) プロセッサ台数による処理時間の変化

図5にプロセッサ台数による処理時間の変化を示す。リレーション・サイズは100000タプル，ステージング・バッファ2MB，クラスタ・サイズ平均50タプルとした。図からわかるように，プロセッサを増やすことにより，処理時間は低減する。また，プロセッサ台数による処理効果は，データのI/O 時間には影響を与えないため，ステージング・バッファ上でのデータ処理のみの時間をprocessingとして示している。図からわかるように，データ処理時間がほぼ理想的に低減していることがわかる。

図6にデータ処理時間の並列処理効果を示すために，パーティション数が0.2 の場合のデータ処理時間をプロセッサ1台の処理時間をもとに正規化したものを示す。この図からほぼ4台まで，並列処理効果が得られることがわかる。これは，1タスク・サイクルの処理としてはデータ処理時間が重い集計演算において機能ディスクシステムが採用している動的クラスタリング方式が極めて有効であることを示している。

3．おわりに

機能ディスクシステム（ＦＤＳ－Ｒ）第2版試作機上における集計演算の性能評価について報告した。すでに提案したＦＤＳ－Ｒ上での関係代数演算処理方式を用いることにより，ステージング・バッファを越えるリレーションの集計演算を測定し，Nested Loop 方式とGRACE Hash方式の処理コストを考察した。また，ステージング・バッファ上の1タスク・サイクルの処理についていくつかのパラメタを用いて性能測定を行い，機能ディスクシステムの有効性を確認した。

システムの拡張をめざし，ディスクを複数台用いたＦＤＳ－Ｒのアーキテクチャについても検討を行う予定である。

［参考文献］

(1)M. Kitsuregawa, et.al., " Functional Disk System for Relational Database, " Proc. of 3rd Int. Conf. on Data Engeneering, pp. 88-95, 1987

(2)中野　美由紀，他：機能ディスクシステム（ＦＤＳ－Ｒ）における問合せ処理方式　電子通信学会データ工学研究会，DE86-25(March, 1987)

(3)M. Kitsuregawa, et.al., "Query Execution for Large Relation on Functional Disk System," Proc. of 5thInt. Conf. on Data Engeneering, 1989

(4)中野美由紀，他：機能ディスクシステム第2版における関係代数演算処理方式とその評価　アドバンストデータベースシンポジウム，pp. 91-98, 1988

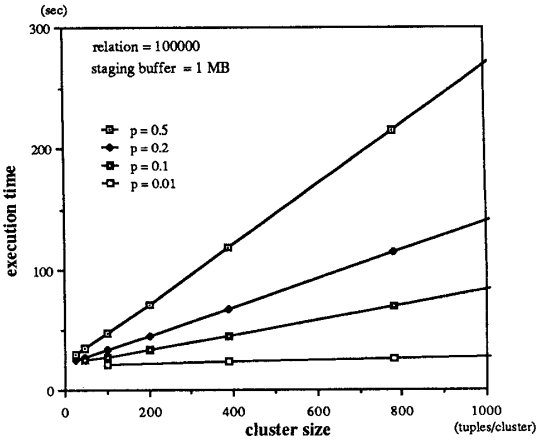

図4．集計演算：クラスタ・サイズによる処理時間の変化

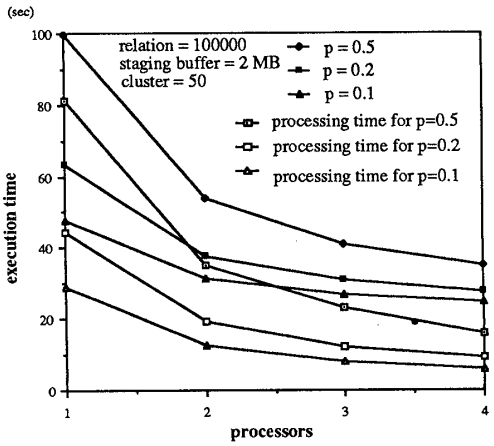

図5．集計演算：プロセッサ台数による処理時間の変化

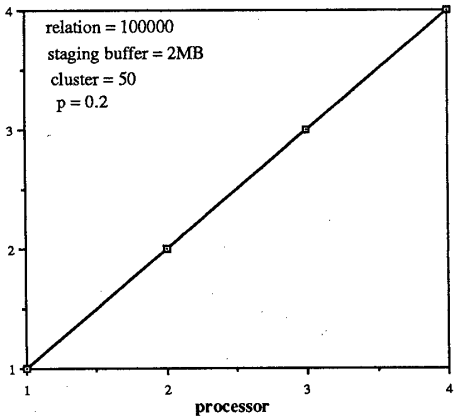

図6．集計演算：並列処理効果